Big data Mechanism Based Analysis of Infant and Child Healthcare: A Case Study for Khorda , Orissa

Meena Moharana[1], Manjusha Pandey[2], Siddharth Swarup Rautaray[3]
[1]Ph.D scholar, School of computer Engineering, KIIT(DU), India, meenamoharana11@gmail.com
[2]Asso.Professor, School of computer Engineering,KIIT(DU), India, manjushafcs@kiit.ac.in
[3]Asso.professor, School of computer Engineering, KIIT(DU), India, siddharthfcs@kiit.ac.in

ABSTRACT

Electronic Health Recording technique generate huge amount of data in healthcare system needs data analytic techniques for proper management of patient's database. Big data technique has been deployed in medical domain as they two are dealing with thousands of data generated in day-to-day transaction. Through different subsystem of Electronic Health Recording technique , patients data generated require proper treatment through data analytic technique. Pediatric system is one of most sensitive and important sub-domain in medical domain need more care in handling child and infant data. Infant and child belong to such age that has tendency of getting affected by such disease like obesity needs early diagnosis to avoid early age disorders. To avoid future life risk conditions need prediction of early age obese factors. Rural, Semi-urban and Urban people follow different life style; disease factors are also different accordingly. To analyses s about different factors of disease, numerous standards are present like BMI for obesity. Taking BMI, the prime attribute further analysis is being needed in prediction of serious disease like pulmonary disease, Type2 diabetics will be caused due to obese.

Key words: Infant and child healthcare, tools and techniques, questioner, data collection

1. INTRODUCTION

Child belongs to age 0 to 5 years is a very sensitive and crucial stage where at each stage of life need proper care and attention. The age needs more care as it is not more communicable rather all is based on observation. It is the stage where maximum possibility of happening the disease either from any infection or the disease due to parental cause. Health care system which deals with the various patients of all age with different cause of occurrence of disease [1]. Accordingly laboratory observation is required. As there are some disease can directly be observed whereas there are lots of diseases need more observations depending on which a physician be able to diagnosis the disease[3],[4],[6]. Healthcare system is a large system based on EHR mechanism. Each day, from the entry of a patients including patients relatives to exit of a patients, all the laboratory observation with physicians advise,

Details of assigned physician with department all the information are generated through EHR system. Huge data flows have been occurred in the system needs proper management with latest analytic techniques. Data analytics technique handles the data flows with maintenance of database in different domain. It has the facility of handling different type of data like structural, semi-structural, continuous and non-continuous data with data transparency and accessible programming tool. R tool provide both statistical and graphical analysis. There are number packages of statistical analysis with different graphical analysis techniques like regression analysis and decision tree analysis. As both i.e medical domain and data analytic techniques dealing with large dataset, we are doing the research by combining this two. Also there are no. of research paper has cited in review section which give a brief idea about how analysis can be done for medical data. It is seen that there are many disease of infectious type like normal flue, any type of skin disease. As child are very much affected by infection due to have low immune system, they need much more care than others. There are many cases where it has seen that some disease of childhood have negative and longer impact on their later age. Hence need proper care for which proper diagnosis of disease with proper treatment [21],[22],[23],[25]. And that is possible when care of data must be managed properly. In our study we under goes filed visit of khorda district in Odisha for primary data collection. Also QoL (Quality of Life) [3],[15]have the effect on to have healthy life. Hence it is necessary of having a well settled data base, with which pediatrician can more effectively find out the case behind disease[32],[27] and can suggest the treatment to patients.

2. REVIEW OF RESEARCH PAPER

In many researches, the biomedical domain has been carried out by applying different analytical techniques. Bio-medical data[11],[16],[17],[24]need more attention as it is the one-to-one procedure of identifying disease to patient based on pathological observation.

Rifat Hossain, et.al., 2018[10], has stimulated the risk factor of causing obesity among the middle-aged people with the help of statistical tools(SPSS) and test the class level attributes according to cross-sectional validation. The author has proposed a risk mining techniques (PRMT) checked

techniques are IBK which is an Instance-Based Learner technique, random forest tree, and logistic analysis. Based on their analysis give a brief idea about the factors that are playing a major role in avoiding obesity in adults.

Before concluding, the factors that are responsible for occurring the particular disease must be known. Many of the time it is found that obesity becomes a prime factor causing many hazardous diseases like cardiovascular disease, type 2 diabetics, high HDL, slow growth in height in children. Saeed Parastar, et, al, 2017[3], has given a brief idea about the development of obesity involves an interaction between genetic and environmental factors, including physical sample sizes and social environment.

