



Predicting Child's Health using Big Data Analytics

Vishal Johri

Department of Computer Science and Engineering
Manipal Institute of Technology, Manipal, India
Bangalore, India.

Varun Singh

Department of Computer Science and Engineering
Manipal Institute of Technology, Manipal, India
Bangalore, India.

Ishan Srivastava

Department of Computer Science and Engineering
Manipal Institute of Technology, Manipal, India
Bangalore, India.

Abstract: An era of open information in healthcare is now under way. We have already experienced a decade of progress in digitizing medical records, as pharmaceutical companies and other organizations aggregate years of research and development in electronic databases. Big Data analytics in healthcare deals with analysis of huge sets of electronic health records which are so complex that they cannot be maintained or analysed by traditional data processing applications. Big Data analytics in healthcare is useful in detecting diseases at early stage when they can be treated more efficiently, predicting outcomes of a surgery, etc. Mother's health during pregnancy and at the time of child's birth affects a range of health related factors in child, perhaps even in long-run. From the previous digital health data of mothers at the time of delivery and health issues in their babies, analysis can be done and prediction model can be created. This model will assist in taking appropriate precautions and treating diseases at early stages which will improve overall health of babies. This paper is structured as follows- first we have listed few health issues in pregnant women which affect their children. Then, we have proposed a model which can help in improving child's health. In the end we will elaborate advantages and challenges faced by the proposed model.

Keywords: Big Data Analytics; healthcare; mother; child; electronic health records; proposed model

I. INTRODUCTION

Data is growing and moving faster than healthcare organizations can consume it; 80% of medical data is unstructured and is clinically relevant. This data resides in multiple places like individual EMRs, lab and imaging systems, physician notes, medical correspondence, claims, CRM systems and finance. Getting access to this valuable data and factoring it into clinical and advanced analytics is critical to improving care and outcomes, incentivizing the right behaviour and driving efficiencies.

Healthcare organizations are leveraging big data technology to capture all of the information about a patient to get a more complete view for insight into care coordination and outcomes-based reimbursement models, population health management, and patient engagement and outreach. The current trend is digitization of this large volume of data. To analyse these huge sets of electronic health records 'Big Data Analytics' come into play because analysis cannot be done by traditional data processing applications. Big Data Analytics in healthcare can help in predicting a disease at early stage so that it can be cured more efficiently and soundly and appropriate precautions can be taken. Moreover, it can be applied to patients' profiles to conclude who will benefit from preventive care and who require diagnosis.

Child's health is affected by mother's health at the time of pregnancy. These effects can be short term as well as long

term. Mother's weight before and during pregnancy, nutritional condition while baby is developing inside mother's womb, overall diet of the mother during pregnancy period, age of the mother, etc can have a great deal of impact on child's health. Additional care is required while treating babies and they fall sick pretty easily because their immune system is not developed fully. Moreover, sometimes it becomes very difficult to treat babies because of the late diagnosis. In many cases if diseases are long-term then precaution is required from a very early stage but it gets delayed because they were not deciphered by doctors. This is not doctors' fault because they see the reports which may look normal till the disease actually affects the child.

To overcome these difficult situations and improve overall health of the child, Big Data Analytics can help a lot. Analysis of mother's health data at the time of pregnancy and diseases/health issues faced by their child can be done. This data can include physician's prescription, pharmacy, electronic health records, machine generated/sensor data, news feeds, articles in medical journals and social media posts including Twitter feeds, blogs, status updates on Facebook and other platforms. Analysis of all this data can be done both for mother and child and a predictive model can be created. This model will help newly born babies to take precautions against the diseases which they can have in future. The early prediction of diseases will help in treating them in more better and efficient way. Big Data Analytics has potential to save lives, improve care and lower costs of treatment of babies by

understanding patterns and trends within data and discovering associations.

