Grenze International Journal of Engineering and Technology, Jan Issue

Maternal Health Risk Analysis using Support Vector Machine Kernels in Machine Learning

Ramesht Chaturvedi[1], Priyanshu Balmiki[2], Rishabh Jaiswal[3], Gavendra Singh[4] and Suchita Kumari[5]
[1-5]Galgotias College of Engineering and Technology/Department of Electronics and Communication, Engineering, Greater Noida, India
Email: rameshtchaturvedi@gmail.com, {priyanshubalmiki123, jaiswalrishabh297, jadaungs, suchita5445}@gmail.com

***Abstract*— The maternal health includes the health conditions of women in three stages that are during pregnancy, childbirth and the postpartum time. We can also say that maternal health is the health of women at the time during pregnancy, at the time of childbirth and during the raising of a child. The creation of life in the form of a baby is a fulfilling experience to a woman but sometimes a large percentage of women develop many health problems and some even die. Machine learning performs well in the medical domain also. This machine learning allows us to understand the situations of patients by delivering medical information, helps in classifying the diseases, detecting anomalies, in predicting future of any treatment on a disease etc. In this paper we implemented our support vector machine model for maternal health risk analysis. Testing accuracy of this model is 72.906 percent. The training accuracy of this model is 73.4895 percent. We aim to improve maternal health during pregnancy and after pregnancy so that overall maternal mortality decreases.**

***Index Terms*— Machine learning, Morbidity, Supervised Machine Learning.**

I. INTRODUCTION

Maternal morbidity simulates that the huge number of maternal deaths are due to various reasons namely Hemorrhage, Hypertensive disorder as well as sepsis. Not only those, but there were also two other factors which increased maternal morbidity that are Abortion, embolism. This had an adverse effect on a large number of people and has almost led to disaster conditions [1]. In recent studies, it has been observed that maternal mortality is rigorous to obestric care and due to which several women who are at their reproductive age reach such a severe stage that there is no possibility of them left for survival. Even the offspring suffers from this at such a young age along with the mother.

This clearly indicates that there is an urgent need in order to reduce maternal morbidity in the world and thus improve the conditions of people [2]. Maternal health risk states that death of the woman takes place due to complications in the pregnancy after 6 weeks of end of pregnancy. As it has been observed that Obstetric hemorrhage has been one of the reasons in increasing maternal morbidity and in recent times it had an eradicating effect of it on countries like united states which is also the part of the pandemic. So, in order to reduce the morbidity that takes place due to hemorrhage, the data of almost 2,00,000 infants gone under implementation of maternal safety bundles thus resulting in reduction of 20% less deaths compared to the last surveys [3]. Machine Learning, which is generally considered as a subset of Artificial intelligence, provides help with its innovation and development to medical healthcare and takes the responsibility to improve the miserable conditions. In today's

Grenze ID: 01.GIJET.10.1.560

world, machine learning provides extensive tools for various data analysis. But the important features that are required for an effective machine learning algorithm are extraordinary performance, in-depth knowledge [4].There are various machine learning algorithms which make it easier to deal with various problems that arise mainly during machine learning. The machine learning algorithms include Supervised machine learning algorithm and unsupervised machine learning algorithm. Machine Learning and its use have been in a continuous pro-active mode that means a constant increase of the uses of machine learning has been seen in the field of medical imaging which consists of computer- aided diagnosis, medical image analysis etc. There are mainly two big models included in Machine learning specially for medical imaging.

One is a massive training artificial neural network and another one is a convolutional neural network. Among both training models it has been observed that MTANNs is more effective in development, also higher performance and it also takes less time to compute the data. With the various applications where machine learning has been used and has been successful overall. But still there is a lot more modifications that need to be done in order to make machine learning processes more effective and usable, modifications such as give high performance training models which can perform on little data, giving accessibility to data by making the conditions better, using various types of image structures, also looking forward about how to interpret results.

