Supervised Learning Techniques for Analysis of Neonatal Data

Rudresh D. Shirwaikar, Nikhit Mago, Dinesh Acharya U, Krishnamoorthi Makkithaya, Govardhan Hegde K*
Dept. of Computer Science & Engg.
MIT, Manipal University
Manipal, India
**nikhitomatic.mit@gmail.com*

***Abstract*— In the current healthcare setup, the use of Machine Learning has been limited as clinicians diagnose and administer treatment manually. Supervised Learning techniques can help solve many prognostic problems and help clinicians in taking decisions pertaining to healthcare. The paper presents selected machine learning techniques that can be applied for medical data, and in particular some supervised learning techniques with their applications on the analysis of neonatal data. The goal of the paper is to review and discuss the methodology, advantages and disadvantages of supervised learning techniques and the use on neonatal data. In addition, this paper also highlights the model evaluation parameters and also suggests the ways to improve the performance of a model designed for neonatal data analysis.**

Keywords—Supervised Learning; Classification; Neonate; Healthcare

I. Introduction

With the advancement of computers and in artificial intelligence, started the concept of machine learning for the applications that could not be programmed by hand such as database mining, handwriting recognition, computer vision and natural language processing. The definition of Machine Learning is, "Field of study that gives computers the ability to learn without being explicitly programmed" [Arthur Samuel (1995)]. Machine learning algorithms have been classified as supervised and unsupervised learning. Supervised Learners are the ones used for predictive modelling whereas Unsupervised Learners are used for descriptive modelling. . In supervised learning the class labels are known i.e. the "right answer is given". For example, consider Breast cancer with two discrete valued outputs (Malignant and Benign) or even in multivalued classification with multiple output types. Classification problem can be stated with single value as with only tumor size or with multi features such as age, tumor size and clump thickness. In Unsupervised learning we are not told what each data point is or in other words class label is not defined. The technique used to find the structure in data is called Clustering, which includes categorizing goggle news, DNA microarray data analysis, social network analysis etc. Different machine learning algorithms are categorized either as supervised or unsupervised learning techniques based on their working and uses.

Machine learning techniques have been used in medical domain from the very beginning for identifying disease patterns [1], prognosis of Colo Rectal Cancer (CRC) surgery [2], proteomic platter identification for detecting cancer [3], drug regime decision making [4] etc. The goal of supervised learning in medical domain is to construct models fed by data specific to patients for the prediction of outcomes of a particular interest and hence support clinical decision making. The paper presents selected machine learning techniques that can be applied for medical data, and in particular some supervised learning techniques which can be used to build predictive models for procedures such as diagnosis of neonatal diseases, identifying risk status based on prognosis information and monitoring neonatal health for planning treatment. Further a methodological review with a comparison of different supervised learning techniques used in medical applications and different model evaluation parameters will be discussed. Besides this it highlights the use of supervised learning techniques for neonatal data analysis with different approaches to improve model accuracy.

II. Supervised learning techniques

The various supervised learning approaches that are used for classification problems are discussed as follows along with their advantages and limitations. Classification techniques are categorized as eager learners and lazy learners. Eagar learners include decision tree, support vector machine, neural networks etc., which approximate the target function globally during training. Instant based learning methods are called lazy learners such as case based reasoning, KNN or K nearest neighbor approach etc., which take less training time and more predicting time.

A. K- Nearest Neighbor Algorithm

Nearest neighbor classifiers are well suited for the classification problem where relationships among features and the target classes are complicated, numerous and difficult to understand, yet the items of similar class type tend to be fairly homogeneous. The KNN algorithm begins with a training dataset made up of examples that are categorized by nominal variables. For every record in the test set, the algorithm will identify k records in the training data set that are "nearest" to each other or the most similar, k being an integer that will be specified previously. The unlabeled test example will be assigned the class of the majority of the k nearest neighbors. Deciding how many neighbors to use for KNN determines how well the model will generalize to future data. The balance between over fitting and under fitting the training data is a problem knows as bias-variance tradeoff. Choosing a large k reduces the variance caused by noisy data, but can bias the learner such as that it runs the risk of ignoring small, but important patterns. The best k value is somewhere between these two extremes. Theoretically it is set somewhere between 3 and 10. One common approach is to set the value of K equal to the square root of number of training observations. Second approach is to choose a larger K but with a weighted voting process in which vote of the close neighbors is considered more authoritative than the vote of faraway neighbors [5]. Although the k-NN algorithm is a simple, fast and effective algorithm, it has a slow classification phase and requires large amount of memory. It also requires additional processing for missing and nominal data.

