VFAST Transactions on Software Engineering
http://vfast.org/journals/index.php/VTSE@ 2014 ISSN(e): 2309-3978;ISSN(p): 2411-6246
Volume 2, Number 1, January-December, 2014

pp. 15-20

ASSOCIATING PERINATAL MORTALITY WITH DIET BY ADAPTING ROBUST CLUSTERING USING LINKS FOR CATEGORICAL VARIABLES

MARIA ABRAR[1], HAMMAD QURESHI[2] AND MAHJABEEN KHAN[3]
[1]Department of Computing, NUST School of Electrical Engineering and Computer Science
10msitmabrar@seecs.edu.pk
[2] Department of Computing, NUST School of Electrical Engineering and Computer Science
hammad.qureshi@seecs.edu.pk
[3]Obstetrics and Gynecology Department, Dow University of Health Sciences, Karachi
mahjabeen.khan@duhs.edu.pk

ABSTRACT. *Perinatal Mortality (perinatal death), is death of a neonate within 6 days (early neonatal mortality) or from 7 – 27 days of birth (late neonatal mortality). Food consumed by an expectant mother is said to have an impact on the pregnancy outcome apart from other factors. For the past few years, perinatal mortality rate has been increasing in developing and under-developed parts of the world. Two-thirds of the world's perinatal deaths occur in only 10 countries, and Pakistan is ranked third amongst these countries. These deaths have not been studied widely, in fact they have been under-reported and these reports have not even been considered in any attempts made to improve birth outcomes in developing nations [1]. Nutritional, socioeconomic, demographic and health advice seeking behavior factors are responsible for higher mortality rates in countries such as Pakistan. Data mining and machine learning can be used to identify factors that are responsible for such high infant mortality rates as it is an important factor indicating progress on Millennium Development Goals. In this paper, we discuss how using ROCK we can cluster expectant mothers as per the food intake and identify major food items causing perinatal mortality.*
Keywords*:* Perinatal Mortality; Categorical attributes; Clustering.

1. **Introduction.** Perinatal mortality or perinatal death refers to the death of a newly born baby within 7 days of its birth or after 28 or more weeks of gestation. Perinatal mortality rates are sensitive indicators of the social and economic levels of any country and high rates reflect the presence of health care and socioeconomic issues. For the past few years, perinatal mortality rate has been increasing in developing and under-developed parts of the world. Perinatal mortality is affected by demographic, environmental, nutritional, socioeconomic and public health-related factors. These socio-economic variables and perinatal mortality is closely linked with each other. Nature of housing, diet intake, sanitary conditions and other economic condition of parents is vital for the health of the neonate. Crowded, unhygienic conditions cause infections leading to perinatal mortality. Age of the mother at the time of birth, birth order and birth interval are important maternal factors which have significant effects on the child's birth. Also infants born to mothers who have a medical history of perinatal mortality have a higher risk [2]. Proper health seeking measure such as vaccinations against preventive diseases and curative measures also affect perinatal mortality rate [2].

Majorly the causes can be divided into neonatal and infectious causes. Many deaths occur due to lack of quality medical advice as there is a severe shortage of gynecologist in most of the developing world. Moreover, complex variables are involved in determining the health of pregnant mothers. In some cases the babies die due to birth complications such as obstructed labor and fetal malpresentation caused by poor obstetric care [3]. Apart from obstetric complications there are external factors involved as well in causing

perinatal deaths. One known factor, said to impact a pregnancy is the food intake by the expectant mother [4] [5]. However, there are no quantitative or automatic methods available to ascertain the impact of food intake on the pregnancy outcome.

Established health care institutions do not provide maternity services which meet the standards of safe pregnancy owing to lack of awareness and understanding in the field of gynecology .They are usually held up by insufficient funds and outdated knowledge, inadequate skills of health care providers, overcrowding of patients, and lack of essential medicines, supplies, equipment, or inadequate hygiene. Medical staff is unable to prevent these deaths and many deaths take place due to a lack of system for detection of any social problems during the pregnancy. Many deliveries take place under the supervision of unskilled labor, 70% [6] of unskilled labor takes place in rural areas leading to health complications and perinatal deaths. A system is urgently required which predicts the outcome of a pregnancy in early weeks of pregnancy and prevents the outcome by taking necessary steps. Building health care systems that can meet the needs of an increasing number of women and infants is a big challenge for a developing country like Pakistan mainly due to lack of funds and untrained medical practitioners.

