IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received September 17, 2018, accepted October 28, 2018, date of publication November 1, 2018, date of current version December 3, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2879115

Cause Analysis of Caesarian Sections and Application of Machine Learning Methods for Classification of Birth Data

SYED ALI ABBAS[1], RABIA RIAZ[1], SYED ZAKI HASSAN KAZMI[1], SANAM SHAHLA RIZVI[2], AND SE JIN KWON[3]
[1]Department of CS & IT, University of Azad Jammu and Kashmir, Muzaffarabad 42714, Pakistan
[2]Raptor Interactive (Pty) Ltd., Eco Boulevard, Centurion 0157, South Africa
[3]Department of Computer Engineering, Kangwon National University, Samcheok 25806, South Korea

Corresponding author: Se Jin Kwon (sjkwon@kangwon.ac.kr)

This work was supported in part by the Basic Science Research through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant NRF-2017R1D1A3B04031440 and in part by 2017 Research Grant from Kangwon National University.

ABSTRACT There are several physical and social factors that are associated to the maternal health and may be considered influential towards the C-Section across the world. Several studies have been conducted in different regions of the world, mostly targeting the pregnant women in specific region. The dynamicity of pregnancy and differences among women with respect to region and social life enforces the researchers to widen the sphere of research at regional level to comprehensively explore the significant risk factors associated to mother and expected child. In this paper, the region of interest is the native city of authors lacking medical facilities and proper pregnant women healthcare infrastructure. As compared to advanced countries, no such study is ever conducted in this region that involves cause analysis of factors resulting in enhanced cases of C-sections and assisting physicians via providing decision support systems based on knowledge induced from machine learning approaches. The aim of this paper is twofold. The first objective is to collect data regionally, in order to conduct local study, first of its kind in this region, and acquire results that are helpful for public health offices in decision making. Second, it is desired to produce different birth classification models and study their applicability on birth data collected previously. The best approach on the basis of correct classification may later be used to produce decision support systems to assist physicians to gain knowledge from the hidden patterns in data. The success of such study is crucial as it will open the doors of interdisciplinary research in two distinct fields of the region.

INDEX TERMS C-section, normal delivery, automated medical diagnosis, machine learning.

I. INTRODUCTION

A Caesarian or C-section is a method through which babies are delivered by surgical incision in the mother's uterus and abdomen. The occasions that require Caesarian are birth of multiples mostly twins, triplets or more, if there is a substantial infant, any previous birth by surgery, or due to any other serious conditions like the baby is in breech or transverse position and so forth. In short the basic reason for the C-section preference in the medical scenarios is, when the health or life of the expected child or mother is at risk [1]. In most of developed countries the rate of C-section is very high. About 23 million C-section were recorded in 2012 across the world [2]. In Pakistan about 50% deliveries are suspected to be conducted in homes, along with a significant growth in C-section deliveries in hospitals and private clinics.

It is important to identify the risk factors associated with C-section deliveries to eradicate the problems or medical complications that affect the mother and caesarian child for the rest of their life. It is believed that the factors that contribute towards C-sections are treatable if discovered well before they impact. It is also observed that expecting women underpass through more or less same kind of experiences during different gestational periods. In such circumstances, the historical data plays a vital role. The historical data may help to conduct predictive analysis of current case under observation. It is further believed that for a physician, it is not possible to catch every contributing factor out of several by just looking at the data. It is very hard for a human brain to identify the patterns in data files. This job is to be done by some learning algorithm, which may memorize the patterns out of data and produce valuable information to physicians.

The application of machine learning towards prediction and classification related with health care is growing gradually. A number of researches have been conducted to strengthen the decision making processes by providing automated solutions based on machine learning methods. It has been observed that the machine learning algorithms are strong in unveiling the hidden patterns in historical data, which are useful in deducting several measures related with subjects under observation. It is believed that machine learning methods and automated solution based on these methods can assist doctors/physicians to take timely decision related with health care. In current study, we aim to analyze the usefulness of supervised machine learning over classification problems, associated with health care, or more specifically to observe the efficacy of supervised machine learning methods to classify subjects either Caesarian sections and Normal delivery subjects.

II. LITERATURE REVIEW

A number of researches have been carried out centered at the medical diagnosis regarding predictive analysis and usage of machine learning methods. The outcomes of such experiments are useful in variety of ways. For example, the reliable computational frame work helps physicians to incur useful information from decision support systems in order to take effective measures in response to the treatment. The reliable predictions acquired using artificial intelligence techniques may help the policy makers and government organization to foresee the reasons towards the problem under observation and take effective measures towards its solution. The sphere of researchers inclined towards medical prediction using machine learning methods is gradually increasing and each and every research contributes to this field. The following section provides the reference of few researches which evidence the utilization of different computational methods, their usefulness and weaknesses towards predictive analysis.