B. *Decision Trees*

A Technique which divides data into smaller parts and uses these parts to predict outcomes by identifying patterns. The Decision Tree Induction is a greedy algorithm which constructs a tree in a top down manner by recursively using the divide and conquer approach. Selection of attribute is a major concern in decision trees which is taken care by attribute selection techniques like gini index, gain ratio, information gain and g-statistics. C4.5, C5.0 and ID3 are the various implementation of decision trees with C5.0 being opinionated about pruning as it takes care of many of the decisions about pruning automatically using fairly reasonable defaults. C5.0 uses an overall strategy of post pruning a tree by first allowing a tree to grow by over fitting training data and latter removing nodes and branches that have little effect on the classification errors. C5.0 algorithm can be improved by adding adaptive boosting in which many decision trees are built and tree vote is taken on the best class for each example. For classifying data in large databases in terms of scalability approach, Rain forest is the preferred choice which constructs an AVC –list consisting of Attribute, Value and Class label set. Rules can also be extracted from decision trees such that a single rule is created for every path from the tree's root node to a leaf node where a conjunction is formed for every attribute-value pair. The leaf node symbolizes the prediction of a class with rules that are both mutually exhaustive as well as exclusive. Sequential covering algorithms such as FOIL and RIPPER are also used to extract rules directly from training data. Decision tree models that are generated using simple CART or J48 are highly interpretable when compared to black box models making them readily accepted by the medicine community [6]. Decision Trees are all-purpose classifiers that do well on most problems and use the most important features. They can handle both numeric and nominal data and can be used on data with relatively few or very large number of training examples. They are more efficient than other complex models. They do deal with some problems like susceptibility to under fitting and over fitting [5].

C. *Bayesian Classification*

Bayesian classifiers utilize training data for calculation of the probability of every class based on the value of their features. When these classifiers are used on test data which are unlabeled, they use the same observed probabilities for the prediction of that class which is most likely for the new features. These classifiers use a very simple technique but are at par with the most sophisticated algorithms. In health care they can be used for diagnosing medical conditions for a given a set of observed symptoms. Bayesian classifiers are best applied to problems where numerous attributes' information is considered at the same time to estimate the outcome probability. While many algorithms ignore features that have weak effects, Bayesian methods utilize all available evidence to subtly change the predictions. They are based on a principle that if large number of features have relatively minor effects, taken together their combined impact could be quite large. *Naïve Bayes Classifier* is an application of the Bayes theorem where the attributes are conditionally independent which reduces the computation cost and only counts the distribution of the class. Naïve Bayes algorithm makes an assumption that all the features in the data set are both equally independent and important. This assumption is rarely true in most of the real world applications. Its major drawback being that it cannot model the dependencies. *Bayesian Belief Networks* is a directed acyclic model of causal influence relationships that represents the dependency among those variables with a joint probability distribution. Bayesian network is constructed based on subjective construction by identifying direct causal structure, synthesis from other specifications such as block diagrams and info flow and learning from data using maximum likelihood principle. Bayesian Belief networks define class conditional interdependencies between the subsets of variables. Naïve Bayes algorithm does well with missing and noisy data. Just like the k-NN algorithm it is simple and fast and is useful when you have very large number of features that are rare and/or nearly independent. It works well with both few and large examples for training. However, it relies on an assumption that is often faulty of equally independent and important features and might not be ideal for those data sets that have a lot of numeric features. One significant demerit of this algorithm may be that estimated probabilities are less reliable than the predicted classes [5].