In short, updating all possible applications would make Machine learning the best version of itself and its uses would increase to a much larger extent [5]. These applications of machine learning are seen to be much more growing in the previous two decades due to the credibility of some different factors like complex data availability, growth in various machine learning tools and other. Machine Learning techniques, which include Support vector machines, mainly focus on dealing with large amounts of data with high dimensions as well. Machine learning techniques are just used to enhance features of previously existing data [6].

The data set that we used for maternal health risk analysis is taken from kaggle. This data set is originally from Daffodil International University, Dhaka, Bangladesh. The data preprocessing step is one of the important steps. This data preprocessing includes the number of steps to change raw data into a clean data set to analyse the data set properly [7]. A number of steps involved in data pre- processing are data cleaning, data integration, data transformation and data reduction. Sometimes these data processing steps need to be implemented iteratively to satisfactorily achieve processed data sets[8].

Data pre-processing is to be implemented in a way that doesn't introduce bias by modifying that will impact the outcome of statistical analysis. The raw data is not proper and contains missing values and noise. These missing values and noise lead to inconsistent data sets and data sets need to be cleaned [9]. After the pre- processing of the data set the next step is exploratory data analysis. In exploratory data analysis the basic aim is to investigate the data set for anomalies, distribution and outliers. In exploratory data analysis, there is an availability of many tools to graphically visualize the data that helps in hypothesis generation. A number of exploratory data analysis techniques are available and used depending on the objective [10].

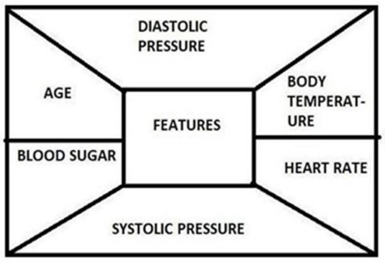

Figure 1. Attributes of the dataset

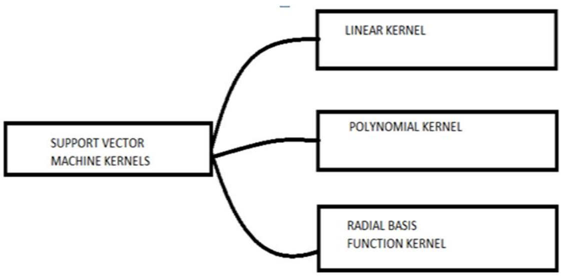

Figure 2. Support vector machine kernels

II. LITERATURE REVIEW

In this Amanda wood and Elliot K. main titled "Reduction of several maternal morbidity from hemorrhage using state perinatal quality collaborative" paper with the use of innovative mentor model technique, it has been observed that in comparison to previous data there has been 20.8% decrease in the maternal morbidity in this collaborative methodology which concludes this technique to be an effective and long-term methodology.

Manjula C. Belavagi and Balachandra Muniyal entitled "Performance evaluation of supervised machine learning algorithm for intrusion detection" paper uses techniques that have been tested with NSL-KDD set and, it has been observed that random forest classifier has been most accurate in the prediction among other Supervised machine learning algorithm for the typical data sets provided.

This Marleen de Bruijne editorial observes and identifies how machine learning techniques has been successful in the image diagnosis area and also clarifies that there is still a scope of improvement in how to train data models, to increase specific features about medical imaging, to predict accurate results effectively.

Ran Neiger and colleagues named, "Long term effect of pregnancy complication on Maternal health" found that there is a powerful relation between obstetric complication in long term effect on Maternal health. It is also found that if the pregnancy has some complications then there is a high chance of developing cardiovascular and metabolic diseases in the mother in future. To lower these risks of developing cardiovascular and metabolic diseases , weight loss, dietary changes and physical activity can be helpful.

In this paper, Akhan and Akbulut ,ultra sonography is able to diagnose 60-70 percent of the anomaly and the remaining anomaly. The prediction result and the ultra sonographic diagnosis were compared and the prediction had 87.5 percent success.

