ORIGINAL ARTICLE

Estimation of Aminophylline effectiveness at Neonatal Intensive Care Unit (NICU) using Artificial Intelligence

Rudresh Deepak Shirwaikar[1], Dinesh Acharya U[2], Krishnamurthy Makkithaya[2], Surulivelrajan M[3]

[1] Department of Information Science and Engineering, BMS Institute Of Technology And Management, Bangalore, Karnataka, India.
[2] Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka, India.
[3] Department of Pharmacy Practice, Manipal College of Pharmaceutical Sciences, Manipal Academy of Higher Education, Manipal, Karnataka, India

ABSTRACT

Introduction: The estimation of drug competence using Artificial Intelligence is presented in various literature for the adult population, but it is still new for drug dosage optimization in neonates. Aminophylline, a methylxanthine is administered as central nervous system stimulant for reducing Apnea episodes in neonates. **Methods:** The paper describes comparative evaluation of Support Vector Machine (SVM), K Nearest Neighbour (KNN), Decision Tree (DT) and Artificial Neural Network (ANN) for predicting drug effectiveness of Aminophylline. The models were evaluated using 100 Aminophylline cases based on various metrics such as sensitivity, specificity, and accuracy. The data used for the analysis was collected from the population pharmacokinetic study conducted at Kasturba Medical College, Neonatal Intensive Care Unit (NICU). **Results:** The evaluation result seemed to favour Multi-Layer Perceptron (MLP) with accuracy of 0.92 Area Under the Curve (AUC) followed by 0.85 (AUC) for Support Vector Machine (SVM). The input parameters in particular maternal, pharmacokinetics, demographic and physiological that were identified in literature as predictor variable played an important role in estimating effectiveness of Aminophylline regimens. **Conclusion:** Artificial Intelligence approach was potentially helpful in analysing drug dosage of Aminophylline and its effectiveness in diagnosing neonatal Apnea.

Keywords: Apnea, Aminophylline, Drug dosage analysis, Artificial Intelligence, Multi-Layer Perceptron

Corresponding Author:
Rudresh Deepak Shirwaikar, PhD
Email: rudresh@bmsit.in; rudreshshirwaikar@gmail.com
Tel: +91 9823014550

INTRODUCTION

Apnea of prematurity a breathing disorder which is a serious concern and a major threat to infant health (1). The repeated apnea episodes accompanied with hypoxemia and bradycardia needed to be identified and treated at the Neonatal Intensive Care Unit (NICU) (2, 3). Specific treatment measures included methylxanthines drug dosage, continuous positive airway pressure, invasive mechanical ventilation, prone positioning, bag mask ventilation and kinesthetic stimulation. Mechanical ventilation for neonatal respiratory distress is the most frequently used intervention, which helped preterm neonates with immature lungs to be able to revive with demands of extra-uterine life (4). Extended ventilation had severe side effects and Caffeine or Aminophylline (methylxanthines) played a clear pharmacological role in detaching the neonate away from ventilation, as well as in improving them form developing recurrent apnea episodes (5, 6). Aminophylline had narrow therapeutic range compared to Caffeine. The appropriate therapeutic range is 10 to 20 mg/L, and less than 10 mg/L (sub therapeutic dose) and greater than 20 mg/L (the supra therapeutic dose). Both the drugs are widely used for the cure of Apnea of prematurity. Based on literature (5), the Aminophylline was found to be as effective as Caffeine in reducing apnea episodes and also cost effective. However, Caffeine has less adverse reactions compared to Aminophylline. In India tachycardia is common with Aminophylline treated infants (5). Therefore, dosage optimization of Aminophylline using careful and efficient technique is required in developing countries like India.

Pharmacokinetics (PK) defines the drug absorption, distribution, metabolism, and elimination in the body with the time course of a specific drug, the major metabolite concentration in plasma and other biological fluids (7). PK when performed in controlled environment is the study of what happens to the drug and its effectiveness when it goes into the body. It is moreover used for drug research such as dose discovery and escalation performed on healthy volunteers in an explained or known environment. The classical pharmacokinetics method for dosage analysis may be unwise or unethical in the infant population. A data analysis method with data taken from groups of individuals or patient populations is called Population pharmacokinetics (8). In Population PK analysis the pharmacokinetics parameters used or identified are derived from the actual patients and not volunteers. The one-compartment model such as the Nonlinear Mixed Effect Modelling is considered the appropriate method for data analysis for each drug (9-12). The advantages of the population pharmacokinetics over the simple pharmacokinetics are: a) analysis done on sparse data with statistical methods; b) the range of blood sampling is less; c) it involves actual patient population.