D. *Artificial Neural Networks*

Neural Network algorithms were developed by neurobiologists and psychologists that were mimetic to the way neurons worked in a human brain. A neural network consists of input and output units that are connected and each connection joining the units has a weight assigned. The learning phase allows the neural network to learn by modifying weights of the connections to predict the correct label that belongs to the input examples. Between the input and output layers there may exist multiple hidden layers primarily used to boost performance but may be computationally expensive. The network performs nonlinear regression with the help of training examples and hidden units to closely approximate any mathematical function. Neural networks can perform the back propagation algorithm which iteratively processes a set of training examples and compares the output of the network with the actual target output. The error is propagated from the output unit(s) to the first hidden layer through all the hidden layers in between and weights of the corresponding connections are modified. The most widely used neural network is called the Multilayer Perceptron. Artificial Neural Networks have been known as "black boxes" that are difficult to interpret; however in recent years software provides measures of sensitivity indicating the importance of individual variables in the Artificial Neural Network model used for prediction of a given outcome [7].

ANNs are very accurate for classification modelling and make very few relationships about data's relationships. Neural Networks use hidden layers which significantly boost the performance. On the contrary, Neural Networks are really slow to train and are computationally very expensive in addition to being prone to under fitting and over fitting the data. The worst disadvantage about Neural Networks is that they can result in an extremely complex black box which can be difficult to interpret [5].

E. *Support Vector Machines*

SVM is a classification algorithm that deals with both linear and nonlinear data sets for classification and prediction problems. It incorporates nonlinear mapping that transforms the original data set to a higher dimension. In the newer dimension it searches for a decision boundary also called a linear optimal separation. With the help of this nonlinear mapping, the two classes on a graph can be distinguished using a hyper plane which is found by the algorithm using support vectors that are the essential training tuples and the margins defined by these support vectors. SVM uses different kinds of kernel functions for nonlinear classification like Polynomial, Gaussian, Radial Bias Function and Sigmoid. SVM can also be used for multiclass classification and regression problems [5].

SVM is not overly influenced by noisy data and is not prone to over fitting. It makes fewer assumptions about variable distribution than do many other methods. It is useful in cases where the useful interactions or other combinations of input variables aren't known in advance. However, there do exist some disadvantages about Support Vector Machines. They require testing of various combinations of kernels and model parameters and can be slow to train if dataset has huge number of features or examples. Support Vector Machines also result in a complex black box that can be difficult to interpret [5].

III. APPLICATIONS OF SUPERVISED LEARNING IN NEONATAL DOMAIN

This section covers the various applications of the supervised learning algorithms in the research work pertaining to the domain of neonatology. A Summary of Supervised Learning approaches in neonatal domain is listed in Table 1.

TABLE I
SUMMARY OF SUPERVISED LEARNING APPROACHES IN NEONATAL DOMAIN

Author	***Publication Year***	***Approach***	***Evaluation Criteria***
Linda Goodwin, Sean Maher [8]	2000	Neural Networks with Demographic variables only	0.64(AUC)
		Neural Networks with Demographic as well as other Indicator variables	0.66(AUC)
		Neural Networks with All variables	0.68(AUC)
		Stepwise logistic regression	0.66(AUC)
		CART	0.65(AUC)
		PPV Rule	0.67(AUC)
		FactMiner with Demographic variables only	0.72(AUC)
		FactMiner with Demographic as well as Indicator variables	0.76(AUC)
Mohammad Sohani et al. [9]	2006	Multilayer Perceptron	91% (Accuracy)
		Radial Basis Function	89% (Accuracy)
		Genetic Algorithm	92.5% (Accuracy)
		ANFIS	83% (Accuracy)
		Gaussian membership function plus Mamdani's fuzzy inference method as well as centroid defuzzification	85.5% (Accuracy)
Sung-Huai Hsieh et al. [10]	2009	Support Vector Machine and Adaptive Feature Selection	100% (sensitivity & specificity)
J.R. Williamson et al. [11]	2011	Equal-prior quadratic classifier	0.79(AUC)
Ferreira et al. [6]	2012	J48 Decision Trees	0.75(AUC)
		Simple CART	0.77(AUC)
		Naïve Bayes	0.88(AUC)
		Bayes Net	0.87(AUC)
		Multilayer Perceptron	0.81(AUC)
		Sequential Minimal Optimization	0.72(AUC)
		Simple Logistic	0.89(AUC)
Arthur Mikhno et al. [12]	2012	Logistic Regression	0.87(AUC)
James R. Williamson et al. [13]	2013	Gaussian Mixture Models	0.80(AUC)