Dulitzki *et al.* [3] identified the relationship between maternal age and C-sections by using multiple logistic regressions. The authors observed that the women in the age of 44 or above are more probable of C-section birth than the women between the ages 22-29. The authors explained that the women with maternal age of 44 or more has more chances of medical complications such that diabetes and hypertension.

Leone [20] emphasized that the prediction of premature/preterm births is viable by monitoring Electrohysterography (uterine electrical signals). They used a dataset comprising 262 term and 38 preterm records in order to classify the records. They compared their research outcomes with existing studies and improvements were identified by the authors with respect of sensitivity and specificity. Cleary-Goldman *et al.* [4] used statistical analysis to identify the relationship between the maternal ages and C-section deliveries. The author divided the age in three groups. The first group of women having age less than 35 years second group of those women, who have ages between 35-39 years and third group of those women, who have age more than 40 years. The author describes that the age of 35-39 are at the increasing risk of miscarriages and felt chromosomes abnormalities. The women of age 40 or above have very high risk of gestational diabetes placenta Persia, placenta abrupt and C-section delivery.

Sana [21] applied mixed models to analyze the relationship between C-sections and socioeconomic factors. They concluded that wealth, age, and education and ultrasonography are combining factors that affect the decision towards the mode of delivery [21].

Adashek *et al.* [5] used multiple logistic regression, t-tests and chi-square to identify relationship between C-section and other factors like age and weight. They identified that if birth weight of expected child is more than 3600 grams and the patient is of age 35 years or less, then the child is more expected to be delivered by C-section.

Ludvigsson and Ludvigsson [6] used student's t-test, Fisher's exact test, leaven's test and multiple linear regressions to analyze the significant factors. Authors observed that if parent's suffered from any coeliac disease the new born will have lower birth weight and shorter pregnancy duration.

Robu and Holban [7] used Naive Bayes algorithms on the analysis and classification of 2086 data instances related to birth data. He used J48, k-NN, Random Forest (RF), SVM, AdaBoost, LogitBoost, JRipp, REPTree, and Simple Cart.

Machine learning and statistical analysis was applied on 9,419 perinatal records in [8]. Their designed prototype expert system provided better accuracy rates compared to those achieved by manual pre-term labor and delivery risk scoring tools.

Another study [8] attempted to improve accuracies achieved by different labor and delivery tools by applying statistical analysis and machine learning on perinatal records. In this attempt, they developed a prototype expert system. In a study [18], authors used multivariate approach to study the effect of psychosocial variables on the complication of deliveries. They identified that the life stress and social support were significantly associated to emotional unbalance. In another research conducted in Pakistan, author used different machine learning techniques to analyze the association of risk factors to the expected mode of delivery. They collected data regionally and applied decision trees and multilayer perceptron to classify the records [22].

The previous sections focused to present a study from literature that guarantees the presence of several physical and social factors that are associated to the maternal health and may be considered influential towards the C-Sections across the world. Furthermore, the involvement of machine learning methods towards the predictive analysis regarding birth classification and related medical diagnosis can't be neglected.

III. METHODOLOGY

A. THE DATA SET

The data was collected from two government hospitals of Muzaffarabad. In the first step, data collected from different sources was cleaned and transformed into shape appropriate for experiments. Around 79 different social, physical and gestational factors including 980 subjects were acquired, however, after dimensionality reduction and data cleansing; 23 factors with 488 subjects were used for classification purposes. The observations in collected data include pregnant women in different trimesters ranging from 17 years of age to 45 years. The frequency of C-sections is higher among young women (age 17-20) and women in late 30s (35 and above). On the other hand, the frequency of normal deliveries is observed higher among women in middle age range i.e. 22 to 28 years. The age wise distribution of women is given in Figure 1. Figure shows that the majority of cases fall within 17 to 30 years.

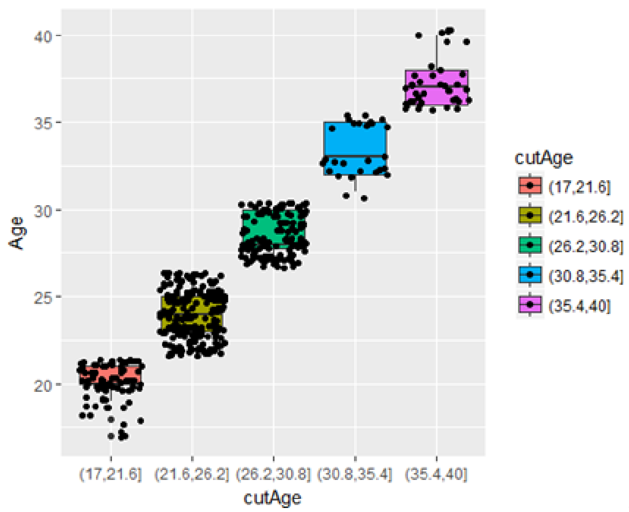

FIGURE 1. Expecting women distribution age wise.