1.1 Significance of Perinatal Mortality. Perinatal mortality is an important health issue in Pakistan which has a great social impact. In a report by WHO (2005), it was stated that 99% of the world's perinatal deaths occur in low-income and middle-income countries [9]. According to another WHO report (2008), two-thirds of the world's perinatal deaths occur in only 10 countries, amongst which Pakistan is ranked 3rd [1]. South East Asian countries have the highest number of fatal births known as still births. In developing countries, average still birth rate is 26 per 1000 live births which are five times higher than the rate in the developed countries. In a WHO Bulletin published in 2005 [1], out of 1280 births studied in urban areas, 82 children died within 7 days and 96 died within 28 days of birth (Table 1 shown).

Births and mortality rates	**Births/ deaths**	**Males**	**Females**	**Total[a]**
All births,[b] n (%)	1280	619 (51.5)	583 (49.5)	1280 (100)
Stillbirths, [c] rate per 1000 births (95% CI)	43	35.5 (20.7–50.4)	25.7 (15.4–38.7)	33.6 (23.6–43.6)
Early neonatal mortality, rate per 1000 live births (95% CI)	39	35.0 (19.5–50.4)	29.3 (14.7–43.9)	34.8 (24.1–45.5)
Late neonatal mortality, rate per 1000 live births (95% CI)	14	5.5 (0.0–11.8)	19.5 (7.5–31.5)	12.5 (6.0–19.0)
Neonatal mortality, rate per 1000 live births (95% CI)	53	40.5 (23.9–57.1)	48.8 (30.2–67.5)	47.3 (34.9–59.7)
Perinatal mortality, rate-1[d] (95% CI)	82	72.6 (51.2–94.0)	56.9 (37.1–76.7)	70.4 (55.7–85.1)
Perinatal mortality, rate-2[e] (95% CI)	96	77.9 (55.8–100.0)	75.9 (53.3–98.5)	82.5 (66.7–98.3)

CI, confidence interval.
[a] Gender not recorded for 78 births.
[b] Includes the 78 infants without gender data.
[c] The rate of stillbirths was calculated from the number of women with known birth outcomes (n = 1280), while neonatal and perinatal mortality rates were derived from the number whose outcomes were known at 28 days (n = 1121) plus stillbirths, where appropriate.
[d] Rate-1 is stillbirths plus mortality within 7 days per 1000 births.
[e] Rate-2 is stillbirths plus mortality within 28 days per 1000 births.

Table 1 Stillbirth and neonatal and perinatal mortality rates, by gender and in total, in a prospective study in an urban Pakistani population, 2003–2005 [12]

According to a 2010 UNICEF report, Pakistan has a perinatal mortality rate of 60-80 per 1000 births [9]. The numbers will be actually higher since these deaths are not reported in some parts of the country. Millennium development goals (MDG) are set each year by UNDP to increase the living standards in

developing countries. According to MDGs goal number 4(reducing child mortality rates), Pakistan needs to reduce perinatal mortality rate (PNMR) by two thirds by 2015 (Target 4-A) [10]. Due to the current medical situation of the country, achieving MDGs in the current state remains impractical and un-probable. Urgent and comprehensive measures need to be taken by our health organizations to reduce the PNMR in Pakistan.

Perinatal death is currently the second leading cause of child death after pneumonia, which will probably become one of the top causes of death by 2015 [11]. Major changes are taking place in the whole world to achieve the MDGs of maternal and child health. Using computer assisted development we can improve the health conditions and help in covering the Millennium Development Goals.