Table 1: Literature survey on disease and tool & techniques

Author, year	Title	Proposed Mechanism	Contribution	Challenges
Rifat Hossain, et.al.,2018	PRMT: Predicting Risk Factor of Obesity among Middle-Aged People Using Data Mining Techniques	PRMT techniques have been proposed. Found out Naive Bayes classifier as the best classifier.	Predict the risk factor of obesity in adult and middle-aged people and found out the class level through different classifier with10-fold-validation technique.	Based on the result analysis risk factor avoids obese in adult and middle-aged people.
Saeed Parastar,et,al,2017	Association of urinary concentrations of four chlorophenol pesticides with cardiometabolic risk factors and obesity in children and adolescents	the cross-sectional study has been done and Multivariate linear regression with BMI z-score and WC has also been proposed.	The development of obesity involves an interaction between genetic and environmental factors, including physical and social environment has been found out. The urinary concentrations of trichlorophenols indicating that there is high exposure to the parents of metabolites.	the cross-sectional study and relationships between the various factors cannot be considered to be causal.
Isolina Riaño-Galán, et.al,2017	Proatherogenic Lipid Profile in Early Childhood: Association with weight status at 4 Years and Parental Obesity	*t*-test or ANOVA test with chi square-test for continuous as well categorical variables has been applied for population-based birth cohort study and logistic regression models has been applied to evaluate the risk of the proatherogenic profile on being overweight / obese.	Four-year-old overweight/obese children have higher lipid risk profiles with obese	Obesity persists from childhood to adolescence and into adulthood is a leading cause of health problems
Tejaswi Rohit Anupindi and Padmini Srinivasa n, 2017	Disease Comorbidity Linkages between MEDLINE and Patient Data	Proposed approximate matching strategy for encoding the patient data using 5 digit ICD9 codes to MEDLINE records	Establish an inference rule between patients data and MEDLINE record links through comorbidity associations s and	through comorbidity associations, the large dataset can be connected which will help in retrieving the valuable information in the healthcare system

the 10-fold-cross-validation techniques. And find out the factors which are responsible for cardiovascular disease indirectly happens because of obesity in middle-aged people. Also, the accuracy of p-value has been established and analysis has been carried out through different data mining

The urinary concentrations of tri-chlorophenols indicating there is high exposure to the parents of metabolites. They have made the cross-sectional study and relationships between the various factors that cannot be considered to be causal and suggested that there is a need for cohort and

case-control studies with bigger. Multivariate linear regression with BMI z-score and WC has also been proposed by the author in his paper,' Association of urinary concentrations of four chlorophenol pesticides with cardio-metabolic risk factors and obesity'. In the paper,' Body Mass Index Trajectories in Early Life Is Predictive of Cardio-metabolic Risk', author Yue Yuan, et.al,2019[4] have been proposed three groups with distinct trajectories in BMI and that identified by the latent mixed models with longitudinal analyses. In their study, they have suggested that ethnic population study must be avoided.

Hannah J Wilkie, et.al,2016,[5], has suggested sleep, ST and PA are important behaviors associated with adiposity and further attention needed for child awareness. Multilevel multiple linear regression and BMI z-score and interactions in their respective research work. Jennifer A. WooBaidal,et.al,2015[12], in his paper," Risk Factors for Childhood Obesity in the First 1,000 Days", he has made a systematic study and review on existing evidence for modifiable childhood obesity risk factors present from conception to age 2 years. And found out that the first 1,000 days are a critical period for childhood obesity development and pre-conception weight and gestational weight gain, tobacco avoidance, and healthy infant weight gain and cause behind all and also what precaution must be taken care of.

Sneha B. Sridhar,et.al[2], has suggested that the effect of genetic predictors for obesity may be independent. But gestational weight gain has impact on both the growth and the birth-weight of fetus. In her research paper she has discussed on independent gestational weight gain and fetal growth as it directly relate to obese of a child after birth which is more likely have myriad health problem that continues to adulthood independent of genetic predictors. They have hypothesized that insulin sensitivity and gestational weight gain, both are inversely related.

Like It is necessary to increase the QoL (quality of life) to avoid the cause of infection. With a wide range of analysis of dermatological conditions it is found that different aspects of QoL which may be affected of having a family member with skin disease to other family members, hence needs QoL. Author Basra, M. K. A., & Finlay, A. Y, has discusses in his research paper about QoL and how can one improve it for a healthy life [3]. Jaclyn Havinga et.al.2019,[12], multivariable multinomial logistic regression and multivariable binary logistic regression are two regression technique has been applied for developing the model. Also

status of SCHN has been shown for long term neonatal disease in finding other hazardous disease in next age if children.