Author	Publication Year	Approach	Evaluation Criteria
Mueller et al. [7]	2013	Artificial Neural Network	0.75(AUC)
		Bagged Decision Tree	0.51(AUC)
		Multivariate Logistic Regression	0.76(AUC)
		Naïve Bayes Classifier	0.63(AUC)
		Support Vector Machine	0.49(AUC)

A. Prediction of Neonatal Diseases

Machine learning has been widely used in research pertaining to neonates. Ferreira et al. [6] used techniques such as neural networks (multilayer perceptron), decision trees (J48), naïve base and simple logistic to predict the presence of hypebilirubinemia to improve neonatal Jaundice diagnosis. A total of 72 variables were collected such as parental, laboratory, demographic and gestational information etc. and only newborns that were healthy with gestation of or more than 35 weeks were included in the study. The levels of transcutaneous bilirubin (TcB) were measured from the birth of the newborn to it's hospital discharge with a maximum 8-hour time interval between these measurements. Supervised learning techniques were applied using internal cross validation 10-folds which was used to possibly know how the learning algorithm quality would be affected in separate sets of data. The results of the algorithms in terms of Area Under Curve (AUC) with both risk factors and (TcB) are summarized in Table 1. The higher accuracy of AUC (0.74) with only clinical risk factors were obtained with bayes net algorithm followed by naïve bayes and simple logistic. Multilayer perception (AUC 0.84) achieved higher accuracy with TcB values followed by naïve base and simple logistic. Finally, the combination of TcB levels at 24 hours and risk factors were tested with simple logistic (AUC 0.89), which achieved higher accuracy among all others. In this research, authors also mention about improving the results by using a tool other than WEKA or by using a bigger sample size. The primary finding of the paper demonstrated that data mining is a valid approach for neonatal hyperbilirubinemia prediction.

Sohani et al. [9] designed a model that would integrate clinical methods with a Neuro – Fuzzy system to help neonatal Jaundice diagnosis. An adaptive Neuro-Fuzzy inference system is used which maps the input space to the output space which can be done with a series of if-then statements. The goal of a fuzzy inference system is to formulate the mapping from input to output using fuzzy logic. Fuzzy logic takes linguistic information from human experts and combines with Artificial Neural Networks called Neuro-Fuzzy (ANFIS). The main motive of ANFIS is to have evolutionary learning similar to the back-propagation algorithm. Genetic Algorithm is also been used for intelligent searching and is based on the principals of biological systems evolution. Evolving Fuzzy Neural Network approach allows for rule insertion and extraction thereby restricting the problem from becoming obsolete and thereby allowing the learning system to learn data in real time. The author also proposed a method with a combination of Gaussian Membership Functions plus Mamdani's fuzzy inference method and centroid defuzzification approach. All the above mentioned methods with their accuracies have been mentioned in Table 1. The various methods mentioned here will help in building a software which will in turn help physicians for diagnosing neonatal Jaundice.

Mikhno et al. [12] developed a prediction algorithm using Logistic regression based on the data collected prior to first extubation attempt to distinguish between the neonates whose extubation was successful and those who had extubation failure. Correlation technique is used for feature selection to eliminate the unwanted features. Overtraining is seen as a problem with his approach as feature selection is not a part of the model building exercise. Precup et al. [14] proposed a Support vector machine method for predicting an optimal time for extubation that will allow clinicians to minimize the use of mechanical ventilation. Support vector machine with Gaussian Kernel being a nonlinear method was found to be better compared to Logistic regression as shown in Figure 1. Muller et al.[7] proposed a decision support tool with a set of algorithms such as Support Vector machine (SVM), Artificial neural networks (ANN), Naïve based classifiers (NBC), boosted decision trees (BDT) and multivariate logistic regression (MLR) for the prediction of optimal time of extubation in premature infants. ANN, MLR and NBC performed best whereas SVM and BDT over-fit the data as shown in Table 1. None of the methods can be considered as acceptable as the best AUC found was with MLR (0.76 AUC) which is below the minimal accepted value of (0.8 AUC). Batch effects found in datasets are responsible for failure in processing data when variable selection is used as an underlying method in algorithms. It is crucial to reduce the dimensionality of dataset prior to model development which would decrease the number of variables that need to be included in the model and may improve the performance of the individual methods. Other reason stated was that the outcome class was highly unbalanced in the data set which reduced the prediction ability of the methods used.