The term Artificial Intelligence (AI) was published at The Dartmouth conference in 1956. Artificial Intelligence empower machines to imitate human behavior. Supervised and unsupervised learning are the two-basic grouping of Artificial Intelligence algorithms. The predictive models with labelled data set are build using supervised learner algorithms, and knowledge generation or descriptive models are build using unsupervised algorithms. Artificial Intelligence has been used in health care in different areas such as hospitalization prediction as a supervised problem (13), risk assessment for cancer prognosis after surgery (14), frequent diseases mining using Apriori algorithm (15), predicting survivability in breast cancer patients (16), neonatal domain for diseases progression prediction and its identification (17-24) and drug dosage prediction (25-27). Unsupervised learning has been used in different applications of health care such as use of clustering for extraction of medical rules in medical databases (28), risk stratification in neonates using clustering techniques and k means clustering to find the likelihood of occurrence of diseases (29).

Classification techniques have been grouped as lazy and eager learner classifiers. Methods such as Decision Tree induction, Support Vector Machine (SVM), Artificial Neural Network (ANN) are eager learning methods with globally approximated target function during training. Lazy learners are instant based learning methods which take less time in training and more time for predicting e.g., K Nearest Neighbours (KNN) and Case based Reasoning. Eager learning methods handle noisy and non-linear data better than lazy learners. Therefore, in most of health care applications where data is noisy and non-linear Decision Tree (DT) induction, SVM and ANN are preferred over others.

Drug dosage analysis of Aminophylline had been carried out using the population pharmacokinetic approach. This approach entailed considerable cost, resource, time and ethical issues. There is no reported study on the use of Artificial Intelligence algorithms on the drug like Aminophylline, a cost-effective medication that can be intensively used in neonates if dosage is optimized properly. The literatures (25-27) demonstrated that supervised algorithms had been fruitfully used for drug dosage analysis and prediction, especially for drugs with a limited therapeutic range and with notable side effects. Also, the algorithm accuracy of the expert systems developed using supervised models was greater than or equivalent to the standard pharmacokinetics models. Therefore, there was the need for a comparative study of Artificial Intelligence techniques such as ANN, DT and SVM to predict the drug adequacy of Aminophylline. It was also noted from literature that in a clinical perspective for determining whether the patient had received adequate prescription, a supervised classification approach to predicting the acceptability of a drug regimen (i.e., effective / ineffective) is more advisable than identifying the peak and trough concentrations of the drug. Proper dosage of Aminophylline was successful in lowering apnea within 2 to 7 days of treatment. Therefore, there was a scope or need of Artificial Intelligence techniques to develop optimal model for dosage optimization of Aminophylline.

MATERIALS AND METHODS

The paper presents a retrospective study on preterm neonates who are administered with Aminophylline dose for the reduction of apnea episodes. High Performance Liquid Chromatography (HPLC) method was used for drug concentration analysis by using a population pharmacokinetics study conducted at the study site Kasturba Medical College, Manipal, Karnataka, India. The required ethical permission was obtained from the Manipal Academy Higher Education ethics committee. The drug concentration data accumulated from the HPLC method was used to define the outcome (effectiveness) for our study. In our study the Aminophylline dosages were analyzed for effectiveness on preborn babies using Artificial Intelligence techniques. The outcome decision classes were defined as (effective (E) vs. ineffective (IE)).

Class E: The neonate had no other associated risk factor and was free from apnea episodes after undergoing seven days of treatment with obtained drug concentration level within the therapeutic range from 10 to 20 mg/L.
Class IE: a) The obtained drug concentration level is not within the therapeutic range from 10 to 20 mg/L. b) The obtained drug concentration level is within the therapeutic range with persistence of apnea episodes even after seven days of treatment. c) For a neonate when the derived drug concentration was not within the therapeutic interval (either < 10 mg/L or > 20 mg/L), apnea continued even after seven days of treatment associated with no other risk factors;

The Figure 1 describes the proposed methodology adopted to build optimal Artificial Intelligence models. It consists of data preparation (collection, exploration and

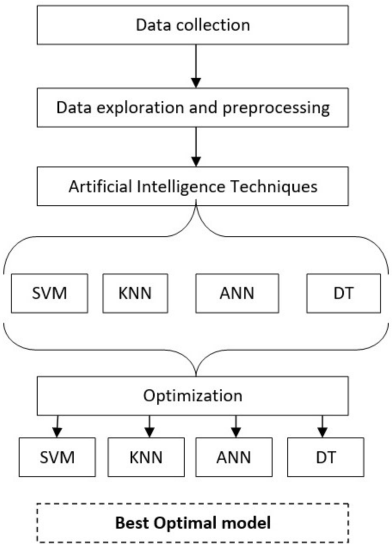

Figure 1: Proposed methodology to build the optimal prediction model

pre-processing), and the identified techniques to build Artificial Intelligence based decision support system for estimating Aminophylline effectiveness.