In many cases it is very difficult decision to make whether to perform surgery on a baby, but with the help of Big Data Analytics prediction can be made regarding the consequences of surgery. So on the basis of these predictive consequences, decision can be made regarding surgery, and preventive measures can be taken if consequences do not show satisfactory signs. The biggest challenge in Big Data Analytics for child's health is error-free and credible[5] prediction apart from huge volume, velocity and variety of data[1]. Appropriate treatment and/or prevention is of utmost importance for babies, so prediction model should be reliable. Voluminous healthcare data for mother and child can include medical records, radiology images, clinical trials, 3D imaging, genomics, etc. Velocity of data is ever increasing and it includes regular monitoring such as blood pressure readings, operating room monitors for anaesthesia, daily diabetic glucose measurements, bedside heart monitors, etc. These days health data is in multimedia format and is unstructured. So the enormous variety of data which can be structured, semi-structured and unstructured, makes it gruelling for data analytics.

In this paper, first we will focus on how mother's health affects child's health. Then we will propose a model which can help in improving child's health based on mother's and child's health data analytics. Moreover, we will highlight advantages and challenges of the proposed model. Then finally we will conclude with the overall picture of Big Data Analytics for improving child's health.

II. EFFECT OF MOTHER'S HEALTH ON CHILD'S HEALTH

Mother's health at the time of pregnancy and at the time of child's birth plays an important role in child's health. These effects of mother's health can be short-term or long-term and appropriate preventive measures and treatments are required to cure babies if they suffer from these implications. There are various health issues in mothers which can cause problems in their children in distinct ways. Sometimes it takes years or otherwise effects can be seen quicker as well.

Mother's weight before and during pregnancy may affect a range of health related factors in her child, perhaps even in the long run. Nutrition conditions that exist while the baby is developing in the womb, and the nutrition he receives during infancy, affects his development and future health. Slow growth during fetal life and infancy which itself is a consequence of poor maternal nutrition, predisposes individuals to coronary heart disease, type 2 diabetes and hypertension later in life[7]. Not only have physical disorders been linked with poor nutrition before and during pregnancy, but neurological disorders and handicaps are a risk that is run by mothers who are malnourished, a condition which can also lead to the child becoming more prone to later degenerative disease(s). 23.8% of babies are estimated to be born with lower than optimal weights at birth due to lack of proper nutrition[13].

In most cases smoking during pregnancy can double the chances that baby will be born too early or weigh less than 5.5 pounds at the time of birth. Babies born to smokers have a slightly higher risk of heart defects, cleft lip or palate, and possibly other birth defects compared to babies born to non-smokers. The nicotine (the addictive substance in cigarettes), carbon monoxide, and numerous other poisons smoker inhales from a cigarette are carried through mother's bloodstream and go directly to baby[6]. Smoking can cause a shortage of oxygen which can have a devastating effect on baby's growth and development. Moreover, it can cause baby's risk of developing respiratory (lung) problems, birth defects and Sudden Infant Death Syndrome. If mothers are regularly exposed to second-hand smoke(also called passive smoke or environmental tobacco smoke) while pregnant, they will have an increased chance of having a miscarriage, stillbirth, tubal pregnancy, low birth weight baby, and other complications of pregnancy.

Mother's diet at the time of pregnancy also plays a major role. Many studies have shown that mothers who eat high-fat food during their last trimester might put their children at increased risk of obesity later in life[3]. Mother's diet before pregnancy can affect many aspects of child's health. Genes clearly play a role in driving an individual's propensity to gain excess weight. If one parent is obese, there is a 50% chance that a child will also be obese. However, when both parents are obese, a child has an 80% chance of being obese[12]. Obesity in child can be caused by various factors like mother's smoking habit during pregnancy, mother's weight gain during pregnancy and mother's blood sugar level during pregnancy.

Mother's who have alcohol at the time of pregnancy affects their child's health in a bad way. Alcohol can cross the placenta to the unborn baby and affect baby's health and development. The alcohol will reach the unborn baby very quickly and its blood alcohol level will be the same as mother's[14]. Miscarriage, stillbirth, premature birth and small birth weight are all associated with a mother's drinking during pregnancy. Foetal exposure to alcohol is also the leading known cause of intellectual disability. Fetal Alcohol Syndrome (FAS) describes the diversity of alcohol effects on a child. The problems range from mild to severe. Alcohol can cause a child to have physical or mental problems that may last all of his or her life.