Lokesh Pawar and Janvi Malhotra entitled "A Robust machine learning predictive model for maternal health risk" the proposed model in this paper is evaluated for different machine learning algorithms like Random forest,J48, multilayer perceptron, etc. The maximum accuracy for this model is 70.21.

III. PROPOSED METHODOLOGY

In our proposed work, we proposed a model based on support vector machine kernels and a comparison of training accuracy and testing accuracy among linear kernel, rbf kernel and polynomial kernel is performed.

The data set that is used in our model has the following attributes namely age, diastolic blood pressure, systolic blood pressure, heart rate, blood sugar and body temperature.These above-mentioned attributes help in analyzing the risk level to the health of the mother. The risk level is divided into three parts namely low risk, mid risk and high risk. The data set has a total of 1014 instances. In the first step we applied a data pre-processing step on this dataset like checking the data for null values. There is no null value found in this data set. After checking for null values, we plotted pair plots to visualize the relationship among attributes of the dataset. The selection of suitable boundaries for classifying the risk level into three zones namely low risk, mid risk and high risk can be performed with the help of these pair plots.

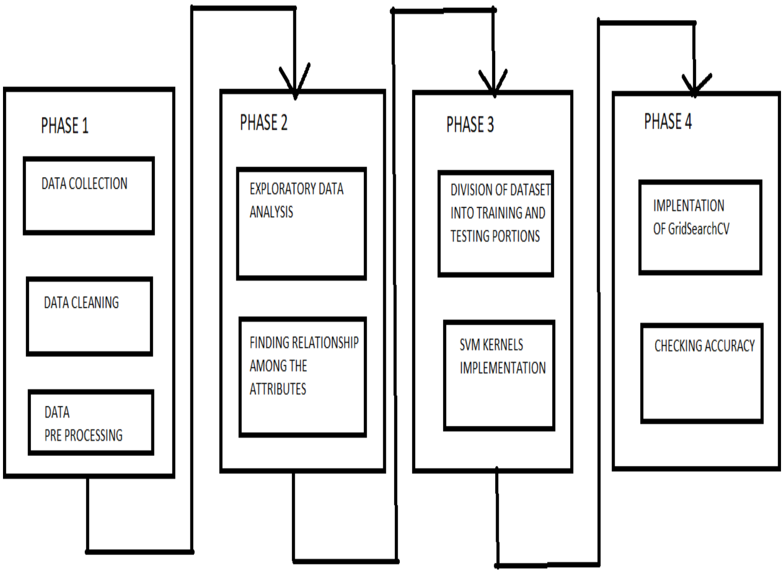

Figure 3. Working model flow diagram

The data set is divided into training portions and testing portions. After this division of the dataset into training and testing portions linear kernel, rbf kernel and polynomial kernel of support vector machine are implemented on the data set.
These three above-mentioned kernels are compared for their testing accuracy. After the comparison, further analysis of the polynomial kernel of the support vector machine is performed on the dataset in order to achieve the maximum testing accuracy.
A loop is then started to evaluate the testing accuracy and training accuracy of the polynomial kernel for different degrees to get the best suited degree of the polynomial kernel of the support vector machine. The best suited degree is selected on the basis of maximum testing accuracy.

IV. Results And Discussion

The relationship among the attributes of the dataset is found with the help of pair plots. These pairplots are very important graphs in determining the relationships among the attributes of the dataset. The benefit of using pair plots is that they plot the relationships among the attributes in a single figure. This makes the comparison among attributes very easy.
This relationship is seen in Fig. 4. The data set is divided into two portions. The training portion has 80 percent of the total dataset and the testing portion has 20 percent of the total dataset.
After the division of the dataset, the SVM kernels are implemented on the dataset and their testing accuracy is calculated and compared. From table number I, it is clear that the rbf kernel and polynomial kernel are better than the linear kernel.