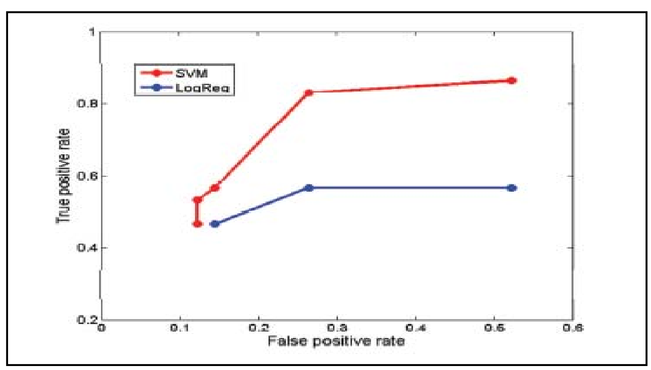

Fig. 1. ROC curves using SVM and Logistic regression [14]

Apnea of prematurity (AOP) is a common problem associated with preterm neonates. Williamsons et al. [11] used physiological signals extracted from interbreath (IBI) intervals and heartbeat (RRI) intervals to predict apnea in preterm neonates. A pre-apnea score which indicated the likelihood that the patient was in a pre-apnea versus a normal background state was used as a dependent variable. These scores were integrated over time to produce an apnea prediction score which was indicative of the probability a patient would experience an apnea during a defined window. A Machine Learning technique called the equal-prior classifier was used to learn the mappings from features of the training set into 2 classes of the patient's state: preapnea or interapnea. The equal-prior quadratic classifier was evaluated on each patient using 40-fold cross validation. The equal-prior quadratic classifier statistic is the 2-class log-likelihood ratio. The Area under the Curve (AUC) was found to be 0.79 with IBI and IRI which outperformed the result where the intervals were considered individually. A larger dataset with a nonlinear algorithm such as SVM could provide even more conclusive results. Williamson et al. [13] also proposed a real time algorithm for apnea prediction, based on cardio-respiratory and movement features that were extracted from multiple physiological sensors. Gaussian mixture models (GMMs) are employed in this paper which are the weighted combinations of multiple Gaussian densities. A separate background GMM was trained which was individualized to each patient. Bayesian adaptation was used to form both an interapnea and a preapnea GMM. The model was then evaluated using a 40-fold cross validation technique. Strong prediction accuracy was obtained, with a statistical significance found on 5 out of 6 patients. Temko et al. [15] presented a SVM classifier for patient – independent neonatal seizure detection to discriminate between seizure and non-seizure. The leave – one – out method being an almost unbiased estimation method is used to access the performance of the system for seizure detection. The work also focusses on the post – processing techniques applied on the classifier output. The post processing schemes applied on the output as shown in Figure 2 consist of: a) SVM output b) posterior probabilities using a sigmoid function. c) smoothed average filter d) binary decisions using 0.5 threshold values. e) single seizure binary decision f) OR fusion of (d) and (g) increase in duration of all positive decisions with collar operations. (h) ground truth – indicating one as seizure. To make the proposed system more flexible for clinicians the proposed system allowed the control by choosing different confidence levels. Further, Tempko et al. [16] proposed performance metrics to assess the performance of a seizure detector task which is grouped into event based and epoch based. Epoch based metrics are application irrelevant and for binary classification the decision is represented by a confusion matrix or a contingency table. The performance measures in epoch based metrics include Sensitivity, Specificity, ROC Curve, Precision, Recall and PR Curve. Event based metrics are application specific and the performance measures include Good Detection Rate, False detection per hour (FD/h), Curve of variation of GDR with FD/h and Mean false detection duration (MFDD). Good Detection Rate is defined as the percentage of events that are classified correctly by a system with the help of an expert. It is suited in the medical domain such as neonatology where expert opinion matters in complex predictions of seizures. False Detection Rate per hour is defined as the number of predicted events in 1 hour that have no overlap with actual reference events. The Mean False Detection Duration is computed by calculating the average duration of all the detections that are false produced by the system at a single operating point.