1.1 Related Work. Many studies have been carried out to investigate the causes of perinatal mortality in the past. In Denmark, a study was conducted by Kristensen et al., investigating the impact of various factors like smoking, caffeine, alcohol, and age etc. on pregnancy by discretization of variables such as BMI (Body Mass Index), age, parity, smoking, alcohol intake etc. [13]. It was found that obesity doubles the risk of stillbirth and neonatal deaths. Similarly, same conclusion was derived by Doherty et al, in their research of pregnancies 16-18 weeks old in which body mass index (BMI) was correlated with the outcome of a pregnancy by creating associations (calculating ratios) between body mass index (BMI), weight gain, and preterm delivery [14]. K-Nearest Neighbor (K-NN) was used by Qureshi et al. to predict the outcome of pregnancy after examining the relationship between body mass index (BMI), pre-pregnancy weight and weight gain during pregnancy [15]. The study concluded that weight gain is an important factor in predicting outcome of a pregnancy. Jerzy et al. studied a machine learning technique which predicted pre-term birth on the basis of less and more specific rule induction. According to Jerzy et al., less specific rules are easier to remember and more efficient with respect to more specific rules in predicting pre-term births [16]. Another study by Jerzy et al. concluded that existing manual methods of predicting pre-term births are only 17-38% accurate, hence a new prediction method was used with the help of LERS (Learning from Examples based on rough sets) [17]. This new classification method increased the prediction rate up to (68-90%).

All of the studies discussed in the above analysis were conducted using statistical analysis techniques such as discretization, chi-square tests, frequency and probability estimation and data mining techniques such as regression analysis for prediction or classification using K-nearest neighbors. Our data consists of categorical attributes for which a clustering technique for categorical variables may be used. Our proposed technique would cluster mothers as per the food items they consume and hence lead to insights in to the causes of high infant mortality.

2 Our Contribution. We have studied the variables affecting perinatal mortal mortality with the help of data mining by applying categorical clustering to the medical records of pregnant mothers. We have adapted RObust Clustering using linKs for Categorical Variables and determined the outcome of pregnancy using ROCK. This algorithm predicts the outcome of a pregnancy after evaluating quantitative data (such as food in our case).

2.1 Material. We applied our technique on real life data belonging to patients from a small town Gadap (population: 448,490) in Karachi, Pakistan. The study consisted of 3 councils from a total of 8. This population belonged to low socioeconomic group who had monthly income less than 3000/= per month (US $ 0.01 per rupee). The data consists of a variety of patients from the district which belonged to different sects of the society. Patients with singleton pregnancies and without any history of diabetes mellitus or hypertension were made eligible for this analysis. The main emphasis of the data lies on the food intake taken in by all the 1039 patients. There are six food groups and 34 food items included in the groups (shown in Table 1). The data is separated into training and test groups. The training (90% of the total data, Alive and Perinatal Mortality separately) and test group (10% of the total data, alive and Perinatal Mortality separately) is randomly selected with replacement for each iteration. The data has been collected from variety of patients (10% living in that district) which belong to different sections of the society and the data is collected in the time period of June 2008 to May 2009. All patients with any medical problems like diabetes mellitus or hypertension were not included. (Shown in Table 2)

The data was cleaned by detecting and removing corrupt records. These records can be:

• Any missing attribute
• Noisy attribute in the data (any invalid entry in the data)

The cleaned data will then be analyzed and used to calculate the outcome of the pregnancy. Our data consists of categorical attributes for which a clustering technique for categorical variables may be used. Our proposed technique would cluster mothers as per the food items they consume and hence lead to insights in to the causes of high infant mortality.

Meat	Beef, Chicken, Fish, Mutton
Vegetables	Beans, Brinjal, B-root, Potato, Mix Vegetables, Carrot, Green Onion, Methi, Palak, Saaq, Sarsoon
Fruits	Apple, Banana, Grape, Guava, Malta, Orange
Diary	Kheer, Milk, Yoghurt
Cereal	Chana, Chana Daal , Masoor, Legumes, Moong, Arhar
Fats/Oil	Butter, Ghee, Oil, Paneer

Table 2 Food Intake

2.2 Methods. The process of grouping a set of physical or abstract objects into classes of similar objects is called clustering [21]. Clustering is a very useful technique to find out hidden patterns and discover interesting ordinality in the data. A cluster thus contains objects which are similar to each other. Similarity between different items in a data base is measured using different techniques of clustering. ROCK (Robust Clustering Using Links), developed by Guha et al. (2005), focuses on interconnectivity amongst features whereas it ignores cluster proximity [18]. It is closest to our problem since it works best with categorical data by using the concept of links instead of distances.