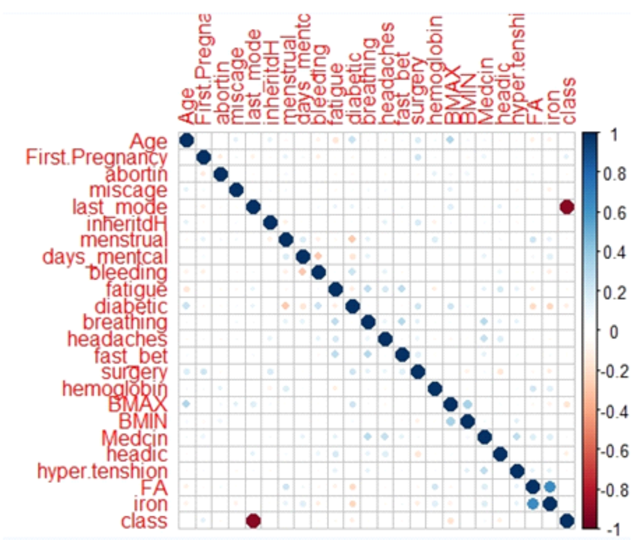

FIGURE 2. Correlation matrix of factors in study.

Among the 23 elements, the dominant factors include age, blood pressures, hemoglobin, mode of last delivery, miscarriages, abortions, hypertension, folic acid, diabetes etc. The relationship between these factors is provided in Figure 2. Data also reveals that the women with C-sections tend to have

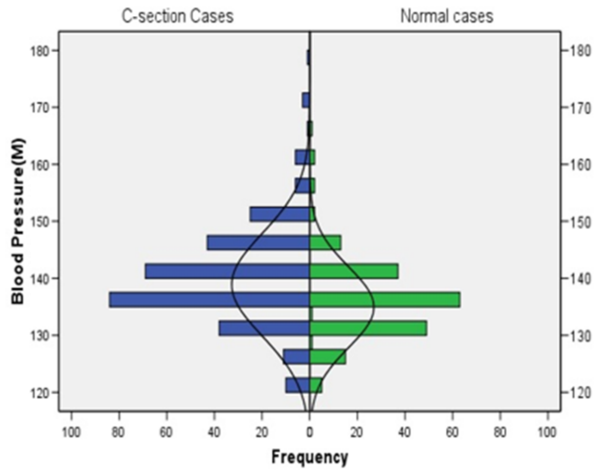

FIGURE 3. Blood pressure maximum vs blood pressure minimum for C-section and Normal delivery Cases.

slightly higher blood pressures as compared to women who delivered naturally, as depicted in Figure 3.The data reveals that age and blood pressure (upper) is positively correlated. The frequency distribution of maximum blood pressure with respect to age is given in Figure 4. Higher blood pressure cases are subject to higher age. The minimum blood pressure scatter in Figure 5 shows that most of the cases tend to fall in 60mmHg to 90mmHg range.

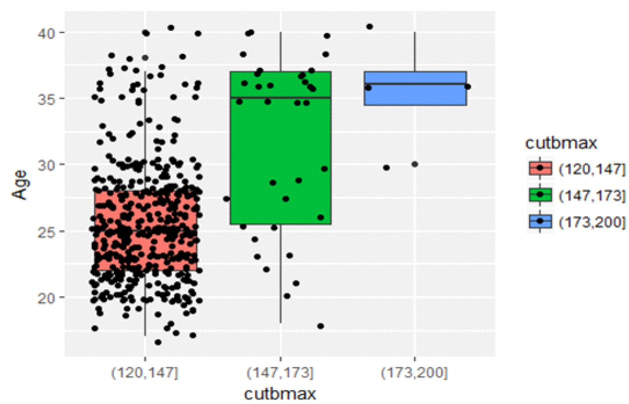

FIGURE 4. Maximum blood pressure with respect to Age.

Diabetic cases also tend to increase with age. It is observed from the data that the last mode of the birth significantly affects the current mode of the birth. If a woman has delivered via C-section in her last pregnancy, then she will most probably deliver via C-section in her upcoming pregnancy. There is not a single observation in our data that evidences the normal delivery following a C-section. Iron deficiency and need to use high doses of folic acid is attributed with increasing age as well. Hemoglobin tends to be lower among women in teen age. It increases in mid and late 20's and decrease again as age increases, as depicted in Figure 6.