0 to 5 years age [1],[13]is a very crucial age as in this age children are very much found of outing, playing with dust, junk food and easily affected through many infectious disease which will becomes more harmful if not cured properly. Table1. depicts how different analytic technique has been used in healthcare system for the better diagnosis of disease. Different author has put their idea and enhanced the research with different techniques of data analytics. We have studied number of research paper, where we found medical domain and data analytics[7][8][9] both are dealing with huge data generated in every day transaction hence need more research on the respective.

3. PROPOSED PROCESS MODEL

Process Model is the graphical structure which helps in depict the flow of work, and data flow in respective domain to take the ongoing research in to a higher step. The process model gives the idea about whole the technique in which the system works.

In our proposed process model, initially data has gathered from the domain in which we are working on. Based on the prime factor data has collected in the sub domain of our research domain. Then it is grouped in to two main parts where we see whether the data collected is digital medium or non-digital medium. As we need to process the data, it is necessary for grouping of data. Many of the input data are in dual format. Like some data are in character while some are in numerical value. Hence need data digitization. Non-digital data were to data digitization where digital form data directly jumps to next step which is preprocessing.
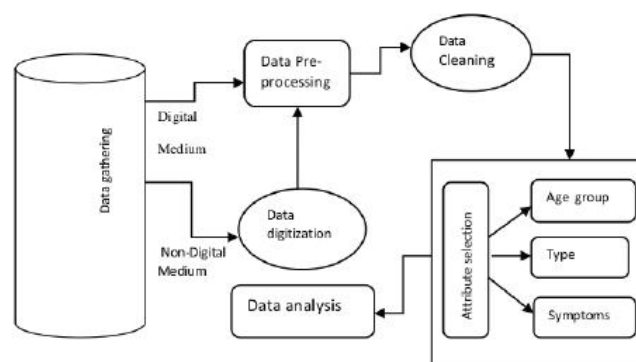

Figure 1: Proposed Process Model

Most important part is data preprocessing in which we have to go for preprocessing of collected data. As some database

contains null value or irrelevant data value which need removal in data cleaning stage. Then after we can be eligible for attribute selection. Attribute selection has a major role in predicting or analyzing the disease and its cause.

Attribute selection plays a crucial role in data analysis. If attribute is well managed then prediction or analysis through selected prime attribute becomes more accurate. For the ongoing research it is very important to do the attribute selection because child life risk is much more than adult as all prediction is based on observation.

4. EXPERIMENTAL SETUP

We have going through different analytics technique. As both the domain dealing with flow of every day data transaction need proper management. R tool provides solid foundation for the application of R language. It has both statistical analysis and graphical analysis for the data of each and every field not specific to one. Due to its simplicity in coding and flexibility of deploying any algorithm it is very popular among data analyst and researchers.

R has the facilities like transparent and accessible programming tool which has collection of libraries for data science. Most important, it has the wide integrated tool that helps in modeling of problem statement. That can be made communicable through different graphs, statistical chart and integrated codes. There are many application areas in R, for which it is much popular among the researcher. In R real world application like data science, statistical computing and machine learning are very popular among different researcher to do their research in different domain on collected data. It has no. of use in predictive and machine learning having different packages like non-linear regression and linear regression, classification of both in linear classification and non-linear classification or clustering of data, decision tree for doing analysis. There are 91000 packages in R for doing statistical analysis. R give a broad plat form for the researcher or data scientist of collecting real world data in doing both statistical and predictive analysis with different machine learning analysis. Also we have seen in our literature survey[5],[6],[11] how different machine learning technique has been applied as per their problem statement both in predicting and analyzing over domain specific data.

For our research we have a well set upped data analytic center of excellence where R tool has installed on the system. As we are working on health care system which generate huge data[30] about patients and their disease need proper management and R tool will be handle such date in a very efficient manner with different statistical and graphical analysis. R tool due to its flexibility and having number of machine learning technique, it will be effective in doing the analysis.

5. DATA COLLECTION AND ANALYSIS

As the needs of ongoing research, we have set up questioners for collecting data from a recognized pediatric center having personalized pathological laboratory. In the initial phase of the collection of data in the healthcare system, we have faced the
language issues. To avoid such a problem, the questioner that has designed in two languages i.e first is in standard American English and second is in local language for the clear understanding of incoming parents with the respective patient i.e kindergartens. For the collection of data, we used the urban area as our research field. Mostly lifestyle in urban is luxurious, lavished hence influence the intake calorie capacity on an everyday basis. It has negative effects that affect the health of every age group which will be the cause of many predictable diseases.

The questioner has designed in two different languages, first, one in English and second in Odia which is the local language of that area from where initial data has collected for the research. Class I, Class II and Class III. Class I belongs to the group of sample children. There are the attributes of Class I, we have taken into consideration as the major attribute. Age is the attribute that classifies the group of sample children as ongoing research which bounds to age from 0-to-5 of sample child.