During the pregnancy time if mother is older, she has more time to develop medical disorders such as diabetes, thyroid disease, high or low blood pressure or fibroids which can affect her pregnancy and the health of the newly born child. Older mothers are at a higher risk of having Down syndrome baby. Children with Down syndrome can have certain physical characteristics, like eyes that slant upwards and outwards, a single crease running across the palm of the hand, low-set ears and small hand. Certain medical conditions which are more likely to occur are heart defects, which occur in one in three children having Down syndrome, digestive tract defects, and sometimes problems with sight and seeing. Some pregnant women over the age of 35 experience pregnancy-induced diabetes (or gestational diabetes). In others the positioning of the placenta can affect pregnancy and also premature birth of the baby[11].

A child born to a mother who was depressed during pregnancy, has a greater chance of depression than a child born to a non-depressed mother. Baby's amygdala, a brain structure important for the regulation of emotion and stress, can be altered by mother's depression at the time of pregnancy[10]. This alteration in amygdala has been assessed in children many years after birth as well. So this signifies that the timings of these alterations is unpredictable.

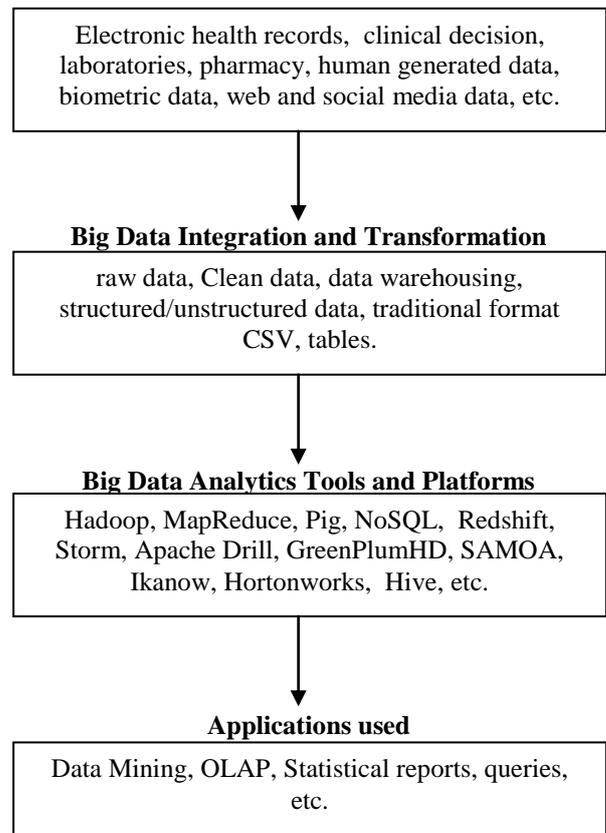
Although many pregnant women with high blood pressure have healthy babies without serious problems, it can be dangerous for both the mother and the fetus. The effects of high blood pressure range from mild to severe. High blood pressure can harm the mother's kidneys and other organs, and it can cause low birth weight and early delivery of child. Some studies have shown that high blood pressure in pregnancy could affect child's IQ in old age[8]. Moreover, there is a greater chance of high blood pressure in a child born to a woman having the same.

A common problem among the babies of pregnant women with diabetes is a condition called "macrosomia," which means "large body"[9]. In other words, babies of diabetic women are apt to be considerably larger than others. This occurs because many of these babies receive too much sugar via the placenta, because their mothers have high blood sugar levels. The baby's pancreas senses the high sugar levels and it produces more insulin in an attempt to use up all the extra sugar. That extra sugar is converted to fat, making a large baby[9]. Gestational diabetes can affect baby in a number of ways like low blood glucose, difficulty breathing and development problems.

Above, we have highlighted few of the mother's health conditions which can affect the health of their children.