Data preparation

The data used for the work is a component of Aminophylline Population Pharmacokinetic study performed on target population at the study site. A retrospective study based on previous research using population PK approach at the study site was designed as a data collection tool. High Performance Liquid Chromatography (HPLC) method was used to generate concentration of drug Aminophylline for analysis. The data used for the study was collected from the medical records from the study site. The data collected from the study site over 2 years consists of 115 preterm neonates diagnosed with apnea administered with Aminophylline dosage. With reference to prior literature and consultation with concerned medical experts the predictor variables identified were: a) heart rate, respiration rate, desaturation and other physiological parameters.; b) demographics such as cry at birth, gestation age, weight at birth and apgar score.; c) delivery mode, steroids surfactant given to mother at the time of delivery and any other drugs with side effects.; d) PK variables such as concentration of drug, infusion rate, time when blood is pricked, time of dosage, dose at loading or start, and maintenance dose. The effectiveness of drug or ineffectiveness of drug was defined with binary code of 0 or 1 respectively. The 40 input features identified from the literature and medical consultation were further optimized with a statistical relevance method known as Forest of trees (30). Gini index (31) as a measuring parameter was used by Forest of trees method to find most accurate 22 features with reducing feature importance.

Principal Component Analysis (PCA) is a device to understand multi-dimensional distributions in lower dimensions. It identifies the eigen vectors of distribution along the maximal variance axis and used them to discover different uncorrelated independent variables / features called the principal components with the greatest corresponding eigen values. The clarity and usefulness of high dimensional data, into lower dimensional feature space is provided using principal components of PCA. In non linear distributions PCA was found to be ineffective as it is unable to detach the data points reduced from a complex multivariate K-dimensional feature space to k dimension lower dimension representation (32). An extension of linear PCA algorithm is Kernel principal component analysis capable of inferring a non-linear representation using kernels and addresses the sparsity problems that most of health care datasets encounter (32). In this work a linearly correlated data with reduced noise was noted with Kernel PCA representation of the data as compared uncorrelated noisy linear PCA representation of the data. Further, based on results it was inferred that data was nonlinear in nature and noisy, as was with most health care datasets. Therefore, it may be classifiable reliably on hyper-plane using a high-dimensional non-linear classifier. The data Kernel Principal Component Analysis (PCA) was used to visualize and understand the linearity of the data.

The statistics methods such as mean, median, variance was used to describe the data. The results showed the data to be noisy with various features having missing values. The presence of inherent noise in the data proved the methods such as mean, mode and nearest neighbor inconsistent and unreliable techniques to be used. Therefore, few of the observations with missing features values were not considered from the dataset. This ensured that approximation error, which can lead to additional noise is not introduced in the model. Further some of the continuous value features having missing data and medically useful were changed to new category (discrete) with "not know" as an additional group. All the human or technical errors feature values while collection of data was grouped into a new group called as "not know". After the preprocessing, the data sized reduced to 100 from initial 115 patient's data. Analyzing the data further it was found that data is skewed with 44:66 ratio where 44 observations had class label as effective (E) and the other 66 had class label as ineffective (IE). Hence it is interpreted as 44 observation had a positive outcome and remaining 66 had an outcome as 1, as negative or ineffective. The preprocessed data was given as input to train the model with 70% as training set and 30% was used to validate or test the model.

Artificial Intelligence algorithms

Algorithms such as Support Vector Machine (SVM), Decision Trees (DT), K nearest neighbor (KNN) and Artificial Neural network (ANN) were used to build classification models to predict aminophylline efficacy. Furthermore, all the models were trained and optimized based on the steps defined from the literature (32-38). The various assessment evaluation parameters were used to compare the results obtained from all the models for estimation of Aminophylline regimens.

SVM is a classification and prediction method for both linear and non-linear data (33). Support vectors define the data points that defines margin. When the number of features is large, support vectors provide a compact way to build classification model. The nonlinear classification is performed with the use of kernel functions such as radial bias, sigmoid, gaussian, and polynomial. Kernel functions are used to evade over fitting when the training data contains noise and to map the data into higher dimensional space. In our study linear and non-linear kernel (radial bias), SVM was trained as described in Figure 2. The dataset contained 22 input features and one output predictor. The kernel function sigma (ξ) and the margin factor (C) were used as optimization parameters for maximum performance as shown in Eqns. (1) and (2). The best combination is found using a 10-fold cross validation technique.

$$\min \frac{1}{2} \| \hat{w} \|^2 + C\sum_{i=1}^{n} \xi_i \qquad (1)$$

$$y_i(\hat{w}.\hat{x}_i + b) \geq 1 - \xi_i, \forall \; \hat{x}_i, \xi_i \geq 0 \qquad (2)$$

Where, (C) represents a cost value, $\hat{w}$ is a normal vector decision hyper plane, $\ddot{x}_i$ data points denoted by $\ddot{x}_i$ and a constant b. Minimizing the error function denoted by Eq. (1) will give us the optimal solution.