B. METHODS USED FOR CLASSIFICATION

Now a day, one of the useful applications of computers and information technology is to assist physicians in medical

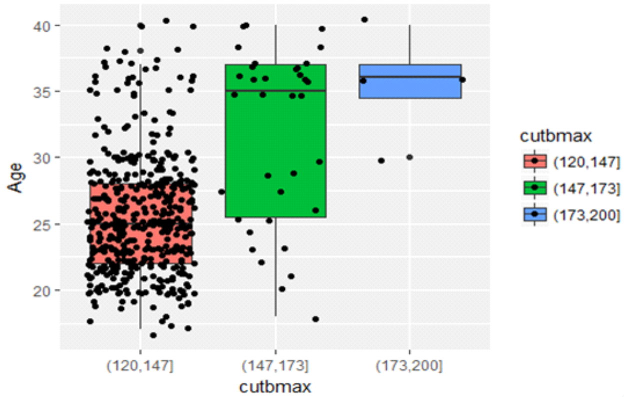

FIGURE 5. Frequency distribution between age and maximum blood pressure.

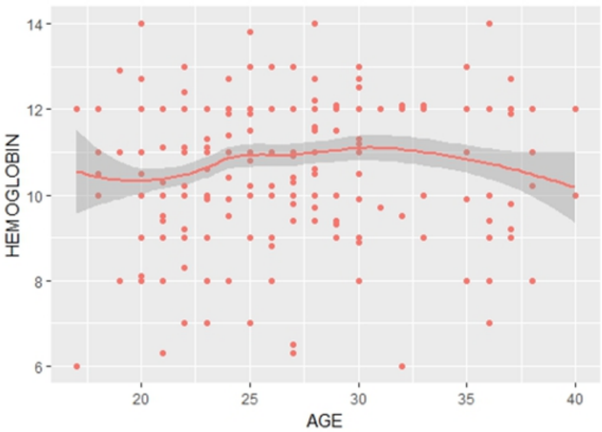

FIGURE 6. Frequency distribution between age and hemoglobin.

diagnosis by providing useful information by analyzing several attributes at once. Several data mining and machine learning methods are available in the industry serving this said purpose. As discussed in literature, there are several researches that have been conducted in predictive analysis using machine learning methods. The most contributing factors are then separated for the birth type classification purposes. In current study, to classify the subjects either C-sections or normal deliveries, classifiers from the different families are assembled. The classifiers used in the current study are Random Forest (RF), Linear Discriminant Analysis (LDA), K-nearest neighbors (KNN), naive Bayes (nb), Neural networks (Nnet), Adaboost, and Support Vector Machine (SVM). The classification parameters are adjusted and to avoid over fitting, 10 fold cross validation [17] is used for the classification of subjects under observation. The classification tasks have been done using R-studio, version 3.2.2.

The Random Forests [9] is a decision tree based machine learning classifier in which the input is provided at the top of the tree which traverses to lower branches of the tree. Every tree is trained on a bootstrap sample of the original training data and determines the split by searching the subset of input variable that are randomly selected. In case of categorical variables, results are achieved on voting majority. In case of mixed variables, result is a weighted average of all of the terminal nodes that are reached.

Support Vector Machines [10] are useful for both, classification and regression. SVM attempts to separate every data element in n-dimensional feature space. Afterwards, SVM computes a hyperplane that separates those data elements into their respective classes.

The k-Nearest Neighbors [11] is a non-parametric method used for both regression and classification. The input is a feature space containing k number of closest training. In case of output, if classification is desired, KNN classifies every object on the basis of majority poll of its neighbors. In case of regression based output, the neighbors to the object are used to provide an average that becomes the regression outcome for that object.

Naive Bayes [12] is a technique used for classification in light of Bayes theorem with a supposition of freedom among indicators. Simply, a Naive Bayes classifier expects that the nearness of a specific feature in a class is inconsequential to the proximity of some other element. For illustration, a natural item can be assumed to be an apple on the probability that it is red in color, curved, and about 3 inches in breadth. Irrespective of the chance that these highlights depend on each other or upon the occurrence of other features, these properties independently add to likelihood that the organic product is an apple and that is the reason it is called 'Naive'.

Linear Discriminant Analysis [13] tries to discriminate between two classes by looking for the best combination of predictors. It estimates the probability of the object with respect of given inputs and class. The object is assigned to that class that gets the highest probability.