III. PROPOSED MODEL FOR PREDICTING CHILD'S HEALTH

In the proposed model we will see the role of Big Data Analytics to improve child's health. Following is the proposed model:



The Big Data Analytics Model is somewhat similar to any other analytics project but the main difference is how the processing is done. The concept of distributive processing is not new but its use in analysing huge data sets of mother and child is extremely important because it can help in making well informed health-related decisions in a sound and efficient manner. The first stage shown in the model depicts different kinds of data sources. For our use case this data can come from internal as well as external sources. Electronic health records, physician's prescription letter, laboratories, pharmacy, web and social media data, readings from remote sensors and meters, biometric data, etc related to mothers and children can be used. Human generated structured and unstructured data including physicians notes, EMRs, e-mail and paper documents can also be used. Unstructured Multimedia data and 3D Imaging are also used as data sources. All this data include mother's data at the time of pregnancy and children's data.

The data collected is raw and in the second step cleaning, integration and transformation of data is done. With the help of ETL(extract, transform and load), data from various sources can be cleansed and readied[5]. Integrating data from one or more disparate sources creates a central repository of data, a data warehouse (DW). Creating a data warehouse requires mapping data between sources and targets, then capturing the details of the transformation in a metadata repository. The data warehouse provides a single, comprehensive source of current and historical information. Apart from storing current and historical data, data warehouses are used for creating trending reports. A service-oriented architectural approach combined with web services can also be used in which data stays raw and services are used to call, retrieve and process the data[5]. Data can be structured, semi-structured or unstructured and

Big Data Sources

Big Data Analytics tools should be able to handle all kinds of data.

The third step shows how the transformed and integrated data is handled by various tools and platforms. The most notable platform for Big Data Analytics is Hadoop[2]. Hadoop is a free, Java-based programming framework that supports the processing of large data sets in a distributed computing environment. Hadoop belongs to the class-NoSQL technologies and has the potential to process extremely large amounts of data mainly by allocating partitioned data sets to numerous servers (nodes), each of which solves different parts of the larger problem and then integrates them for the final result. MapReduce is a programming model and an associated implementation for processing and generating large data sets with a parallel, distributed algorithm on a cluster[2]. The "MapReduce System" (also called "infrastructure" or "framework") orchestrates the processing by marshalling the distributed servers, running the various tasks in parallel, managing all communications and data transfers between the various parts of the system, and providing for redundancy and fault tolerance. Apart from Hadoop there are many other open source platforms and tools available like Apache Drill, Dremel, GreenPlum HD, etc.

Finally, in the last stage some typical applications for Big Data Analytics for improving child's health are shown. Data mining (sometimes called data or knowledge discovery) is the process of analysing data from different perspectives and summarizing it into useful information. Then later this information can be used as a predictive model. Data mining combines many methods from artificial intelligence, statistics and database management. OLAP (online analytical processing) as the name suggest is a compilation of ways to query multi-dimensional databases. OLAP consists of three basic analytical operations: consolidation (roll-up), drill-down, and slicing and dicing. Visualization is an extremely important and difficult theme across all the applications. From the fields such as statistics, computer science, applied mathematics and economics, a wide variety of techniques and technologies has been developed and adapted to improve Big Data Analytics in healthcare[5].

So a predictive model based on the past health data of mother and child is ready which can help the newly born babies. Health data of new mother will be given as input and it can predict health issues which can arise in a newly born child. This will succour not only in quicker and better treatment but also save a lot of cost because precautions and treatment can be started at a very nascent stage. So, basically, this model will help in improving overall health of children.

IV. ADVANTAGES OF THE PROPOSED MODEL

The model proposed in the above section can have a remarkable influence in improving the health of child. It is based on the past health characteristics of mother and child and provides prediction for newly born babies. Based on the health issues in mother, it can tell what all precautions are required in child and it can help in treating diseases at very early stage.

For example if mother is not having sufficient nutrition it can affect child with coronary heart disease, type 2 diabetes and

hypertension later in life. So instead of waiting till child is affected by these diseases, appropriate precautions can be taken and treatment can be started timely which will help in dealing with them in a more efficient and well planned manner. Similarly, many studies have shown that mothers who eat high-fat food during their last trimester might put their children at increased risk of obesity later in life. This model will help in predicting obesity in a newly born child. If it can be taken care at initial stage only then later child can live a happy and healthy life.