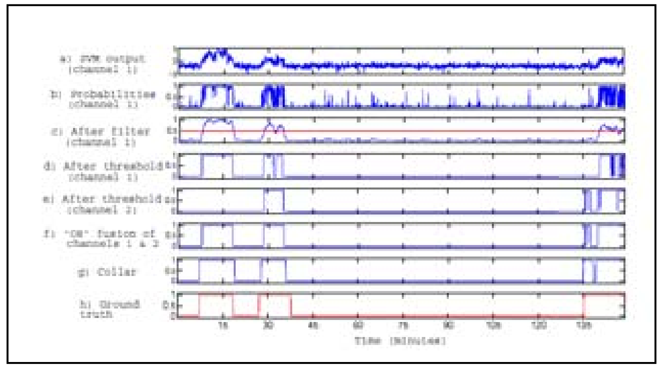

Fig. 2. Post Processing schemes [15]

Fergus et al. [17] worked on detection of whole-brain seizures using machine learning. They used EEG classification for the generalization of detection across all the regions of the brain. 342 EEG segments were extracted from 24 EEG records of patients who were suffering from epilepsy. The study focused on discriminating between seizure and non-seizure EEGs. 686 scalp recordings were taken from 23 patients. Feature selection methods such as statistical significance, principal component analysis, linear discriminant analysis(LDA), LDA forward search, LDA backward search, independent search and gram-schmidt analysis were applied and the top 20 uncorrelated features were extracted from all the regions of the scalp readings. For each of these methods, AUC was calculated and these methods were evaluated. Linear discriminant analysis backward search method proved to be the best with AUC = 91%. Top five uncorrelated features were also extracted from each of the 5 regions covered by the EEG scalp electrodes to allow the classifier to detect the focal seizures in different parts of the brain. Classification algorithms such as Support Vector Machines, Decision Trees, Parzen Classifier, K-Nearest Neighbor, Linear discriminant classifier, Quadratic discriminant classifier, uncorrelated normal density based classifier, polynomial classifier and logistic classifier were applied to the data. Techniques like Holdout Cross-Validation where training data was 80%, ROC curve, Sensitivity and Specificity were used as validation methods.

B. *Screening and monitoring neonates with risk stratification*

Identifying neonates at risk is a major concern in Neonatal Intensive Care Unit. Support Vector Machine has been proposed to determine whether the neonate has some kind of metabolic disorder disease [10]. The data was preprocessed with normalization and Pearson's Correlation Coefficient was used for feature selection. SVM with Radial Based Function (RBF) kernel was used for training the data. Grid search was applied to find the best pair of (Penalty, Gamma) so that the classifier can most accurately predict the unknown data. K-fold validation technique was also applied and the experiment was repeated 5 times. The final accuracy of this proposed method along with feature selection and cross validation provided accuracy approaching 100% as listed in Table 1.

Zernikow et al. [18] used an artificial neural network approach for mortality risk prediction trained on admission data of preterm infants. ANN performs slightly better than Logistic regression with AUC 0.95 vs. AUC 0.92 as the nonlinear association was not detected by logistic regression. However, this approach could not predict individual mortality risk as the important factors such as physiological parameters that may contribute to preterm mortality have not been considered.

Aggarwal et al. [19] proposed a multiclass SVM classifier which classifies the clinical conditions of a neonate as risk classes. Directed acyclic graph SVM is used to perform binary classification comprising of one v/s one SVM by comparing only one class with all the other classes to cover all set of pairs and finally reaching to the end leaf as shown in Figure 3.