Neural Networks [23], learns by comparisons of already known real record with the classification based on its own processing. They have the capacity to feed back the error values from the previous classification back to the network, where this error value is used to adjust the weights for next iterations. During this learning phase, the network trains by adjusting the weights to predict the correct class label of input samples.

Adaptive Boosting [24] can be used in combination with other types of learning algorithms to improve performance. The output of the weak learners is combined into a weighted sum that represents the final output of the boosted classifier. AdaBoost is sensitive to noisy data and outliers.

IV. RESULTS & DISCUSSION

Each classifier was used to generate scores for the entire dataset using a 10-fold cross validation (CV) scheme. In cross validation method, the training part of the data is partitioned into "n" number of sub sets of equal length, so that each subset serves once in training.

The accuracies of different classifiers are presented in Table 1. The validation rate with respect to the training of different classifiers is provided in Figure 7.

Random Forest (RF) has produced the highest mean accuracy of 91.8%. The rest of classifiers i.e. LDA, SVM, Nnet,

TABLE 1. Accuracy and kappa measures of methods.

	Accuracies				
	Min.	1st Qu.	Median	Mean	3rd Qu.
Adaboost	0.805	0.872	0.918	0.901	0.938
LDA	0.805	0.870	0.918	0.906	0.945
KNN	0.756	0.790	0.822	0.828	0.863
RF	0.888	0.891	0.906	0.918	0.944
SVM	0.805	0.871	0.918	0.906	0.945
NB	0.783	0.862	0.89	0.882	0.912
NNET	0.805	0.871	0.917	0.901	0.938
	Kappa Accuracies				
	Min.	1st Qu.	Median	Mean	3rd Qu.
Adaboost	0.585	0.728	0.815	0.784	0.864
LDA	0.585	0.724	0.825	0.800	0.883
KNN	0.436	0.521	0.589	0.600	0.683
RF	0.762	0.766	0.789	0.822	0.881
SVM	0.585	0.724	0.825	0.800	0.883
NB	0.541	0.684	0.756	0.742	0.816
NNET	0.598	0.726	0.814	0.786	0.864

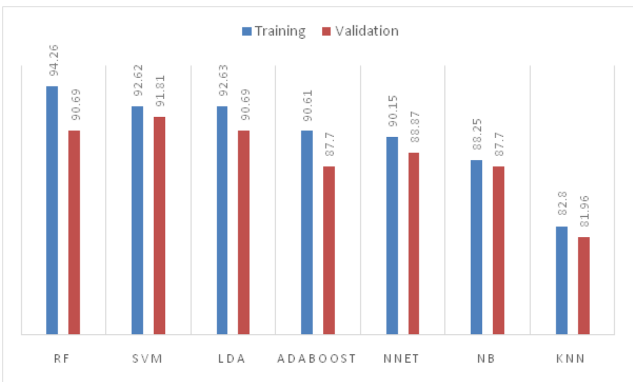

FIGURE 7. Training vs validation accuracies.

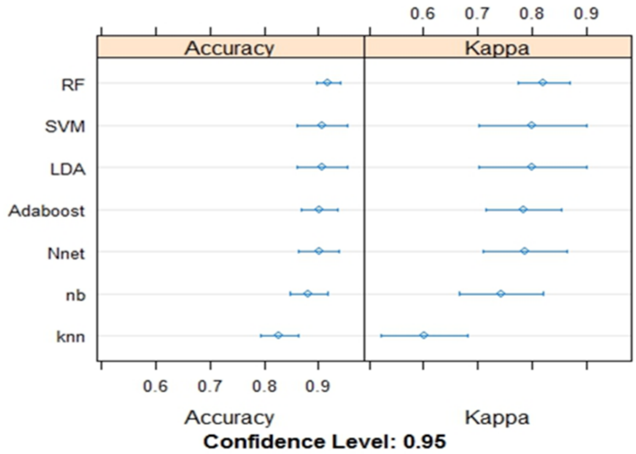

FIGURE 8. Accuracy and Kappa Statistics for methods.

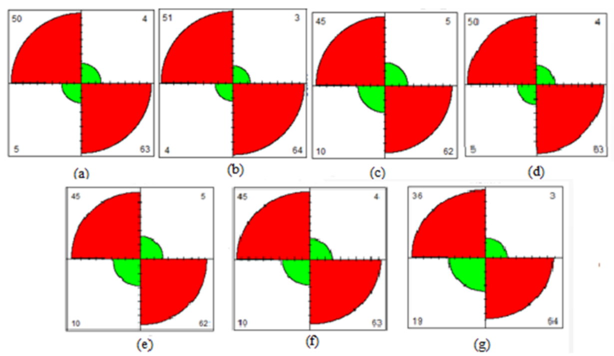

FIGURE 9. Correct vs incorrect classification observations for one run. (a) LDA (b) RF (c) NNET (d) ADABOOST (e)NB (f)SVM (g)KNN.

nb,AdaBoost and KNN have mean accuracies of 90.6%, 90.6%, 90.1%, 88.2%,90.1% and 82.8% respectively. The details of accuracies and Inter classifier variation is provided in Table 1 and depicted in Figure 8.