While designing Questioner , we have distinguished the group by male and female in the sex column of sample child. During the data collection, information about the weight of the sample child has been taken into consideration. The weight at birth along with weight at present has been recorded. Including weight, the height of the sample child has observed and recorded as it is a major factor in calculating the BMI (body mass index) [14],[18],[19],[20]of any sample child. As infant and child are very easily infected, hence needs fully vaccinated. To keep on track, Vaccine All/remaining column has been added to the questioner.

Many times it has found that lifestyle harms a healthy lifestyle [30],[34],[35]. Intake calorie-dense and ethnicity/living style are two more attributes have been added into the questioner.

These are the necessary attribute and have given the idea about living status and habits of food or calories[36][37][38][39] in their day to day life.

Class III records the details about clinical history or observation of father along with paternal family members if any. In the respective class III, income source, living style, habits of intake food, anyone is suffering from any inherited disease or any disease currently suffering from, that whole family members including father of sample child is being recorded through designed questioner.

Table 2: Designed Questioner for collecting data

Class I ବର୍ଗ-୧ Child(ଶିଶୁ ବର୍ଗ)	Sex (ଲିଙ୍ଗ)	Weight In birth (ଜନ୍ମ ବେଳେ ଓଜନ)	Weight at present (ବର୍ତ୍ମାନ ଓଜନ)	Height at present (ବର୍ତ୍ତମାନ ଉଚ୍ଚତା)	Vaccine All/remaining (ଟୀକା)		Intake calorie-dense (OE) (ପୋଷଣ ଗ୍ରହଣ)		Ethnicity/living style(OE) (ଜୀବନ ଶୈଳୀ)	
Class II ବର୍ଗ-୨ Mother (ମାତୃ ବର୍ଗ)	Gestational period(ଗର୍ଭ ଧାରଣା ସମୟ)					Normal period(ସାଧାରଣସମୟ)				
	Age (ବୟସ)	Height (ଉଚ୍ଚତା)	weight (ଓଜନ)	Ethnicity (OE) (ଜୀବନ ଶୈଳୀ)	Clinical observation(OE)(କ୍ଲିନିକାଲ ଅବଲୋକନ)		Ethnicity (OE) (ଜୀବନ ଶୈଳୀ)	Carrier(y/n)if then why (ରୋଗ ସୂତ୍ରଧାର)	Clinical observation (OE) (କ୍ଲିନିକାଲ ଅବଲୋକନ)	
Class-III ବର୍ଗ-୩ Family (ପରିବାର)	Status(living style, source of income, family history) Open-ended(OE) (ପରିବାର ଜୀବନଶୈଳୀ, ରୋଜଗାର ପନ୍ଥା,ଅତୀତରେ ହୋଇଥିବା ଗମ୍ଭୀର ରୋଗ)									

Class II belongs to the mother of that sample child whose clinical information has been taken into consideration through our designed questioner. In this section, we have focused on mother clinical issues again dividing the group into two additional sub-groups as a gestational and normal period. We have recorded the issues that were arising during the gestational period of the mother of particular to sample child.

Table 3: Attributes description

Attribute	Attribute Description
Class-I	It is the section of sample child
Age	Age of 0-5 years child
Sex	Sex of child and infant
Weight in birth	Weight of child and infant
Weight at present	Weight of child and infant in time of data taken

Height at present	Height of child and infant at the time of data taken
Vaccine All/remaining	Give the idea about vaccination of particular sample child
Intake calorie dense	Details of food taken by the patient
Ethnicity/living style	Social attachment, activity, and lifestyle of the patient
Class-II	**Mother categorized under the class-II of respective sample child**
Gestational period	Gestational period of the mother of patients
Age	Age of mother at the time of gestation
Weight	Weight of mother at the time of gestation
Ethnicity(OE)	Living status of the mother during the gestation period
Clinical observation(OE)	Any medical issues at the time of gestation
Normal period	The period describing normal days except the gestation period of the respective mother
Ethnicity (OE)	Living status of the mother
Carrier(y/n)if then why	Whether she would be a carrier of any parental disease
Class-III	**Family status including father of respective patients under this division**
Status(living style, source of income, family history)	Recoded the whole family lifestyle including income source, living style, habits of intake food, anyone suffering from any inherited disease or current disease

5.1 Data Collection

In medical domain, all the healthcare system adopted electronic health records (EHR) model for the synchronization of flows of all the information related to incoming and outgoing patients with all the details. We have deployed EHR techniques for the collection of data from different areas of different QoL of people. EHR helps in recording the information from the entry of patients to exit of patients. It registered all the details of patients with assigned physician. All the observational records along with prescribed treatment including patient's history regarding health issues are stored electronically. We have designed a questioner having attributes of both children and respective parents for the purpose of collecting data including physicians

suggestions if any. We have followed the EHR methodology [35],[40] of collecting of both biological, observational information related to child and his/her parents. Though obese[26],[28],[29] is the prime attribute and children belongs to age group 0-5 years, there are so many factors like pre-gestational and post-gestational information has its own value of having a child early age obesity[31]. We have recorded the information and some sample data with the figures given has given below. Among all whatever data has taken from urban, semi-urban and rural areas are validated by a reputed pediatrician.