As shown in the Figure 2 the default cost function C=1 was used to build the linear SVM classifier, evaluated

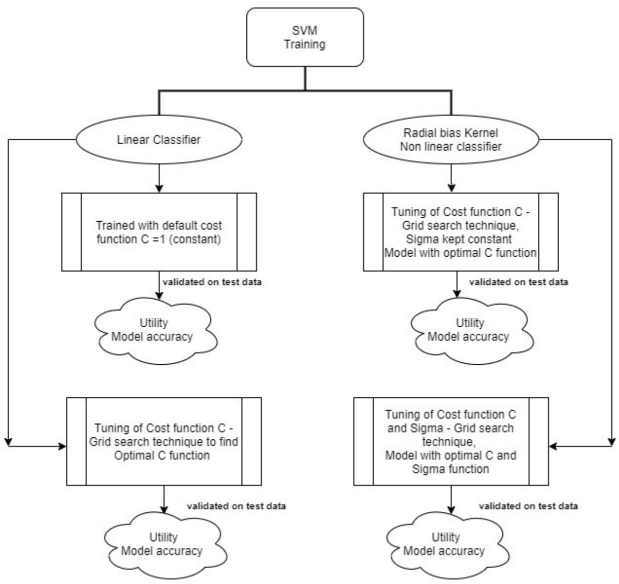

Figure 2: Steps to build optimal SVM model

on a test dataset. Further, a grid search technique was used for tuning different values of the cost function C. The best model optimal cost function C was selected based on Kappa metric with the highest accuracy. The accuracy of SVM linear (C=1) and optimized (C) SVM was recorded on the test data. In parallel, SVM radial bias kernel function classifier was trained. Different types of kernels were tried but the performance of the model had no impact on the choice of the kernel function. Therefore, with SVM radial bias kernel, the sigma parameter in Eq. (1) was kept constant and the optimized tuned cost (C) value, the highest accuracy was used. Based on the final optimized values of the sigma and cost, the model was checked for accuracy on a test data set. Further, based on grid search technique the most optimal combination of sigma and cost function was noted. The kappa value with the highest accuracy was used to identify the optimized tuning parameters and the model accuracy was noted with the test data set (20 %). The Figure 2 explained the optimizing of C and sigma values. The maximum accuracy was 0.89 and when C = 2 and sigma = 0.028 and for the model SVM for Aminophylline effectiveness is described in Figure 3. The model is built with the obtained tuned sigma and C values.

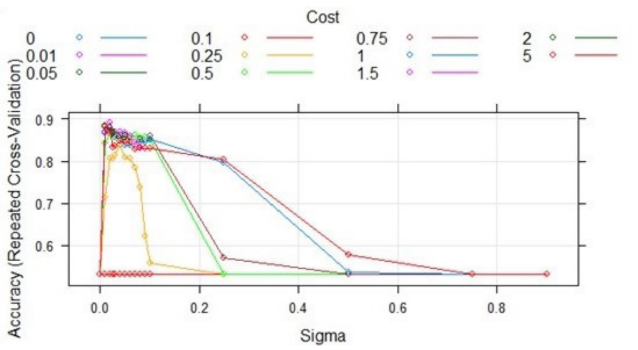

Figure 3: Process of tuning sigma and C parameters

ANN is a black box versatile learner which can be used for any learning task of classification or prediction (34). The neural network is defined with the choice of activation functions, network architecture and training algorithm. The choice of activation function is decided based on type of data. The most possible activation functions are Rectified Linear Unit (ReLU), Sigmoid, Softmax and Tanh. Non-linear activation functions help to bring non-linearity in the system. Network architecture consists of number of input nodes, number of hidden layers and output layers. Neuron in hidden layer depends on the number of input nodes, training data size, noise in data and complexity of learning task. The back-propagation algorithm is used to train an ANN with a strategy to back propagate error which in turn modifies the assigned random weights to reduce the total error of the network (35). The gradient descent decides how much the error will be reduced or increased for a change in weight. The idea here is to have balance between the variance and the bias in the model. Multi-Layer Perceptron (MLP) is type of neural network which uses feed-forward mechanisms consisting of one or more hidden layers, with non-linear or linear activated neuron as its core unit (34, 35). In the work, the MLP network is optimized for improving the performance of network by heuristically selecting, fine-tuning various parameters and hyper parameters of the network based on the steps mentioned below and as described in literature (36).