This model also has the health data of children. It will help in predicting if surgery is required or preventive measures are good enough. Surgery on small children is extremely complex and dangerous but if it is possible to predict the outcome then appropriate decision can be taken. This model will help in taking required optimal decision. Application of this model to children's profiles will help in identifying if they would benefit from proactive care or lifestyle changes, for example, those children at risk of developing a specific disease.

Moreover, it will help in comparative effectiveness research to determine more clinically appropriate and cost-effective ways to diagnose and treat children. This model will speed up the treatment process and at the same time help in saving lot of money in treatment. It will help in faster development of more accurately targeted vaccine. This model will provide parents with the information they need to make informed decisions and more effectively manage their children's health and more easily adopt and track their healthier behaviours.

V. CHALLENGES

There are some challenges in this proposed model. The criteria for platform and tools may include availability, continuity, ease of use, scalability, ability to manipulate at different levels of granularity, privacy and security enablement, and quality assurance[5]. Mostly the platforms are open source and so the advantages and limitations of open source applies. We need continual technical advances to store and efficiently access the rapidly expanding amount of data. There are also challenges that are particularly salient in healthcare. Concerns about privacy and security are paramount, although these are increasingly being addressed by new authentication approaches and policies that better safeguard patient-identifiable data[4].

The challenge that requires significantly more attention is that prediction is error-free and credible because it will deal with the health of small children. Mistake in prediction can make a very bad impact on the health of the child. The problem in healthcare lies with the quality of data. To derive insights from data, it is critical that they be accurate and relatively complete. When data are systematically biased through either errors or omissions, the correlations that give rise to new insights will be missed or spurious, and could result in misguided confidence or scarce resources dedicated to chasing down dead ends[4]. Critical clinical data like symptoms, physical signs, orders, and progress notes, still rely heavily on human entry and will do so for the foreseeable future. The opportunities to introduce errors are rife (in structured fields, it is easy to select the option above or below the one that was intended). Beyond simple data entry errors, there are bigger, systemic problems with clinical data in electronic format.

Continued investment in technical solutions will undoubtedly improve data accuracy, but without fundamental changes to how care is documented, we should be circumspect about our ability to get rid data of systematic errors[4].

Incomplete data is common in clinical practice and reflect highly fragmented healthcare system where patients see multiple clinicians whose EHRs(Electronic Health Records) do not communicate. Despite significant policy interest, we have yet to achieve any meaningful level of interoperability and without it, creating a comprehensive picture of child's care will be nearly impossible[4]. Handling fragmented data is almost as difficult as handling inaccurate data. Data availability can also be a big problem. The goal is to make identified information available for care and de-identified information available for system improvement, but healthcare data does not cross institutional boundaries due to political conflict and lack of interoperability.

Real-time Big Data Analytics is of utmost importance. The dynamic availability of numerous analytics algorithms, models and methods in a pull-down type of menu is also necessary for large-scale adoption. The important managerial issues of ownership, governance and standards have to be addressed. Data cleansing which is a process of detecting and correcting (or removing) corrupt or inaccurate records from a record set, table, or database, is also a big challenge because of high volume, velocity and variety of data[1]. After cleansing, a data set will be consistent with other similar data sets in the system. The inconsistencies detected or removed may have been originally caused by user entry errors, by corruption in transmission or storage, or by different data dictionary definitions of similar entities in different stores. Moreover, healthcare data is rarely standardised and often fragmented. The different forms of data can be incompatible because they are generated by legacy systems[5]. So these challenges should also be addressed.

There is a considerable shortage of skilled professionals and data scientists who can assess healthcare needs and impacts, write algorithms and work on platforms like Hadoop. To minimize this shortage, many Universities and Institutions are coming up with programmes on Big Data Analytics. For budding data scientists there are enormous opportunities.

The potential of Big Data Analytics for improving child's health is limitless. But we need to understand the issues of data quality and address them effectively. The solutions are not easy. However, ignoring these challenges could quickly lead us from the hope for big data to the disappointing and wasteful results of bad data. These issues require specific attention from policy makers and practitioners, and may be lessened by promoting greater interoperability and reducing burdensome documentation requirements[4].