5.2 Data Samples

During the data collection period, we came across many people which they have given some relevant information about homemade remedies that they prefer before the pediatrician. That also we have included in our questioner as any homemade remedies for the child.
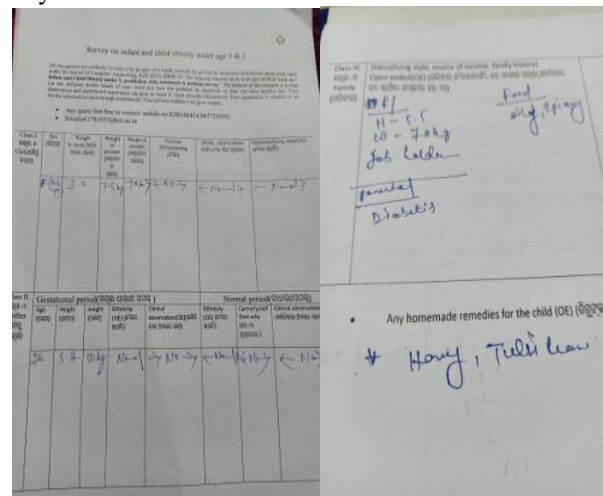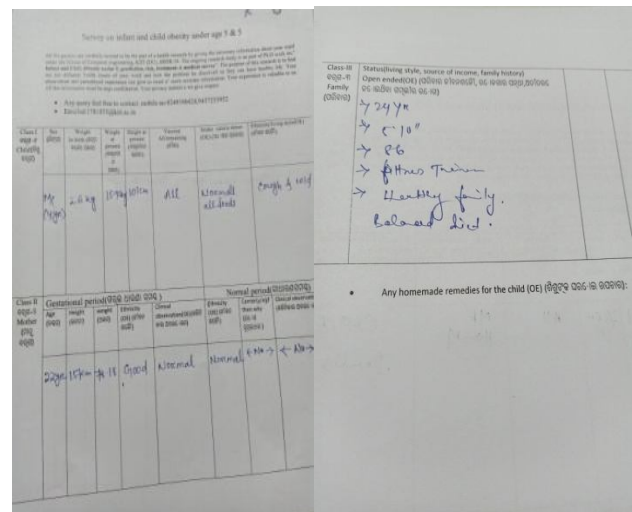

Figure 3(a)

Figure 3(b)

Figure3 (a) & (b): Sample data collected from the urban area

There is some data sample collected for the respective work given below. Figure 3(a) & (b) and Figure 3(c) & (d) showing data samples of patients in urban and semi-urban areas having different ages between 0-5 years of age. To have a well-formulated process model we need verities in collected
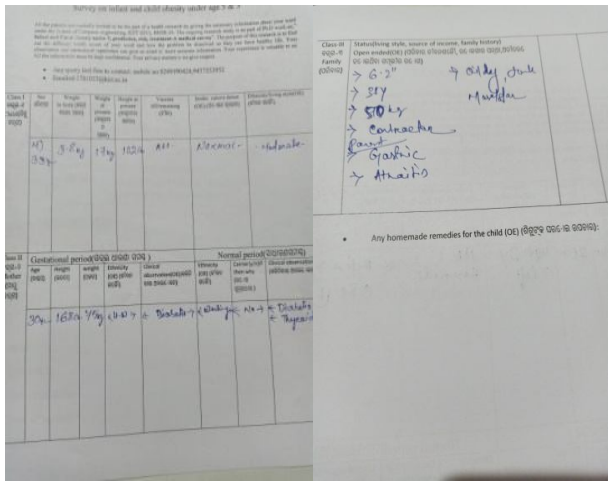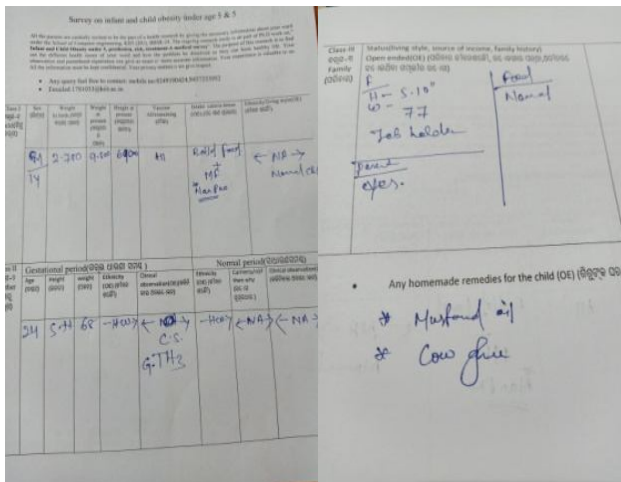