Step 1: A basic MLP is built with one hidden layer trained with a learning rate of 0.01 and 0.9 as momentum with 5,000 epochs, cross validated at every twenty-five epochs. The various activation functions such as Sigmoid, Tanh, Hard Tanh, Rectified Linear Unit (ReLu) and leaky ReLu were tried. The evaluation performance parameter Area Under the Curve (AUC) was used to find the optimal model with the best activation functions. Based on result the MLP with Rectified Linear Unit (ReLu) was found to the most accurate model which could be credited to the inherent lower bias and the reduced computational cost of the model belonging to linear activation approximation.

Step 2: Continuing with the optimal activation function the model is further trained to identify the best performing optimization algorithm on the data set. The Stochastic Gradient Descent (SGD) out performed Linear Gradient Descent (LGD) and Conjugate Gradient Descent (CGD). The performance increase seen with SGD could be due to gradual decrease in error rate based on the concept of the iterative averaging approach of the gradient in batches.

Step 3: Continuing with the optimal activation function and Gradient descent algorithm the model is trained to find the optimal layers of the network. The maximum performance was noted when the number of layers increased up to 4, with a decrease in the performance as we added more layers as variance increases.

Step 4: Continuing with the optimal activation function, Gradient descent algorithm and size of network the model is trained to find the effect of regularization functions such as L1 and L2 on the network. The bias variance trade off problem in MLP could be minimised with the use of regularization functions. However, on the dataset in our problem, varying L1 and L2 did not show any positive results.

Step 5: The final step dealt with choice of the Updater algorithms known as Adam, RMS Prop, SGD, Nestero Momentum, Momentum, and Ada Delta. This were used to improve the network performance and the optimized learning of the network and on the dataset. The Ada Delta updater algorithm was found to be most performing with the previous optimized model characteristics.

The MLP was built with the found optimal parameters mentioned in step 1 to step 5; the input dataset was normalized using zero mean and unit variance.

In KNN algorithm the new cases are classified based on similarity measures. The Euclidean and Manhattam distance functions were used to compute similarity (37). The K in KNN represents neighbors and number of K tells the model performance on future values. The sample data set was validated with various number of K values to overcome bias variance trade off problem, associated with over fitting and under fitting. The normalization techniques such as min max technique was used to build KNN model for predicting apnea. The K values used varied from 1 to 10 which was used as a technique to validate and optimize KNN. The model built with different k values with highest accuracy and less error prone was used as the final model to predict neonatal apnea.

Random forest (RF) is a technique which adds distinction with Decision Tree by combining random variable selection procedure with principles of bagging. It is also known as Decision Tree Forest tree classifier. The data set which has missing elements or noisy content with both continuous or categorical variables can be accurately handled with Random Forest (RF) classifier (37, 38). In the work the RF classifier with cross validated dataset, the optimization technique for setting of tuning grid was optimized with the concept of using randomly selected feature at each split as the square root of the total number of input variables.

RESULTS

The section deals with the results obtained from our study where all the developed Artificial Intelligence models where evaluated using various parameters such as accuracy, sensitivity, specificity and compared with Area Under the Curve (AUC). The Random Forest (RF) model was built with various composition of randomly sampled input features (mtry). At the start the square root of number of input features was considered and compared with other values (2, 5, 10, 22). Further, the DT model was also built and the results were compared with the RF models on the study population. The results obtained from DT was found to be low (0.56 AUC). The most accurate RF model was with the highest accuracy (mtry = 22), with 0.80 AUC score. This proves the importance of ensembles and Random Forest (RF) over Decision Tree (DT) based on the results obtained on study population. Among all KNNs with AUC = 0.70, the model with k = 10, nearest neighbor was found to be the most accurate.

SVM was trained with various optimizations such as SVM linear kernel, SVM linear kernel tuned with (c= 0.75), kappa = 0.88. Further, SVM nonlinear kernel (Radial bias) with constant sigma and tuned with (c=0.5), kappa =0.74, followed with tuning of sigma and C (sigma =

0.028 and c= 2). The results obtained from the Support Vector Machine for aminophylline effectiveness are mentioned in the Table I. The KNN model was trained with some of the optimal k values as described in Table II. Compared to all the K values (3, 5, 7 and 10), the model built with 10 nearest neighbors was found to be accurate with AUC of 0.70.