Figure 4. Relationship among attributes

The testing accuracies of different SVM kernels are listed in Table I. After this for the value of C=100, we obtained the testing accuracy and training accuracy for the polynomial kernel for different degrees. This is listed in Table II.

TABLE I. Performance Comparison (Kernels)

S. No	Kernelname	Testing Accuracy	C
1	linear	59.61	100
2	polynomial	68.97	100
3	rbf	69.95	100

After this we use Grid Search CV for the following purposes to find out the best value of C, to find out the best polynomial degree and to select the suitable kernel between rbf kernel and polynomial kernel. After the implementation of this Grid Search CV we obtained the best value of C as 100, the best value of polynomial degree as 5 and the suitable kernel as polynomial. The testing accuracy for C=100, Degree=5 and polynomial kernel is 72.9064 percent.

TABLE II. PERFORMANCE COMPARISON (DEGREES)

S. No.	Polynomial Degree	Testing accuracy	Training accuracy	C
1	1	66.9951	65.2281	100
2	2	66.9951	68.4340	100
3	3	68.9655	71.1467	100
4	4	71.9212	72.5031	100
5	5	72.9064	73.4895	100
6	6	70.9360	74.1060	100
7	7	64.0394	73.2429	100
8	8	62.0690	69.1739	100
9	9	57.6355	68.3107	100

V. CONCLUSIONS

The work of a doctor is easier in today’s world because of the implementation of machine learning in the medical domain. Machine learning not only saves time and money but also delivers suitable interpretations for diseases. This machine learning is now fulfilling the role of an assistant to the doctor. The model proposed in this paper is reliable and has a satisfactorily good testing accuracy.
In future the testing accuracy can be further increased by many approaches like changing the size of the data set. The authors can conclude on the topic discussed and proposed.

REFERENCES

[1] Say, L., Chou, D., Gemmill, A., Tunçalp, Ö., Moller, A. B., Daniels, J., ... & Alkema, L. (2014). Global causes of maternal death: a WHO systematic analysis. The Lancet global health, 2(6), e323-e333.

[2] Pacagnella, R. C., Cecatti, J. G., Osis, M. J., & Souza, J. P. (2012). The role of delays in severe maternal morbidity and mortality: expanding the conceptual framework. Reproductive health matters, 20(39), 155-163.

[3] Main, E. K., Cape, V., Abreo, A., Vasher, J., Woods, A., Carpenter, A., & Gould, J. B. (2017). Reduction of severe maternal morbidity from hemorrhage using a state perinatal quality collaborative. American journal of obstetrics and gynecology, 216(3), 298-e1.

[4] De Bruijne, M. (2016). Machine learning approaches in medical image analysis: From detection to diagnosis. Medical image analysis, 33, 94-97.

[5] Wuest, T., Weimer, D., Irgens, C., & Thoben, K. D. (2016). Machine learning in manufacturing: advantages, challenges, and applications. Production & Manufacturing Research, 4(1), 23-45.

[6] Main, E. K., Abreo, A., McNulty, J., Gilbert, W., McNally, C., Poeltler, D., ... & Kilpatrick, S. (2016). Measuring severe maternal morbidity: validation of potential measures. American journal of obstetrics and gynecology, 214(5), 643-e1.

[7] Belavagi, M. C., & Muniyal, B. (2016). Performance evaluation of supervised machine learning algorithms for intrusion detection. Procedia Computer Science, 89, 117-123.

[8] Pawar, L., Malhotra, J., Sharma, A., Arora, D., & Vaidya, D. (2022, August). A Robust Machine Learning Predictive Model for Maternal Health Risk. In 2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC) (pp. 882-888). IEEE.

[9] Neiger, R. (2017). Long-term effects of pregnancy complications on maternal health: a review. Journal of clinical medicine, 6(8), 76.

[10] Tukey, J. W. (1977). Exploratory data analysis (Vol. 2, pp. 131-160).