VI. CONCLUSION

Big Data Analytics has enormous potential to improve overall health of a child. This is so because it provides sophisticated technology to get insight of the previous medical records of mother and child and predict the health of the newly born baby. Looking at various applications and usefulness of Big Data Analytics we can be sure that in the future we'll see the

rapid, widespread implementation and use of it across the healthcare organization and the healthcare industry[5]. Mother's health at the time of pregnancy affects child's health and in the proposed model we are taking advantage of that and developing a predictive model. Moreover, the model also has health data of the child. It will help in guiding if surgery is required or preventive measures are good enough. It will help in treating disease at nascent stage because of which treatment will be less costly and more efficient.

The proposed model will provide parents with the information to help them take pertinent decisions for their children's health. Children can lead a healthy life if precautions are taken at very early stage and diseases are treated more efficiently and methodically. Moreover, it will provide comparative effectiveness research to determine more clinically relevant method for treatment of a child.

There are several challenges highlighted above, which should be addressed. As Big Data Analytics used more and more and become mainstream, issues such as privacy and security, establishing standards, governance and continually improving technology will gain attention. One of the biggest challenges is error free prediction because it deals with the health of small children. Apart from these, challenges like incomplete or fragmented data should also be addressed. Ignoring these challenges could quickly lead us from the hope of improving healthcare with Big Data Analytics to the disappointing and wasteful results.

Big Data Analytics in healthcare is at an initial stage of development but rapid advances in platform and tools can help it in maturing. Once data quality, conformity and fidelity have been optimised, and electronic health records have been integrated worldwide, the applications of Big Data Analytics are limitless in healthcare industry.

VII. REFERENCES

- [1] Cynthia Burgard, "Big Data and Analytics Key to Accountable Care Success", IDC Health Insights, Sponsored by: IBM, pp. 3-4, October 2012.
- [2] Paul Zikopoulos, Chris Eaton, Dirk deRoos, Thomas Deutsch and George Lapis, Understanding Big Data, pp. 53-80, McGraw-Hill, 2012.
- [3] Katie Golde, New study links high-fat diet late in pregnancy to child health risks <http://news.medill.northwestern.edu/chicago/news.aspx?id=226993>
- [4] Julia Adler-Milstein and Ashish K. Jha, "Healthcare's Big Data Challenge ", American Journal of Managed Care, vol. 19, July 2013.
- [5] Wullianallur Raghupathi and Viju Raghupathi, "Big data analytics in healthcare: promise and potential" Health Information Science and Systems, vol. 2, February 2014, doi:10.1186/2047-2501-2-3
- [6] Smoking During Pregnancy, <http://www.webmd.com/baby/smoking-during-pregnancy>
- [7] Chris Kresser, Health begins in the womb - and even before, <http://chriskresser.com/health-begins-in-the-womb-and-even-before>
- [8] High Blood Pressure in Pregnancy, <http://www.nhlbi.nih.gov/health/resources/heart/hbp-pregnancy.htm>

- [9] Diabetes and Pregnancy, <http://www.webmd.com/diabetes/pregnancy-diabetes-and-pregnancy>
- [10] Science Daily - Depression in mothers, <http://www.sciencedaily.com/releases/2013/12/131204090956.htm>
- [11] How does increased age affects pregnancy, <http://www.medic8.com/healthguide/pregnancy-birth/pregnancy/baby-over-35/affect-pregnancy.html>
- [12] Obesity In Children And Teens, http://www.aacap.org/AACAP/Families_and_Youth/Facts_for_Families/Facts_for_Families_Pages/Obesity_In_Children_And_Teens_79.aspx
- [13] Nutrition and pregnancy, http://en.wikipedia.org/wiki/Nutrition_and_pregnancy
- [14] Pregnancy and Alcohol - risks and effects on the developing baby, <http://www.cyh.com/healthtopics/healthtopicdetails.aspx?p=114&np=122&id=1950>