Figure 3(c)

Figure 3(d)

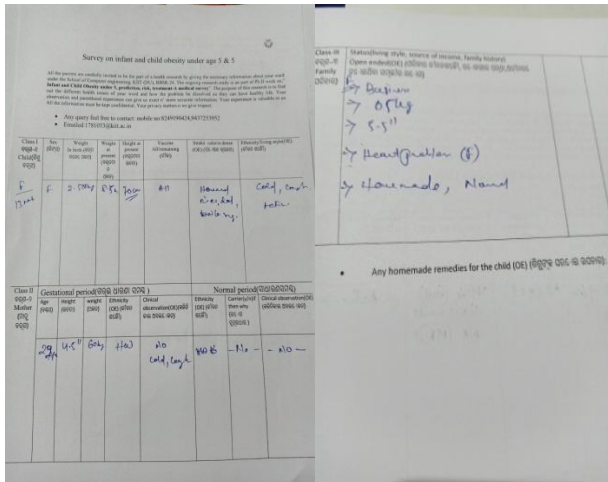**Figure 3(c)&3(d):** Sample data collected from the semi-urban area
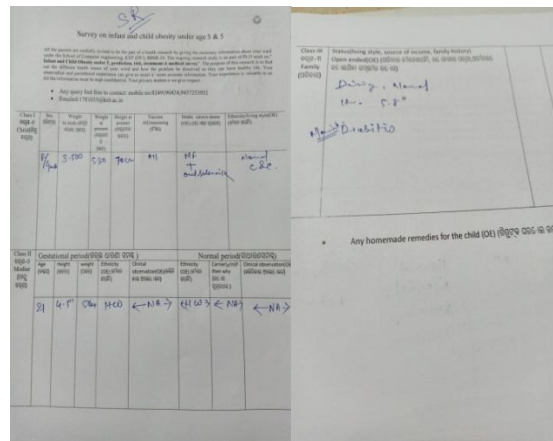

Figure 3(e)

Figure 3(f)

Figure 3(e) & 3(f): Sample data collected from the rural area

data. In Rural areas people are less illiterate need much knowledge regarding do and don't about health issues. Given Figure 3(e) and 3(f) are showing the data sample of rural areas in khurda district. The three different data has been collected for enhancing our research more appropriately. While collecting data, all the patients share their information with a positive vibe which was very much helpful in collecting information. Sometimes a problem arises in a rural area during the collection of information. There are many patients are unable to give the information because of less illiterate and need to demonstrate each attribute very clearly. But they had given the information with positive participation in the survey.

5.2 Data Analysis

QoL is different according to the civilization of people. From the collected data we have found that so many factors affected the QoL of people. According to the collected data samples from different area, we have found their many differences in different factor values. If we take the Age to height, in a particular age their height values different for different civilization. From the analysis, it has come to on notice that intake calorie is directly proportional to have a good height. All the observations also validate by the pediatrician. Though child obesity is the prime factor of our research, we have given much effort on child issues including their birth issues to their maternal background. Various maternal factors have analyzed as these are the root cause of child birth. Pre-gestation and post-gestation period both have the impact on child health issues. From the analysis of data sample if the mother of respective child gains too much weight during the gestation period or she has pre gestational weight, it is more like to have early obese. It is better to avoided early age obesity in child for a healthy life. All the factors like maternal and paternal factor has analyses. In our taken sample data , it seen that most of the parents belongs to urban areas follow a lavish lifestyle. Hence there is more chance to have child

obese compared to other two semi-urban and rural areas. It may be due to genomic or hereditary cause or ethnicity of life and intake habits of calories of parents. The child who are suffered from child obese, there is more possibility of having life risk disease in their adulthood. As per different area child issue are different, needs more attention with more analysis on child and maternal factors which is our research goal. Upon collected data samples, processing of data will be carried out to have a well managed predictive model.