Table I: SVM results for drug effectiveness for Aminophylline

Algorithms	Sensitivity	Specificity	Accuracy	AUC
SVM Linear	0.69	0.68	0.70	0.69
SVM Linear optimized	0.71	0.68	0.74	0.70
SVM nonlinear	0.83	0.67	0.80	0.79
SVM nonlinear optimized	0.87	0.70	0.89	0.85

Table II: KNN results for drug effectivess for Aminophylline

Algorithms	Sensitivity	Specificity	Accuracy	AUC
KNN with k = 3	0.50	0.45	0.40	0.46
KNN with k = 5	0.68	0.58	0.51	0.50
KNN with k = 7	0.70	0.63	0.62	0.60
KNN with k= 10	0.76	0.65	0.75	0.70

All the models were trained with optimization tuning methods and the optimal model with highest accuracy on individual algorithms is presented in the Table III. The evaluation results seemed to favor Multi-Layer Perceptron (MLP) with rectified linear unit as an activation function, 4 hidden layers and optimization algorithms as Stochastic Gradient Decent, for drug effectiveness as decision class with 0.92 AUC followed by non-linear kernel-based Support Vector Machine with sigma constant at 0.028 and tuned C=2. Furthermore, all the developed machine models were able to predict positive class (effectiveness of Aminophylline) better than ineffectiveness demonstrated by greater sensitivity and lower specificity.

Table III: Comparative evaluation results for drug effectives for Aminophylline

Algorithms	Optimization	Sensitivity	Specificity	Accuracy	AUC
SVM	SVM non-linear kernel with sigma constant at 0.028 and tuned C=2	0.87	0.70	0.89	0.85
KNN	K = 10	0.76	0.65	0.75	0.70
ANN	MLP with 4 hidden layers, Stochastic Gradient Decent, ReLU	0.98	0.80	0.93	0.92
DT	Random Forest with(trials = 24)	0.80	0.77	0.82	0.80

Note: Support Vector Machine (SVM), K nearest neighbor (KNN), Artificial Neural Network (ANN), Decision Tree (DT), Multi-Layer Preceptron (MLP).

DISCUSSION

Overall, our evaluation results for Aminophylline suggested the following: a) a classification model (classes) derived from literature and neonatologist experience for predicting effectiveness of Aminophylline was more useful than estimating drug concentration or dosage. b) the input parameters such as physiological, pharmacokinetics, demographic and maternal that were identified in literature as predictor input features played an influential role in evaluating and estimating Aminophylline regime. c) the effectiveness of optimizing Artificial Intelligence models was noticed for all the classifier models with improved performance along optimized tuning. d) MLP was proved to be more effective in handling the single centric small data set which was nonlinear and unbalanced compared to other models in the study.

Previous literatures (5) compares Aminophylline effectiveness over Caffeine for prevention of apneic spells in premature neonates. Caffeine could be the preferred drug in many NICU due to its less side effects, however more studies on aminophylline effectiveness with reduced side effects could change the usage of Caffeine for Aminophylline as it is cheaper than Caffeine. The paper deals with the development of Artificial Intelligence framework to predict the dosage, and effectiveness of the drugs aminophylline. The model should be able to optimize the administration of life saving drugs to patients who are very sensitive to dosage. It will reduce the cost, time and ethical considerations which are the drawbacks when employing a pharmacokinetic model. The proposed drug dosage predictor model should be useful for designing molecular diagnostic/healthcare framework to tailor treatment of individual patients. To train any predictive model the basic requirement is availability of timely, accurate and reliable data. If we manage to collect huge amount of data and train the model to reach a certain maturity level, then we can propose to integrate this drug predictor model into the healthcare framework to tailor treatment of individual patients. The future of drug discovery and optimization depends on population approach for data analysis and Artificial Intelligence techniques will be potentially useful for optimizing drug dosage and also for drug discovery.

CONCLUSION

Effective management of drug dosage is need of the hour for infant patients being admitted and cared in a NICU. Aminophylline therapeutic regimens helped avert the development of continuous apneas episodes and reduced the continuation of artificial mechanical ventilation. The Artificial Intelligence classification algorithms were used to administer the clinical use of Aminophylline, the methylxanthines. Also, to predict aminophylline efficacy for its effectiveness in the study population. The analysis favored the use of ANN architectures such as MLP with optimized 4 hidden layers, Stochastic Gradient Decent, ReLU to improve the performance of Artificial Intelligence model (decision support system) built with non-linear, unbalanced and lesser data sets. To generalize the results further research was needed. A multi-center study with a greater number of cases was indicated. Further with more data advanced deep leaning methods could be explored. Classification algorithms were potentially useful and supportive for a neonatologist to optimize Aminophylline regime. Based on results reported herein it could be concluded that such techniques would help neonatologist in analyzing drug dosage and diagnosis of neonatal apnea.

