A COMPARATIVE STUDY ON MODE OF DELIVERY AND ANALYZING THE RISK FACTORS OF CESAREAN DELIVERY USING K-NEAREST NEIGHBOR, SVM AND C5.0 CLASSIFICATION TECHNIQUES

D.KAVITHA¹, DR T.BALASUBRAMANIAN ²

¹Assistant Professor, Dept. of CSE, KPR Institute of Engineering and Technology, Coimbatore, Tamil Nadu, India
²UG Student, Dept. of CSE, KPR Institute of Engineering & Technology, Coimbatore, Tamil Nadu, India

ABSTRACT

This paper depicts human services in decision making by applying machine learning algorithms on medical data. Health care industry produces huge amount of data that controls complex information relating to patients and their medical conditions. Data mining techniques have the effectiveness to determine relationships or hidden patterns among the objects in the medical data. Most supervised machine learning classification and advancement methods are employed for making decisions. This work focuses on predicting the mode of birth at an early stage by diagnosing the various risk factors. The modes of delivery are vaginal and cesarean. This analyzing helps to predict the birth mode and reduce the cesarean delivery. We examine this system on the collected data and find the best prediction. The physicians can apply this system for making better decisions in emergency cases. In this work, machine learning algorithms are applied for diagnosing the mode of delivery.

Keywords: Machine Learning algorithms, C5.0, SVM and K Nearest Neighbor.

I. INTRODUCTION

Data mining is applied in various areas like pattern recognition, decision making, measuring customer satisfaction, quality performance, improving the efficiency of healthcare, etc. [2]. Today, gathering a wide variety of data about different diseases has noteworthy significance. For many reasons, health centers gather these data. The generic and main reason is simplifying medicament works. Utilizing this data prompts understanding the most ideal administration in least time. The present human services require efficient techniques and research procedures to decrease the expenditure of the medical service, detecting the diseases earlier and saving the lives. Healthcare services predict the patient conditions and their behaviors using the machine learning techniques [6]. New discoveries stated that medicinal services prompted the gathering of huge amount of qualitative data. According to the analysis of McKinsey it is found as big data and Machine Learning algorithms are used for making better decisions, streamlined advancements and improved proficiency of clinical preliminaries.

Data innovations are in effect continuously connected in medicinal services associations so as to react to the requirements of physicians and assist in making decisions at emergency cases. Artificial Intelligence and Machine Learning Techniques are applied in making decisions at emergency cases as well as in primary health services. Artificial Intelligence can be utilized to enable the physicians to analyze the patients particularly in situations when it is difficult to predict the results [2].

This paper includes seven sections. After the introduction, the second section presents machine learning algorithms in medical services. The third section describes the classification algorithms such as Support Vector Machine (SVM), C5.0 and k-Nearest Neighbors (K-NN). Section four describes maternal health care. The materials and methods are described in section five. Section 6 explains the results and discussion. The last section describes the conclusion.
II. MACHINE LEARNING IN HEALTHCARE

In today’s medical services Machine learning is unavoidable. Machine learning serves as a tool for physician’s knowledge. As the population rate increases recording and analyzing the huge amount of patients information is becoming a challenging task. In healthcare system, machine learning is used to manipulate the data automatically which makes the system more robust and dynamic.

Artificial intelligence and machine learning algorithms are used in diagnosing the diseases better and also in advance, treat diseases more absolutely and draw efficient conclusions in patient’s health services [7]. Recent findings have established as algorithms created using machine learning can function similar to a medical practitioner. Recently, machine learning is used in analyzing and predicting the diseases such as heart diseases, cancer, tumor, diabetes, liver diseases and other diseases and processing the medical images. It focuses on new developments in treatments and provides a robust service.

III. CLASSIFICATION ALGORITHMS

Computational algorithms can be developed using machine learning classifiers which can access data and provide better predictions. In Supervised learning, it is necessary to train the algorithm with required data which is already labeled with correct answers. Using the classification algorithms, classes of the data points are determined. Classes are also called as targets or labels or categories. Predictive modeling in classification is the process of relating function (f) from input values (X) to yield desired output values (y). This section describes the machine learning algorithms: Support Vector Machine (SVM), C5.0 and k – Nearest Neighbor (k-NN).

3.1 SUPPORT VECTOR MACHINE (SVM)

Support Vector Machine (SVM) is a supervised learning model with associated learning algorithms used for classification and regression analysis. Pattern classification can be performed using SVM. It is a supervised machine learning algorithm and follows the risk minimization principal. Support Vector Machine is used for both classification and regression, but is mainly used for classification challenges. In Support Vector Machine, the data items are plotted as points in n-dimensional space. Based on the values of the data items, hyper-plane is formed and then the classification is performed by differentiating the classes. In SVM, the hyper plane acts as a boundary between the positive and negative classes.

3.2 C5.0 DECISION TREE CLASSIFICATION

C5.0 is a Classification method which extracts useful information from the huge data. The job of C5.0 is predicting the target class using the values of the given attributes. C5.0 converts trees into set of rules, as it is its important feature. At the point when a standard rule set like this is utilized to group a case, it might happen that few of the principles are pertinent. If the relevant standard rules predict various classes, that could be classified in two different methods: the rule can be taken as the highest confidence; or the rules can be aggregated for making decisions. C5.0 uses the second approach. Each appropriate rule chooses the predicted class by assigning numeric values to the attributes, adding the values and finding the targeted class. The class with the highest value is preferred as the final prediction.

