Comparison of Some Prediction Models and their Relevance in the Clinical Research

Nihar Ranjan Panda[1,2], Kamal Lochan Mahanta[2], Jitendra Kumar Pati[3], Pavankumar Reddy Varanasi[1] and Ruchi Bhuyan[1,*]

[1]*Department of Medical Research, IMS and SUM Hospital, SOA deemed to be University, Bhubaneswar, Odisha, India*

[2]*Department of Mathematics CV Raman Global University, Bhubaneswar, Odisha, India*

[3]*Department of Mathematics Gandhi Engeeniearing College, Bhubaneswar, Odisha, India*

Abstract: In healthcare research, predictive modeling is commonly utilized to forecast risk variables and enhance treatment procedures for improved patient outcomes. Enormous quantities of data are being created as a result of recent advances in research, clinical trials, next-generation genomic sequencing, biomarkers, and transcriptional and translational studies. Understanding how to handle and comprehend scientific data to offer better treatment for patients is critical. Currently, multiple prediction models are being utilized to investigate patient outcomes. However, it is critical to recognize the limitations of these models in the research design and their unique benefits and drawbacks. In this overview, we will look at linear regression, logistic regression, decision trees, and artificial neural network prediction models, as well as their advantages and disadvantages. The two most perilous requirements for building any predictive healthcare model are feature selection and model validation. Typically, feature selection is done by a review of the literature and expert opinion on that subject. Model validation is also an essential component of every prediction model. It characteristically relates to the predictive model's performance and accuracy. It is strongly recommended that all clinical parameters should be thoroughly examined before using any prediction model.

Keywords: Predictive modeling, Risk estimation, Probability, Public health, Clinical outcomes.

INTRODUCTION

Predictive modeling is a popular tool for predicting illness, risk, progression, and consequences. The predictive modeling methodology works with a wide range of traditional statistical methodologies, probability rules, and optimization methods. These models may be trained using existing data sets and used to predict test data output [1]. These models are not limited to public health, but also have a wide variety of applications in other industries such as banking, education, and real estate [2, 3].

In the twenty-first century, robust methodologies and improvements enhanced data capture and acquired relevance in data science and machine learning. As a result, researchers and analysts were driven to create or implement new methodologies. To this goal, the popularity of predictive modeling is astonishing since it can properly anticipate every single disease and risk of disease. However, we cannot be contingent just on prediction models since there may be additional clinical criteria at stake; so, clinical judgment is also compulsory.

The preponderance of these prediction models has used a similar analytical approach to disease diagnosis and prognosis. The major distinction is the disease's outcomes, i.e., whether the disease is present or not (diagnostic), and if there is any possibility of developing the disease in the future (prognosis). Cross-sectional research designs are applied to plan diagnostic clinical prediction models, whereas longitudinal study designs are utilized to develop prognostic clinical prediction models [4, 5]. Normally, researchers and scientists validate the predicted accuracy of the model before adopting it in health research [6, 7]. Predictive modeling approaches are created by learning a dataset with known labels and then predicting the outcome of unlabeled cases. The purpose of this article is to explore several prediction models used in health research.

Clinical prediction algorithms are routinely used for radiological parameters [6]. As we all know, imaging tests have regularly been used to diagnose any disease. Ultrasound imaging for gastrointestinal patients, for example, is utilized to assess risk stratification for this kind of disease. These imaging test results may now be employed as critical factors in a clinical prediction model. Radiological imaging techniques were created to aid in the early identification of disease and the prevention of disease. However, these imaging modalities are crucial in the clinical prediction model.

*Address correspondence to this author at the Department of Medical Research, IMS and SUM Hospital Siksha O Anusandhan Deemed to be University, Bhubaneswar, India; Tel: 9437213002;
E-mail: ruchibhuyan@soa.ac.in

E-ISSN: 1929-6029/23

Clinical prediction models are primarily concerned with public health, medical research, and clinical practice. These are the three primary fields in which clinical prediction models are most commonly utilized. Prediction of the future incidence of any particular disease is crucial in public health and is often used to target preventative interventions. For example, the most prevalent prediction model known as Framingham risk functions is frequently used to identify individuals who have experienced cardiovascular problems.

Similarly, in clinical research, the prediction model may be used in a wide range of ways. The majority of clinical models are concerned with clinical and demographic variables. These models may be used for medical tests by disease probability. Using a clinical prediction model, we can deliver appropriate therapies to patients.