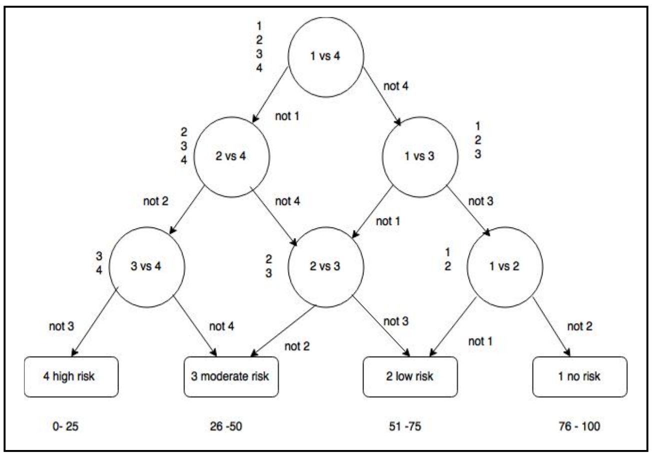

Fig. 3. Directed Acyclic Graph SVM approach for multi classification [19]

IV. DISCUSSION

The paper investigates different supervised learning techniques in health care with their strengths and weaknesses. The data in health care domain being highly imbalanced and unprocessed have to pre-processed with different techniques such as data exploration, data transformation, normalization, rescaling and feature selection. For handling class imbalance which is common in medical data sets certain methods include over sampling, under sampling, threshold moving and ensemble approach. Ensemble approach works with a principle of combining multiple learners to create a strong learner. It is built on the two considerations. First one is based on the allocation function dealing with either a homogeneous or heterogeneous classifier used for the construction of a model and the second one deals with the combination function which governs the disagreement among predictions which are reconciled using majority vote or weighing each model output with past experience. Ensemble approach can lead to diversity among predictors with a superior predictive capacity. The papers also describes different applications of supervised learning techniques in neonatal domain.

V. CONCLUSION

The brief review of literature with regard to use in neonatal disease diagnosis and monitoring of health with the methodological details and utility in terms of performance evaluation parameters has been specified in the paper. The various classifier evaluation parameters mentioned in literature are accuracy, specificity, sensitivity and ROC curves. The best performance metric to evaluate any algorithm is a ROC curve that calculates the AUC or the Area under the curve which is plotted between sensitivity and (1- specificity). It takes into account both the sensitivity and specificity as compared to general accuracy which is better in cases where the output classes are imbalanced especially in medical examples. Lastly, there is no one supervised learning classification algorithm that is suitable for every single data set. Every supervised technique has its own merits therefore the best way is to analyze and preprocess the data with the preprocessing techniques and then build the model with the most suitable techniques. Based on the available literature it can be concluded that Neural Networks, Support Vector Machines and Decision Trees have been considered really efficient in the neonatal domain.

ACKNOWLEDGMENT

Authors are deeply indebted to Manipal Institute of Technology and Manipal University for providing an opportunity to demonstrate the research work.

REFERENCES

[1] Ilayaraja and T. Meyyappan. Mining medical data to identify frequent diseases using apriori algorithm. In Proceedings of International Conference on Pattern Recognition, Informatics and Medical Engineering (PRIME), Salem, Februaryr 2013. IEEE.

[2] T. Oliveira, E. Barbosa, S. Martins, A. Goulart, J. Neves and P. Novais. A Prognosis System for Colorectal Cancer. IEEE2013.

[3] Sheng-Yong, Qu Huang, Lin-Li, C. Qi-ZhiTeng, Chang-Ying Ma, Hui Zhang, RuBai, Qi- ZhiTeng, Ming-Li Xiang, and Yu-Quan Wei. An integrated scheme for feature selection and parameter setting in the support vector machine modeling and its application to the prediction of pharmacokinetic properties of drugs. Artificial Intelligence in Medicine, 46(2):153–163, 2009.

[4] Hu, Paul Jen-Hwa, Chih-Ping Wei, Tsang-Hsiang Cheng, and Jian-Xun Chen. Predicting adequacy of vancomycin regimens: A learning-based classification approach to improving clinical decision making. Elsevier, Decision Support Systems, 43:1226–1241, 2007.

[5] Brett Lantz, Machine Learning with R, Packt Publishing, February 2014.