The inter classifier variation or more sophisticatedly Kappa Statistics [14] is suitable in situations where two or more independent classification methods are evaluating the same thing. The calculation is based on the difference among "observed" agreement (how much agreement is actually present) compared to "expected" agreement (how much agreement would be expected to be present by chance alone). In literature, it is often claimed that in the presence of different number of observations in the class or multiple classes in data set, alone classification accuracy might mislead to judge the performance of the classifier. In such circumstances, the classification accuracy is aided with calculating a confusion matrix that provides a way to understand what classification model is getting right. Once we acquire predicted values from classifiers, we can generate confusion matrix by getting correct predicted values by the class and incorrect predictions made by classifiers for each class. The confusion matrix for classifiers used in the study is provided in Figure 9.

The accuracy of a classifier in order to distinguish C-sections cases from normal cases is evaluated using Receiver Operating Characteristic (ROC) curve analysis [15]. It shows all likely blends of the relative occurrences of correct and incorrect results. ROC provides aims to provide a relationship between the percentage sensitivity and specificity. If the curve is close to the upper left corner, then it represents higher accuracy of the classifier [16]. The ROC comparison of classifiers used in current study is provided in Figure 10.

As mentioned earlier, ROC curves provide the percentages between sensitivity and specificity of the classifiers. The sensitivity or recall is the fraction of relevant retrieved instances over the total amount of relevant instances. On the other hand, the specificity or precision is the fraction of relevant instances over retrieved instances. Higher values associated with precision represents a low false positive rate that means the classifier is returning accurate results. The higher values in recall represent a low false negative rate that means the

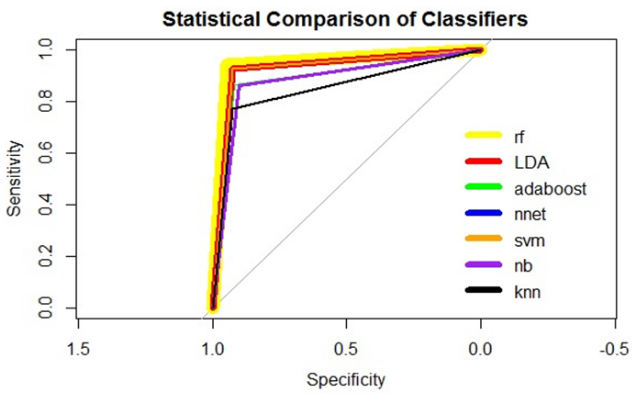

FIGURE 10. ROC curves for classifiers used in study.

classifier is returning a majority of positive results. For our study, the precision and recall values are presented in Table 2.

TABLE 2. Precision and recalls of different classifier (single run based).

Precisions and Recalls (single Run observation)				
Methods	C-sections Cases		Normal Delivery Cases	
	Precision	Recall	Precision	Recall
Adaboost	0.940	0.926	0.925	0.909
RF	0.955	0.941	0.944	0.927
KNN	0.921	0.771	0.923	0.654
LDA	0.940	0.926	0.925	0.909
SVM	0.940	0.875	0.918	0.818
NB	0.925	0.861	0.900	0.818
NNET	0.925	0.861	0.900	0.818

On the basis of results provided, it is claimed that RF performs well as far as this dataset is concerned. Other classifiers have provided remarkably well results, if compared with the existing study performed in the region before [22].

V. SUMMARY

The aim of the current study was twofold. Firstly, an attempt is made to identify the risk factors associated with C-sections among the women in city Muzaffarabad, the Capital of Azad Jammu and Kashmir. In this regard, the data is collected from two government hospitals, i.e., Abbas Institute of Medical Sciences and Combined Military Hospital Muzaffarabad. Secondly, we intend to conduct a predictive analysis on the basis of collected data. As the aim is to predict the birth mode, either C-section or normal delivery, the best way to achieve is to perform classification on data in order to induce knowledge that may help physicians to incur useful information from decision support systems in order to take effective measures in response to the treatment.