A binary variable will have only two states: 1 and 0. In this study 1 will denote food item is present and 0 denotes the variable is absent in the patients diet. In our perinatal mortality data we can easily denote alive and dead outcome by a simple binary value of 1 or 0. Categorical data is similar to binary variables but with more than two states. These states can be numbers, symbols or any other textual values related to real time data. Similar objects can be identified within categorical data by calculating the ratios of the mismatches between the items. So our categorical data is converted to binary data. During this process of data transformation our 34 food items are converted to binary values of 1 and 0. ROCK is then applied to this binary matrix. Since the perinatal outcome is already present in our data set, we divide the matrix into two parts: C_A and C_B. All patients whose perinatal outcome is alive are moved to C_A and the remaining grouped in C_B.

ROCK computes the links between all the pair of points in each group C_A and C_B and creates clusters based on the interconnectivity of points. Any pair of point is defined as a neighbor if they satisfy the similarity threshold and hence considered close to each other. Links are created between neighbors who are identified using the similarity function. The similarity measure can be any distance function. In our research, we have used Jaccard coefficient as the similarity function. This generates uniform sized clusters for categorical data which become the first level clusters. A threshold is taken as input for calculating the neighbors. If the similarity value of pair of points exceed the threshold they are considered as neighbors. The total common neighbors for the items are the links between the neighbors. All the points in the same cluster will have a larger number of neighbors' hence higher value for the links. So the points with the most number of links are clustered together.

The two set of selected food items from ROCK are then compared with each other to find their significance factor of Alive and Perinatal outcomes; this highlights the items present in abundance in both groups. Significance factor is calculated by comparing the food item and its links in both the data sets having Alive and Perinatal Mortality outcomes. Higher significance factor shows higher effect of the food item in the specific cluster and low significance factor show negligible impact on the outcome.

ROCK clusters the food data according to the connectivity. The links from ROCK show the connectivity of food items within the Alive and Perinatal Mortality group. Food items which are found more abundantly in

the group are more responsible for creating links which highlights the items playing an important part in group C_A and C_B respectively. The food items with a higher number of links are marked as potential causes of perinatal mortality or good for pregnancy respective to the group they belong to.

3. Results and Discussion. According to ROCK results, food items were marked good for pregnancies and some of them were marked as a risk. As a result of the first iteration, with the threshold 0.50, we easily identified the major food items which affect the perinatal outcome. One of the major clusters found in the first iteration and their variations in C_A and C_B are shown in Table 3. A clear cluster is seen in Diary food group (Table 3) which shows the presence of Kheer (famous diary product of Pakistan which consists of milk and rice) and yoghurt in successful pregnancies. Another significant cluster was found in Fats/Oil food group. Paneer was a positive contributor in alive pregnancies and ghee was found abundantly in perinatal mortality.

Food	C_A	C_B
kheer	16%	13%
milk	68%	78%
yogurt	16%	9%

Table 3 Food Group: Diary Product

Food	C_A	C_B
ghee	38%	53%
oil	6%	5%
paneer	24%	3%

Table 4 Food Group: Fats/Oils

According to these details, food items were marked good for pregnancies and some of them were marked as a risk. According to the result of the first iterations Moong, Paneer, Yogurt, Kheer were given the highest ranks for Alive cluster and Chana Daal and Ghee was given the highest ranks for perinatal mortality. This highlighted the important food items during pregnancy and filtered out the food items which cause perinatal mortality.

Confirming our result, medical communities have also marked protein, fatty acids, iron and multivitamins as an important factor in birth outcomes [19] [20]. Comparing these studies with the result of this study, yoghurt and kheer which is a source of protein, paneer source of fatty acids and cereals which contain Iron have been considered positive for a neonate and decrease the chances of perinatal mortality.

4. Conclusion. Perinatal deaths are a result of lack of awareness and inadequate care during the ante-partum, intra-partum and post-partum period. In recent years, several data sets have been brought forward regarding perinatal mortality in Pakistan and study the causes. Although many efforts have been carried by our government and non-governmental sector to reduce perinatal mortality such as LHW (Lady Health workers), but still insufficient prenatal care, harmful newborn practices and home deliveries care are highly prevalent. This leads us to important questions of why practices and behaviors have not changed. Developed countries of the world have seen a good decrease in perinatal mortality due to investments in reproductive health and socio-economic conditions of the countries, but corresponding progress in low income and developing countries has been slow.