3.3 K-NEAREST NEIGHBOURS (K-NN)

K-Nearest Neighbors algorithm is one of the simplest and essential classification algorithms in machine learning, and it belongs to the supervised classifier. It is a non-parametric method used for both classification and regression. In k-Nearest Neighbor method, the outcome is predicted by creating the model [2]. It uses the $k$ closest classes in both classification and regression and the output depends on either of the cases. It focuses on predicting the target class by finding the nearest neighbor class. In k-NN, the classification is made based on the number of closest neighboring points.

IV. MATERNAL HEALTH CARE

Nowadays the rate of Caesarian section is being increased which is a surgical procedure of delivering a baby. Among European nations, Italy has the most elevated rate of cesarean section (36.8%) and in the Campania, this rate achieved 60% [3]. Taiwan has the next highest rate of cesarean delivery [4]. In Australia, the delivery of babies through cesarean section is almost doubled in the last 25 years [8]. During the last decades, there has been an alarming and dramatic increase in the number of cesarean births in both developed and undeveloped countries [10]. The main factors lead to cesarean section are maternal height, weight, age, twin gestations, arrest of labor, decreased level of amniotic
fluid [5].

Figure 1: A methodology of decision making system based on machine learning classifiers

V. MATERIALS AND METHODS

This study aims at comparing the effectiveness of the machine learning classifiers in making the decisions based on the patient conditions. In this study, the decision making process includes the following process: data collection, testing the classification algorithms and evaluating the classification algorithms as shown in figure 1.

5.1 DATA COLLECTION

The data were collected through questionnaire from various sources such as Government & Private hospitals, primary Health Centre and Public (GPP) during the period December 1, 2015 to December 31, 2017. The data source includes the information about 2021 women who had given birth during the data collection period. The database was comprised of thirty three features. Among those features, six were found to be the most significant factors which lead to cesarean delivery. The dominant factors are age, maternal height [1], multiple gestations, child weight, water level (amniotic fluid) in the uterus at the time of delivery and arrest of labor during the delivery time which are described in table 1. Breech presentation is also an additional factor for cesarean delivery [9].

Table 1: Feature Description

<table>
<thead>
<tr>
<th>Classification</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binary value( if age&lt;30 then value is 0 else 1)</td>
<td>Age</td>
</tr>
<tr>
<td>Binary value( if height&gt;5 then value is 0 else 1)</td>
<td>Maternal height</td>
</tr>
<tr>
<td>Binary value( if babies&lt;1 then value is 0 else 1)</td>
<td>Multiple gestations</td>
</tr>
<tr>
<td>Binary value( if weight&lt;3 then value is 0 else 1)</td>
<td>Child weight</td>
</tr>
<tr>
<td>Yes, No(1,0)</td>
<td>Water level (amniotic fluid)</td>
</tr>
<tr>
<td>Yes, No(1,0)</td>
<td>Arrest of labor</td>
</tr>
</tbody>
</table>
5.2 TESTING THE CLASSIFICATION ALGORITHMS

Classification is the process of training the machine learning algorithms. Machine learning focuses on training the algorithms to automate decision making process from the given dataset. Testing of algorithms is performed by importing and loading the maternal dataset in R programming.

5.3 EVALUATING THE CLASSIFICATION ALGORITHMS

To evaluate the algorithms, the dataset is divided into training dataset and testing dataset. During the first stage, the algorithms were applied on training dataset and the performance was evaluated. Later, the algorithms were applied on testing dataset to make predictions.

VI. RESULTS AND DISCUSSION

During the dataset training, the positive values of the attributes were assigned 0 and the negative values were assigned 1 and classified the prediction class as vaginal delivery and cesarean delivery. The machine learning algorithms were compared based on the performance of Kappa statistics, sensitivity and specificity and described in table 2.

In this work, R programming open source software was used for calculating the performance of machine learning algorithms. Here the Kappa statistics, Sensitivity and Specificity increased based on the accuracy and less error rate in our training dataset. As you see, C5.0 yields high accuracy rate (100%) while compared with SVM and KNN. Therefore, it is suggested that C5.0 can be a useful algorithm in making healthcare decisions.

Table 2: Evaluation Results on the Training Data

<table>
<thead>
<tr>
<th></th>
<th>C5.0</th>
<th>k-NN</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy in %</td>
<td>100</td>
<td>90</td>
<td>89</td>
</tr>
<tr>
<td>Kappa statistics</td>
<td>100</td>
<td>80</td>
<td>79</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>100</td>
<td>82</td>
<td></td>
</tr>
<tr>
<td>Specificity</td>
<td>100</td>
<td>99</td>
<td>99</td>
</tr>
</tbody>
</table>

Figure 2: Experimental Result of SVM
VII. CONCLUSION

Predicting the mode of delivery is an important task as the life of mother and child is dependent on this decision. Machine learning technique has brought great significance in medical data analysis and guides the physicians in decision making. Machine learning classifiers such as Support Vector Machine, k-Nearest Neighbor and C5.0 were utilized in this study for predicting the mode of delivery. The performance of the machine learning algorithms was calculated by testing the algorithms with the maternal dataset. The algorithms classified the mode of delivery into two classes: vaginal delivery and cesarean delivery. Support Vector Machine classified the mode of delivery with an accuracy of 89%, K-Nearest Neighbor with 90% and C5.0 with 100%. Based on the experimental results, it is found that C5.0 performs well with the best accuracy and can be applied in medical data analysis. It is also found that maternal height, age, twin gestations, baby weight, breech presentation, decreased level of amniotic fluid and arrest of labor at the delivery time are the leading factors for cesarean delivery. Machine learning algorithms can be helpful in making a prior decision and reducing the rate of cesarean delivery. In this work, the algorithms were applied for predicting the mode of delivery and finding the major factors for cesarean delivery. Similarly, machine learning algorithms can be applied to analyze different diseases.
Reference