In first step the bivariate analysis which is not discussed in the manuscript was performed in order to identify dependency among the factors. It has been observed that majority of C-sections have been reported in age groups less than 20 years (86%) and above 36 (96%) years old. The normal delivery cases with respect to age are observed between 25 to 30 years old women. The analysis revealed that age and last mode of the birth significantly affects the mode of the expected birth. The chances of C-sections in earlier ages and late 40's are higher than among women in middle ages. On the other hand the mode of previous delivery affects the outcome of expected delivery as well. It is evidenced that the women delivered via C-sections tend to have comparatively higher blood pressure as compared to women delivered normally. There are other risk factors that may be associated with C-sections for example heavy breathing, Headache, body pain, fits, previous surgery bleeding in pregnancy, diabetes etc.

The second objective was achieved by using the involved attributes, few mentioned above, for the classification purposes. We applied 10 fold cross validation upon random forests, linear discriminate analysis, support vector machine, naive Bayes and K-nearest neighbors, Adaboost and Neural netto conduct classification and upon the accuracy based results, the best technique was nominated. In current study, for such kind of data set, random forest proved to be a better method as compared to others, as evidenced by accuracy results, specificity and sensitivity. It is claimed that an automated application developed on the basis of Random Forest algorithm will provide an opportunity to physicians use this platform to make prediction for expecting woman by extracting information from historical data based application. This is the first comprehensive study of such kind in the region and its success will open new horizons of interdisciplinary research, hence strengthening the women health and child care field.

ACKNOWLEDGMENT

Special thanks to Dr. Seemab Zafar (Assistant professor and OB-Gyn) and Dr. Maryam Iqbal (OB-Gyn) at Abbas Institute of Medical Sciences, Muzaffarabad, for their guidance during data collection process.

COMPETING INTERESTS

The authors declare that there is no conflict of interests regarding the publication of this paper.

REFERENCES

[1] F. P. Guignard, "A gendered bun in the oven. The gender-reveal party as a new ritualization during pregnancy," *Stud. Religion/Sci. Religieuses*, vol. 44, no. 4, pp. 479–500, 2015.

[2] G. Molina *et al.*, "Relationship between cesarean delivery rate and maternal and neonatal mortality," *JAMA*, vol. 314, no. 21, pp. 2263–2270, 2015.

[3] M. Dulitzki, D. Soriano, E. Schiff, A. Chetrit, S. Mashiach, and D. S. Seidman, "Effect of very advanced maternal age on pregnancy outcome and rate of cesarean delivery," *Obstetrics Gynecol.*, vol. 92, no. 6, pp. 935–939, 1998.

[4] J. Cleary-Goldman *et al.*, "Impact of maternal age on obstetric outcome," *Obstetrics Gynecol.*, vol. 105, no. 5, pp. 983–990, 2005.

[5] J. A. Adashek, A. M. Peaceman, J. A. López-Zeno, J. P. Minogue, and M. L. Socol, "Factors contributing to the increased cesarean birth rate in older parturient women," *Amer. J. Obstetrics Gynecol.*, vol. 169, no. 4, pp. 936–940, 1993.

[6] J. F. Ludvigsson and J. Ludvigsson, "Coeliac disease in the father affects the newborn," *Gut*, vol. 49, no. 2, pp. 169–175, 2001.

[7] R. Robu and Ş. Holban, "The analysis and classification of birth data," *Acta Polytechn. Hungarica*, vol. 12, no. 4, pp. 1–20, 2015.

[8] M. M. van Dyne, L. K. Woolery, J. Gryzmala-Busse, and C. Tsatsoulis, "Using machine learning and expert systems to predict preterm delivery in pregnant women," in *Proc. Conf. Artif. Intell. Appl.*, Mar. 1994, pp. 344–350.

[9] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[10] V. Vapnik, "The support vector method of function estimation," in *Nonlinear Modeling: Advanced Black-Box Techniques*, J. A. K. Suykens and J. Vandewalle, Eds. Boston, MA, USA: Kluwer, 1998, pp. 55–85.

[11] W.-J. Hwang and K.-W. Wen, "Fast kNN classification algorithm based on partial distance search," *Electron. Lett.*, vol. 34, no. 21, pp. 2062–2063, 1998.

[12] R. R. Bouckaert, "Properties of Bayesian network learning algorithms," in *Proc. Conf. Uncertainty Artif. Intell.*, R. L. de Mantarás and D. Poole, Eds. San Francisco, CA, USA: Morgan Kaufmann, 1994, pp. 102–109.

[13] A. M. Martínez and A. C. Kak, "PCA versus LDA," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 2, pp. 228–233, Feb. 2001.

[14] J. Cohen, "A coefficient of agreement for nominal scales," *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 37–46, 1960.