Definitions

The presence of disease status (yes/no) is an example of a dependent variable. Similarly, independent variables are predictors that aid in the development of a clinical prediction model. While considering disease status (yes/no) as a dependent variable, independent factors such as age, gender, smoking status, disease history, and BMI can be included.

Study Design, Sample Size, and Development of Clinical Prediction Model

They are classified into two types based on the type of investigation or study design: 1) observational and 2) experimental [8]. These research designs are further classified as cross-sectional, case-control, randomized control trials, non-randomized control trials, and so on (Table **1**). The clinical prediction model may be applied and verified in a variety of research designs. Cross-sectional research designs were most commonly employed in diagnosis, while clinical prediction models and longitudinal study designs were utilized for prognostic clinical prediction models. For clinical prediction, some researchers employ a case-control study design. Cohort studies are the ideal research design for prognostic analyses. For example, symptomatic or asymptomatic individuals were included in the research and followed up on for a certain period. This type of study can be done retrospectively or prospectively; however, it is most commonly referred to as a prospective longitudinal cohort study. However, retrospective studies and clinical prediction models work with an existing database, such as a health survey or hospital database. Similarly, randomized control trial studies might be called prospective study designs. This type of clinical prediction model is mostly based on patient inclusion and exclusion criteria, as well as single-blinded or double-blinded randomization.

Study Design

Sample Size

A crucial component of a clinical prediction model is the sample size. Good clinical models typically favor a bigger cohort or a predetermined sample size. Larger cohorts are employed to create the prediction model since the sample size directly affects the model's dependability. Additional metrics for data fit must be taken into account since failure to do so might result in

Table 1: Classification of Different Study Design and their Applications

S. No	Study Category	Definition	Application
1	Cross-sectional	This kind of study design deals with the prevalence and frequency of any particular disease.	Diagnostic and clinical prediction model
2	Cohort study	This kind of study usually deals with prospective (follow-up patients) or retrospective (records). Usually aims to find out the relative risk of developing a disease.	Diagnostic and prognostic clinical prediction model
3	Case-control study	Comparison of two groups: cases and controls are made to find out the odds ratio or risk factor for a particular disease.	Diagnostic clinical prediction model
4	Randomized Control Trails	This kind of study is usually done by giving interventions to a particular group and comparing it with the standard treatment group. Methods of randomization are implemented in this study.	Diagnostic clinical prediction model

either an excessive or inadequate fit of the clinical prediction model. Prior sample size estimation can be used to lower data bias or inaccuracy. The study's findings are used to estimate the sample size. So, in clinical prediction models, we can find out three basic outcomes. i.e., binary outcomes, time-to-event outcomes, and continuous outcomes. Sample sizes are calculated using different formulas and different parameters according to the study outcomes[9].

Data Preparation

The most crucial factors that take up a lot of effort before attempting to develop predictive models are data preparation and presentation. Essentially, data preparation is the process of arranging data to make it suitable for analysis. Data preparation is concerned with preparing the data for analysis. To begin, we collect raw or unstructured data that must be adjusted or coded to create a valid prediction model.

Missing data

Sets impede model development and data visualization. Dealing with missing data is a significant difficulty. As a result, it is critical to comprehend the approaches for dealing with missing data. Today's programming languages offer significant gains in dealing with missing data. One-step sweep techniques, imputation of mean methods, multiple imputation methods, and ignoring the predictor's methods are the most often used missing data handling approaches[10].

Development of Prediction Model

Linear Regression

Regression analysis is a machine learning approach that is widely used in statistics to figure out the relationship between two or more variables. The dependent and independent variables must have a linear relationship. One of these factors is known as the predictor variable, and its value is obtained by experimentation, such as predicting the duration of stay of patients using other predictors. The other variable is known as the response variable, and its value is determined by the predictor variable. In linear regression, the straight line is shown using a single equation. A straight line is typically used to describe a linear relationship between two variables on a graph.