ACKNOWLEDGEMENT

The authors are deeply indebted to doctors and medical staff of Neonatal Intensive Care Unit of Kasturba Medical College, Manipal Academy of Higher Education (MAHE), Karnataka for all the help in data collection and completion of the research problem.

REFERENCES

1. Zhao J, Gonzalez F, Mu D. Apnea of prematurity: from cause to treatment. Eur J Pediatr. 2011;170(9): 1097–1105.
2. Mishra S, Agarwal R, Deorari A K, Paul V K. Apnea in the newborn. The Indian Journal of Pediatrics. 2008;75(1): 57-61.
3. Lee H, Rusin C G, Lake D E, Clsark M T, Guin L, Smoot T J, et al. A new algorithm for detecting central apnea in neonates. Physiological measurement, IOP Science. 2012;333(1): 1-17.
4. Bancalari E, Sinclair J C. Mechanical ventilation. In: Sinclair JC, Bracken MB, eds. Effective Care of the Newborn Infant. Oxford University Press Inc. 1992: 200 – 220.
5. Shivakumar M, Jayashree P, Lewis L, Bhat Y R, Kamath A, Shashikala. Comparative Efficacy and Safety of Caffeine and Aminophylline for Apnea of Prematurity in Preterm (≤34weeks) Neonates: A Randomized Controlled Trial. Indian Pediatrics. 2017; 54(4): 279-283.
6. Aranda J V, Louridas A T, Vitullo B B, Thom P, Aldridge A, Haber R. Metabolism of theophylline to caffeine in human fetal liver. Science, 1979; 206(4424): 1319-1321.
7. Sop13: Pharmacokinetic data analysis. Onkologie. 2003;1(26): 56–59.
8. Sop14: Pharmacokinetic data analysis. Onkologie, 2003;1(26): 60–69.
9. Lee T C, Charles B, Steer P, Flenady V, Shearman A. Population pharmacokinetics of intravenous caffeine in neonates with apnea of prematurity. Clin. Pharmacol Ther. 1997;61(6): 628-640.
10. Charles B G, Townsend S R, Steer P A, Flenady V J, Gray P H , Shearman A. Caffeine citrate treatment for extremely premature infants with apnea: population pharmacokinetics, absolute bioavailability, and implications for therapeutic drug monitoring. Ther Drug Minit. 2008; 30(6): 709-16.
11. Laubsher B, Greenough A, Dimitriou g. Comparative effects of theophylline and caffeine on respiratory function of prematurely born infants. Sci Ireland Early Human Devel. 1998; 50(2): 185–192.
12. James J E, Charles B, Philip R K. Caffeine and neonatal apnea. J Caff Res. 2012;22(2): 55–61.
13. Dai W, Brisimi T S, Adams W G, Mela T, Saligrama V, Paschalidis I C. Prediction of hospitalization due to heart diseases by supervised learning methods. International Journal of Medical Informatics. 2015;84(3): 189 -197.
14. Oliveira T, Barbosa E, Martins S, Goulart A, Neves J, Novais P. A Prognosis System for Colorectal Cancer. Proceedings of 26thInternational symposium on Computer based Medical systems, Jun 2013, Porto, Portugal, IEEE, pp.481-484.
15. M. Ilayaraja and T. Meyyappan. "Mining medical data to identify frequent diseases using Apriori algorithm," Proceedings of International Conference on Pattern Recognition, Informatics and Medical Engineering (PRIME), Feb 2013, Salem, India, IEEE, pp.1-6.
16. Delen D, Walker G, Kadam A. Predicting breast cancer survivability: a comparison of three data mining methods, Artificial intelligence in medicine. 2005;34(2): 113–127.
17. Mikhno and C. M. Ennett,"Prediction of extubation failure for neonates with respiratory distress syndrome using the MIMIC-II clinical database," Proceedings of 34th Annual International Conference on Engineering in Medicine and Biology Society (EMBC), Aug-Sep 2012, San Diego, California, USA, IEEE, pp.5094–5097.
18. D. Precup, C. A. Robles-Rubio, K. A. Brown, L. Kanbar, J. Kaczmarek, S. Chawla, G. M. Sant'Anna and R. E. Kearney, "Prediction of extubation readiness in extreme preterm infants based on measures of cardiorespiratory variability,"Proceedings of 34th Annual International Conference on Engineering in Medicine and Biology Society (EMBC), Aug-Sep 2012, San Diego, California USA,IEEE, pp.5630–5633.
19. Mueller M, Almeida J S, Stanislaus R, Wagner C L. Can machine learning methods predict extubation outcome in premature infants as well as clinicians?, J Neonatal Biol, 2012;2(2).
20. C Baumgartner, Bohm C, Baumgartner D, Marini G, Weinberger K, Olgemoller B, Liebl B, Roscher A A. Supervised machine learning techniques for the classification of metabolic disorders in newborns. Bioinformatics, Oxford University Press 2004;20(17): 2985– 2996.
21. Williamson J R, Bliss D W, Browne, Indic P,