[15] C. E. Metz, "Basic principles of ROC analysis," *Seminars Nucl. Med.*, vol. 8, no. 4, pp. 283–298, Oct. 1978.

[16] M. H. Zweig and G. Campbell, "Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine," *Clin. Chem.*, vol. 39, no. 4, pp. 561–577, 1996.

[17] P. Refaeilzadeh, L. Tang, and H. Liu. *Cross-Validation*. Accessed: Apr. 2018. [Online]. Available: http://www.public.asu.edu/ltang9/papers/ency-cross-validation.pdf

[18] J. S. Norbeck and V. P. Peterson, "Life stress, social support, and emotional disequilibrium in complications of pregnancy: A prospective, multivariate study," *J. Health Social Behav.*, vol. 24, no. 1, pp. 30–46, 1983.

[19] P. Fergus, P. Cheung, A. Hussain, D. Al-Jumeily, C. Dobbins, and S. Iram, "Prediction of preterm deliveries from EHG signals using machine learning," *PLoS ONE*, vol. 8, no. 10, p. e77154, 2013, doi: 10.1371/journal.pone.0077154.

[20] T. Leone, "Demand and supply factors affecting the rising overmedicalization of birth in India," *Int. J. Gynaecol. Obstetrics*, vol. 127, no. 2, pp. 157–162, 2014.

[21] A. Sana, S. Razaaq, and J. Ferzund, "Automated diagnosis and cause analysis of cesarean section using machine learning techniques," *Int. J. Mach. Learn. Comput.*, vol. 2, no. 5, p. 677, 2012.

[22] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proc. Nat. Acad. Sci. USA*, vol. 79, no. 8, pp. 2554–2558, 1982.

[23] Y. Freun and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, 1997.

SYED ALI ABBAS received the Ph.D. degree in computer sciences from Chongqing University, China. He has served public and private sectors in different capacities. He started his career at Higher Education Commission of Pakistan as a Research Associate. Later, he switched to academia by joining the University of Azad Jammu & Kashmir as a Programmer. He served as a Lecturer in computer sciences with the University of Kotli, Azad Kashmir. He is currently serving as an Assistant Professor with the Department of CS&IT, University of Azad Jammu & Kashmir. His research interests include software metrics calibrations in software engineering, image encryption, and design and performance evaluation of machine learning algorithms.

RABIA RIAZ received the master's degree in computer sciences from Fatima Jinnah Women University, Rawalpindi, in 2002, and the Ph.D. degree in electronics engineering from AJOU University, South Korea, with specialization in network security, in 2009. She has more than ten years of teaching and research experience in national and international organizations. She is currently an Assistant Professor with the Department of Computer Sciences and Information Technology, University of Azad Jammu and Kashmir. She has been the Head of the Department of Software Engineering and Department of Electrical Engineering. Her research interests include security of wireless and sensor networks, security and privacy of social networks, cryptography, and network security.

SYED ZAKI HASSAN KAZMI received the Ph.D. degree in computer sciences from The University of Azad Jammu and Kashmir Pakistan. He has served public and private sectors in different capacities. He started his career as a Programmer in software house. Later, he switched to Information Technology Board Government of Azad Jammu & Kashmir as a Database Administrator. He is currently serving as an Assistant Professor with the Department of CS&IT, University of Azad Jammu & Kashmir. His research interests include software engineering and biomedical signal processing.

SANAM SHAHLA RIZVI received the B.C.S. degree in computer science from Shah Abdul Latif University, Khairpur, Pakistan, in 2003, the M.C.S. degree in computer science from KASBIT University, Karachi, Pakistan, in 2004, the M.S. degree in software engineering from Mohammad Ali Jinnah University, Karachi, in 2006, and the Ph.D. degree in information and communication engineering from Ajou University, Suwon, South Korea, in 2010. She has been an Assistant and Associate Professor with the Department of Computer Sciences, MCS NUST and Preston University Pakistan. She is currently the Head of Research and Development, Raptor Interactive (Pty) Ltd., South Africa. Her research interests include flash memory storage, data management, database systems, CPU scheduling algorithms, and wireless sensor networks.

SE JIN KWON received the M.S. and Ph.D. degrees in computer engineering from Ajou University, South Korea, in 2008 and 2012, respectively. He was a Research Professor with the Department of Information and Computer Engineering, Ajou University, from 2013 to 2016. He was a Post-Doctoral Researcher at the University of California at Santa Cruz, Santa Cruz, CA, USA, in 2016. He is currently an Assistant Professor at the Department of Computer Engineering, Kangwon National University, Kangwon-do, South Korea. His current interests include non-volatile memory systems, reliable storage systems, and large database systems.