In this study we used computer assisted diagnostic tools to support the gynecologist in studying pregnancy and predicting perinatal mortality. Using our technique, we can determine which food items give a positive affect and which are responsible for the perinatal mortality .In a country like Pakistan such studies are necessary to help the medical society in not only decreasing the perinatal mortality but to increase the health awareness in all medical society. Our algorithm can be successfully used to cluster pregnant mothers as per the major food items in take and their relation to mortality may be ascertained. In a developing country like Pakistan such researches are necessary to decrease the perinatal mortality and to increase the health awareness in the medical society.

REFERENCES

[1]. http://www.who.int/bulletin/volumes/87/2/08-050963/en/index.html

[2]. M. Arshad Mahmood (2002), Determinants of Neonatal and Post-neonatal Mortality in Pakistan, The Pakistan Development Review 41:4 Part II pp. 723–744

[3]. World Health Organization, 2006, “Neonatal and perinatal mortality : country, regional and global estimates”

[4]. Abu-Saad, Kathleen, and Drora Fraser (2002) "Maternal nutrition and birth outcomes." Epidemiologic Reviews 32.1: 5-25

[5]. Lechtig, Aaron, et al (1975) "Effect of food supplementation during pregnancy on birth weight." Pediatrics 56.4 : 508-520.

[6]. http://www.pakservice.com/Health/health_home.html

[7]. Geneva: World Health Organization (2005) , “The world health report 2005: make every mother and child count”

[8]. http://www.who.int/bulletin/volumes/87/2/08-050963/en/index.html

[9]. http://www.unicef.org/infobycountry/pakistan_pakistan_statistics.html#94

[10]. http://en.wikipedia.org/wiki/Millennium_Development_Goals#Goal_4:_Reduce_child_mortality_rates

[11]. Li Liu, Hope L Johnson, Simon Cousens, Jamie Perin, Susana Scott, Joy E Lawn, Igor Rudan, Harry Campbell, Richard Cibulskis, Mengying Li,Colin Mathers, Robert E Black, for the Child Health Epidemiology Reference Group of WHO and UNICEF, “Global, regional, and national causes of child mortality:an updated systematic analysis for 2010 with time trends since 2000”

[12]. Imtiaz Jehan, Hillary Harris, Sohail Salat, Amna Zeb, Naushaba Mobeen, Omrana Pasha, Elizabeth M McClure, Janet Moore, Linda L Wright & Robert L Goldenberg (2009), “Neonatal mortality, risk factors and causes: a prospective population-based cohort study in urban Pakistan”, WHO Bulletin Volume 87, Number 2, 130-138

[13]. J. Kristensen, M. Vestergaard, K. Wisborg, U. Kesmodel, and N. Secher (2005),“Pre-pregnancy weight and the risk of stillbirth and neonatal death,” BJOG An International Journal of Obstetrics and Gynaecology, vol.112, no. 4, pp. 403–408.

[14]. D. Doherty, E. Magann, J. Francis, J. Morrison, and J. Newnham (2006), “Prepregnancy body mass index and pregnancy outcomes” International Journal of Gynecology and Obstetrics, vol. 95, no. 3, pp. 242–247.

[15]. Hammad Qureshi, Rehan Hafiz, Mahjabeen Khan, Syed Mustafeel Aser Quadri , “Association of Pre-pregnancy Weight and Weight Gain with Perinatal Mortality”

[16]. Jerzy W. Grzymala-Busse, Linda K. Woolery, “Improving Prediction of Preterm Birth, Using a New Classification Scheme and Rule Induction”

[17]. Jerzy W. Grzymala-Busse, Linda K. Woolery, “A Comparison of Less Specific Versus More Specific Rules for Preterm Birth Prediction”

[18]. Guha, S., Rastogi, R., & Shim, K. (2000). ROCK: A robust clustering algorithm for categorical attributes. Information systems, 25(5), 345-366.

[19]. Abu-Saad, Kathleen, and Drora Fraser (2010). "Maternal nutrition and birth outcomes." Epidemiologic Reviews 32.1 : 5-25

[20]. Lechtig, Aaron, et al(1975) "Effect of food supplementation during pregnancy on birthweight." Pediatrics 56.4: 508-520.

[21]. Han, J., Kamber, M., & Pei, J. (2006). Data mining: concepts and techniques. Morgan kaufmann