[6] D. Ferreira, A. Oliveira, and A. Freitas. Applying data mining techniques to improve diagnosis in neonatal jaundice. Springer, BMC Medical Informatics and Decision Making, 12(143), December 2012.

[7] M. Mueller, J. Almeida, R. Stanislaus, and C. Wagner. Can machine learning methods predict extubation outcome in premature infants as well as clinicians. PubMed, J Neonatal Biol, 2(2), June 2012.

[8] Linda Goodwin, Sean Maher. Data Mining for Preterm Birth Prediction. Proceedings of the 2000 ACM symposium on Applied computing – Volume 1. Pages 46-51, 2000.

[9] M. Sohani, B. Makki, N. Sadati, K. K. Kermani and A. Riazati, "A Neuro-Fuzzy Approach to Diagnosis of Neonatal Jaundice," *Bio-Inspired Models of Network, Information and Computing Systems, 2006.*

1st, Madonna di Campiglio, 2006, pp.1-4. doi: 10.1109/BIMNICS.2006.361808.

[10] Hsieh, S. H., Chien, Y. H., Shen, C. P., Chen, W. H., Chen, P. H., Hsieh, S. L., ... & Lai, F. (2009, June). Newborn Screening System Based on Adaptive Feature Selection and Support Vector Machines. In *Bioinformatics and BioEngineering, 2009. BIBE'09. Ninth IEEE International Conference on* (pp. 344-347). IEEE.

[11] J.R. Williamson, D.W. Bliss, D.W. Browne, P. Indic, E. Bloch-Salisbury, and D. Paydarfar. Using Physiological Signals to Predict Apnea in Preterm Infants. Conference Record - Asilomar Conference on Signals, Systems and Computers, (pp. 1098-1102). [6190183] 10.1109/ACSSC.2011.6190183.

[12] Mikhno A, Ennett CM. Prediction of extubation failure for neonates with respiratory distress syndrome using the MIMIC-II clinical database. InEngineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE 2012 Aug 28 (pp. 5094-5097). IEEE.

[13] J. R. Williamson, D. W. Bliss, D. W. Browne, P. Indic, E. Bloch-Salisbury and D. Paydarfar, "Individualized apnea prediction in preterm infants using cardio-respiratory and movement signals," *Body Sensor Networks (BSN), 2013 IEEE International Conference on*, Cambridge, MA, USA, 2013, pp. 1-6. doi: 10.1109/BSN.2013.6575523.

[14] D. Precup, Carlos, A. Robles, Karen, A. Brown, L. Kanbear, J. Kaczmarek, S. Chawla, Robert, and E. Kearney. Prediction of extubation readiness in extreme preterm infants based on measures of cardiorespiratory variability. In Proceedings of Thirty fourth Annual International Conference on Engineering in Medicine and Biology Society (EMBC), pages 5630–5633, San Diego, California USA, August-September 2012. IEEE.

[15] Temko, A., Thomas, E., Marnane, W., Lightbody, G., & Boylan, G. (2011). EEG-based neonatal seizure detection with support vector machines.*Clinical Neurophysiology*, *122*(3), 464-473.

[16] Temko, A., Thomas, E., Marnane, W., Lightbody, G., & Boylan, G. B. (2011). Performance assessment for EEG-based neonatal seizure detectors. *Clinical Neurophysiology*, *122*(3), 474-482.

[17] Fergus, P., Hussain, A., Hignett, D., Al-Jumeily, D., Abdel-Aziz, K., & Hamdan, H. (2016). A machine learning system for automated whole-brain seizure detection. *Applied Computing and Informatics*, *12*(1), 70-89.

[18] B. Zernikow, K. Holtmannspoetter, E. Michel, W. Pielemeier, F. Hornschuh, A. Westermann, and K. H. Hennecke. Artificial neural network for risk assessment in preterm neonates. PubMed, Archives of Disease in Childhood-Fetal and Neonatal Edition, 79(2):F129–F134, 1998.

[19] S. Agarwal and G. N. Pandey. Human computer interface design for neonatal intensive care with data mining. In Proceedings of fourth IEEE Conference on Intelligent Human Computer Interaction (IHCI), pages 5630–5633, Kharagpur, India, December 2012. IEEE.