Bloch-Salisbury E, Paydarfar D. Individualized apnea prediction in preterm infants using cardio-respiratory and movement signals. Proceedings of IEEE Conference on Body Sensor Networks (BSN). 2013 May, Cambridge, USA, IEEE.
22. Williamson J R, Bliss D W, Browne, Indic P, Bloch-Salisbury E, Paydarfar D. Using physiological signals to predict apnea in preterm infants. Proceedings of 45th Asilomar Conference on Signals, Systems and Computers (ASILOMAR). 2011 Nov, Pacific Grove CA, USA, IEEE.
23. Shirwaikar R D, Acharya A, Makkithaya K, Surulivelrajan M, Lewis L E S. Machine learning Techniques for Neonatal Apnea prediction. Journal of Artificial Intelligence. 2016; 9(1-3): 33 -38.
24. Shirwaikar R D, Acharya D, Makkithaya K, Surulivelrajan M, Lewis L E S. Hegde G. Supervised Leaning Technique for analysis of neonatal data. Proceedings of 2nd International conference on Applied and Theoretical Computing and Communication Technology (iCATccT); 2016; Bangalore, India: IEEE; 2017.
25. Hu P J, Cheng T H, Wei C P, Yu C H, Chan L F, Wang H Y. Managing clinical use of high-alert drugs: A supervised learning approach to pharmacokinetic data analysis. Systems Man and Cybernetics, Part A: Systems and Humans IEEE Transactions. 2007;37 (4): 481–492.
26. Hu Y H, Wu F, Lo C L, Tai C T. Predicting warfarin dosage from clinical data: A supervised learning approach. Artificial Intelligence in Medicine.2012;56(1): 27-34.
27. Wenqi Y, Widmer N, Micheli G D. Example-based support vector machine for drug concentration analysis. Proceedings of 33rd Annual Conference Engineering in Medicine and Biology Society (EMBC); 2011 Aug-Sep; Boston, USA, IEEE 2011: 153–157.
28. Shirwaikar R D, Mago N, Acharya D, Makkithaya K, Hegde G, Lewis L E S, Shivakumar. Partition and Hierarchical Based Clustering Techniques for Analysis of Neonatal Data. Proceedings of International Conference on Cognition and Recognition, D. Guru, T. Vasudev, H. Chethan. Y. Kumar Eds. Lecture Notes in Networks and Systems. 2018; Singapore: 14.
29. Paul R, Latiful Abu H. Clustering medical data to predict the likelihood of diseases. Proceedings of 5th International Conference on Digital Information Management (ICDIM). 2010 Jul, Lakehead University, Thunder Bay, Canada, IEEE: 44-49.
30. Guyon I, Elisseeff A. An introduction to variable and feature selection. Journal of machine learning research. 2003: 1157-1182.
31. Rodriguez-Galiano V F, Ghimire B, Rogan J, Chica-Olma M, Rigil -Sanchez J P. An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS Journal of Photogrammetry and Remote Sensing. 2012;67: 93-104.
32. Shirwaikar R D, Acharya D, Makkithaya K, Surulivelrajan M, Lewis L E S. Estimation of caffeine regimens: a machine learning approach for enhanced clinical decision making at a Neonatal Intensive Care Unit (NICU). Critical Reviews in Biomedical Engineering. 2018;46(2): 93-15.
33. Yang S Y, Huang Q, Li L L, Teng Q Z, Ma C Y, Zhang H, Bai R, Teng Q Z, Xiang M L, Wei Y Q. 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. 2009;46(2); 155–163.
34. Bose N K, Liang P. Neural Network Fundamentals with Graphs, Algorithms and Applications. McGraw-Hill Series in Electrical and Computer Engineering, 1996.
35. Shanthi D, Sahoo G, Saravanan N. Designing an artificial neural network model for the prediction of thrombo-embolic stroke. International Journals of Biometric and Bioinformatics (IJBB). 2009;3(1): 10-18.
36. Shirwaikar R D, Acharya D, Makkithaya K, Surulivelrajan M, Shrivastava S, Lewis L E S. Optimizing Neural Networks for medical data sets: A case study on Neonatal Apnea prediction. Artificial Intelligence in Medicine. 2019;98: 59-76.
37. Brett Lantz. Machine Learning with R. 2nd ed. India: Packt Publication Ltd; 2014.
38. Han J, Kamber M. Data mining concepts and Techniques. 3rd edition, San Fransisco, Elsevier; 1998