6. CONCLUSION

Infant and child need much attention compared to adult age as their age group is very sensitive to persistent disease that has prolonged survived in them. As data flows in huge amount need proper storage system with advance analytic techniques which are possible in acquired of R tool. It has also seen that R tool is much simpler and user friendly than other tools and techniques. There are two basic causes besides of doing this research on child obesity is to find the prime attribute which are responsible for different life chasing disease and establish a predictive analysis through analytic techniques for the maintenance of database. It will be much helpful to physician in diagnosis of disease with its stage of infection. It will be benefited to both physician, laboratory observer as well patients to extract the information at any time and will be easy for physician to suggest ,advise the patients regarding treatment of patients. As children health issues are very delicate to handle laboratory information should be taken care of very efficiently so that physician can able to do accurate diagnosis with future prediction related to health issues. Hence it is necessary to have well managed, synchronized medical database with latest data analytic techniques as both are dealing with huge data.

REFERENCES

1. Li, Y., An, Z., Yin, D., Liu, Y., Huang, Z., Ma, Y., & Wang, H. (2017). Disease burden of community acquired pneumonia among children under 5 y old in China: A population based survey. Human vaccines & immunot herapeutics, 13(7), 1681-1687.

2. Basra, M. K. A., & Finlay, A. Y. (2007). The family impact of skin diseases: the Greater Patient concept. British Journal of Dermatology, 156(5), 929-937.

3. MacLennan, A. H., Thompson, S. C., & Gecz, J. (2015). Cerebral palsy: causes, pathways, and the role of genetic variants. American journal of obstetrics and gynecology, 213(6), 779-788.

4. Santoro, S. L., Esbensen, A. J., Hopkin, R. J., Hendershot, L., Hickey, F., & Patterson, B. (2016). Contributions to racial disparity in mortality among children with Down syndrome. The Journal of pediatrics, 174, 240-246.

5. Chatzimichail, E., Paraskakis, E., & Rigas, A. (2013, April). An evolutionary two-objective genetic algorithm for asthma prediction. In 2013 UKSim 15th International Conference on Computer Modelling and Simulation (pp. 90-94). IEEE.

6. de Onis, M., Martínez-Costa, C., Núñez, F., Nguefack-Tsague, G., Montal, A., & Brines, J. (2013). Association between WHO cut-offs for childhood overweight and obesity and cardiometabolic risk. Public health nutrition, 16(4), 625-630.

7. Chen, J., Li, K., Rong, H., Bilal, K., Yang, N., & Li, K. (2018). A disease diagnosis and treatment recommendation system based on big data mining and cloud computing. Information Sciences, 435, 124-149.

8. Anupindi, T. R., & Srinivasan, P. (2017, August). Disease Comorbidity Linkages between MEDLINE and Patient Data. In 2017 IEEE International Conference on Healthcare Informatics (ICHI) (pp. 403-408). IEEE.

9. Moharana, M., Pandey, M., & Routaray, S. S. (2020). Why big data, and what it is: basics to advanced big data journey for the medical industry. In Handbook of Data Science Approaches for Biomedical Engineering (pp. 221-249). Academic Press

10. Hossain, R., Mahmud, S. H., Hossin, M. A., Noori, S. R. H., & Jahan, H. (2018). PRMT: Predicting Risk Factor of Obesity among Middle-Aged People Using Data Mining Techniques. Procedia computer science, 132, 1068-1076.

11. Moharana, M., Rautaray, S. S., & Pandey, M. (2019, May). A Survey on Big Data Solution for Complex Bio-medical Information. In International Conference on Computer Networks and Inventive Communication Technologies (pp. 229-237). Springer, Cham.

12. Baidal, J. A. W., Locks, L. M., Cheng, E. R., Blake-Lamb, T. L., Perkins, M. E., & Taveras, E. M. (2016). Risk factors for childhood obesity in the first 1,000 days: a systematic review. American journal of preventive medicine, 50(6), 761-779.

13. Riaño-Galán, I., Fernández-Somoano, A., Rodríguez-Dehli, C., Valvi, D., Vrijheid, M., & Tardón, A. (2017). Proatherogenic lipid profile in early childhood: association with weight status at 4 years and parental obesity. The Journal of pediatrics, 187, 153-157.

14. Yuan, Y., Chu, C., Zheng, W. L., Ma, Q., Hu, J. W., Wang, Y., ... & Mu, J. J. (2020). Body Mass Index Trajectories in Early Life Is Predictive of Cardiometabolic Risk. The Journal of Pediatrics.

15. Wilkie, H. J., Standage, M., Gillison, F. B., Cumming, S. P., & Katzmarzyk, P. T. (2016). Multiple lifestyle behaviours and overweight and obesity among children aged 9–11 years: results from the UK site of the International Study of Childhood Obesity, Lifestyle and the Environment. BMJ open, 6(2), e010677.

16. Kirk Roberts, Matthew S. Simpson, Ellen Voorhees, William R. Hersh.Overview of the TREC 2015 Clinical Decision Support Track. Proceedings of TREC 2015. NIST.

17. Neil R. Smalheiser and Gary Bonifield Two Similarity Metrics for Medical Subject Headings (MeSH): An Aid to

Biomedical Text Mining and Author Name Disambiguation Journal of Biomedical Discovery and Collaboration, 7: e1, 2016

18. Antonelli, J. A., Maalouf, N. M., Pearle, M. S., & Lotan, Y. (2014). Use of the National Health and Nutrition Examination Survey to calculate the impact of obesity and diabetes on cost and prevalence of urolithiasis in 2030. European urology, 66(4), 724-729.

19. Ogden, C. L., Carroll, M. D., & Flegal, K. M. (2014). Prevalence of obesity in the United States. Jama, 312(2), 189-190.

20. Ogden, C. L., Carroll, M. D., Fryar, C. D., & Flegal, K. M. (2015). Prevalence of obesity among adults and youth: United States, 2011–2014.

21. Tsai, A. G., & Wadden, T. A. (2009). Treatment of obesity in primary care practice in the United States: a systematic review. Journal of general internal medicine, 24(9), 1073-1079.

22. Lynch, F. (2016). Understanding hoarding symptoms in children with comorbid ADHD (No. DPsych (Clin)). Deakin University.

23. Scales Jr CD, Smith AC, Hanley JM, Saigal CS, Urologic Diseases in America Project. Prevalence of kidney stones in the United States.Eur Urol 2012;62:160–5.

24. Centers for Disease Control and Prevention: Increasing prevalence of diagnosed diabetes—United States and Puerto Rico, 1995–2010. MMWR Morb Mortal Wkly Rep 2012;61:918

25. Maalouf NM, Cameron MA, Moe OW, Sakhaee K. Metabolic basis for low urine pH in type 2 diabetes. Clin J Am Soc Nephrol 2010;5: 1277–81.

26. Maalouf NM, Sakhaee K, Parks JH, et al. Association of urinary pH with body weight in nephrolithiasis. Kidney Int 2004;65:1422–5.

27. Mohapatra, C., Pandey, M., & Rautray, S. S. (2017, January). Modeling and dynamics of infectious disease: Big data analytics. In 2017 International Conference on Computer Communication and Informatics (ICCCI) (pp. 1-4). IEEE.

28. Curhan GC, Willett WC, Rimm EB, Speizer FE, Stampfer MJ. Body size and risk of kidney stones. J Am Soc Nephrol 1998;9:1645–52.

29. Nowfar S, Palazzi-Churas K, Chang DC, Sur RL. The relationship of obesity and gender prevalence changes in United States inpatient nephrolithiasis. Urology 2011;78:1029–33.

30. World Health Organization. WHO Reference 2007; Growth Reference Data for 5–19 Years. WHO: Geneva, 2010.

31. Swinburn BA, Millar L, Utter J, Kremer P, Moodie M, Mavoa H et al. The Pacific Obesity Prevention in Communities project: project overview and methods. Obes Rev 2011; 12(Suppl. 2): 3–11.

32. Mohapatra, C., Pandey, M., & Rautray, S. S. (2017, January). Modeling and dynamics of infectious disease: Big data analytics. In 2017 International Conference on Computer Communication and Informatics (ICCCI) (pp. 1-4). IEEE.

33. Mathews L, Malakellis M, Sanigorski AM, Kremer P,S winburn BA. Nutrition and Physical Activity in Children and Adolescents. Barwon-South Western Region. Sentinel SiteSeries. Department of Human Services (Victoria): Melbourne,2008.

34. Kelly B, Chapman K, King L, Hebden L. Trends in food advertising to children on free-to-air television in Australia. Aust N Z J Public Health. 2011;35:131–4.

35. Shickel, B., Tighe, P. J., Bihorac, A., & Rashidi, P. (2017). Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. IEEE journal of biomedical and health informatics, 22(5), 1589-1604.

36. Effertz T, Wilcke AC. Do television food commercials target children in Germany? Public Health Nutr. 2011;1–8.

37. Keller SK, Schulz PJ. Distorted food pyramid in kids programmes: a content analysis of television advertising watched in Switzerland. Eur J Pub Health. 2011;21:300–5.

38. Pellai A, Vetrano S, Nobile M, Luti C. The reverse pyramid: a quali-quantitative study about food advertising inside children's television programmes. Minerva Pediatr. 2012;64:15–26.

39. Boyland EJ, Harrold JA, Kirkham TC, Halford JC. The extent of food advertising to children on UK television in 2008. Int J Pediatr Obes. 2011;6:455–61.

40. Sutherland, S. M., Kaelber, D. C., Downing, N. L., Goel, V. V., & Longhurst, C. A. (2016). Electronic health record–enabled research in children using the electronic health record for clinical discovery. Pediatric Clinics, 63(2), 251-268.