The correlation coefficient is closely associated with linear regression. Using the correlation coefficient researchers can evaluate the relationship between continuous variables. Whether the two variables are affected by each other or not. Similarly, linear regression builds a prediction model to find out the continuous response variables by considering the independent variables associated with the study design. As we have seen earlier, if two variables are linearly correlated we can essentially draw a line through their scatter plot which depicts the relationship between them. But the problem is we can draw many lines, and until now have no clue as to which one to choose finally. However, linear regression techniques are widely used predictive modeling techniques for predicting continuous response variables. In past research, a linear regression model has been used to predict any disease cases. S. Rath *et al*. [11] used linear regression analysis to predict the new active cases of the coronavirus disease (COVID-19) pandemic using a multiple linear regression model.

Multiple linear regressions can be defined as an extended version of linear regression. where the relevant parameters can be added to the study design and the basic principles remain the same. Instead of considering one independent variable x, we have multiple predictor variables x1, x2, x3......Xp. The target y can still be written as a linear combination of these predictors: The model can be performed by the below equation.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots \beta_i x_i + e_i$$

where $\beta_0, \beta_1, \beta_2 \ldots\ldots \beta_i$ are the regression coeff.

However, multicollinearity plays a big role in the model-building procedure. Usually, we would prefer there should not to be a very high linear relationship among the explanatory variables. Multi-Colinearity means, one or more of the predictor variables are linearly dependent on a few other predictor variables. There is a very important consequence of this in MLR. We discard or remove the variables which have a very high vif(variance influence factor) value. Usually, the cut of value for vif is less than or equal to 5. If any of the predictors have if greater than 5 we would like to discard it from the analysis.

Logistic Regression

Logistic regression which is also known as a logistic regression classifier is the most dominant predictive model nowadays. These types of predictive models are based on probability [12, 13]. The main advantages of these models are it does not require any prior assumptions to build the model. However, the response variables must be dichotomous. Usually, for a model-building strategy, we first do a univariate analysis. Which is usually performed by the chi-square test. There we have to verify which variables are statistically significant with the outcome variable. The significant predictors were usually determined if we found the p-value < 0.05. Finally, all the significant variables in the chi-square test were used to perform a multivariate logistic regression analysis.

Priyanka *et al.* [14] used a logistic regression model to predict the diabetes status of the patient without making any diagnosis. The author used feature selection, data processing, normalization and cross validation methods in the paper. Dental caries is a common problem nowadays. The older age group is most affected by dental caries. Hence, it's very essential to identify, what are the major factors affecting dental caries in old age. Liu Lu *et al.* [15] considered the Geriatric Residents of a certain community in China to predict dental caries among the people. The author implemented a logistic regression prediction model to predict dental caries. In this research, the authors considered age, gender, lifestyle, smoking habit, alcohol consumption, Domestic water access, Use of dental floss, Use of a toothpick, Use of fluoride toothpaste, Number of true teeth, having toothache in previous years, etc. to create a model. The consistency of his prediction model was found to be 76.50%. Rakesh *et al*. [16] considered environmental factor prediction in preterm birth. In this research, the authors developed two predictive models, i.e., logistic regression and decision tree. However, logistic regression performs better than decision trees in their research. The author considered age, GSH (Glutathione), MDA (Malondialdehyde), α-HCH, γ-HCH, total HCH, and total DDT (dichlorodiphenyltrichloroethane) as predictors in this particular study.

Stillbirth is a big issue among pregnant women; many fetuses die during pregnancy as a result of stillbirth. Kidanemariam *et al*. [17] used a logistic regression model to determine the variables influencing stillbirth. He performed a univariate analysis and then incorporated all of the important factors in the final logistic regression model. The author discovered that the mother's age, region of stay, place of residence, education level, parity, antenatal care utilization, place of delivery, mode of delivery, body mass index (BMI), and anemia level are all significantly related to stillbirth; thus, all significant predictors were included in the final model for analysis. The model's accuracy was quite good (99.9%). We have reviewed the prediction model for the classification problem, also known as the logistic regression model. It provides the concept of odds and adjusted odds ratios, which is essential in health research.

Decision Tree

A decision tree is a promising predictive modeling algorithm nowadays used in the clinical research domain [18]. In the past, it was developed to predict

Table 2: Comparison of Predictive Modeling Algorithms

Predictive modeling	Advantages	Disadvantages
Linear Regression	1. These techniques are easy to implement while predicting continuous response variables. 2. Output is clear and easy to interpret and understand.	1. Needs lots of statistical assumptions 2. Cannot detect any non-linear pattern in the data set.
Logistic Regression	1. No prior assumptions are needed for this. 2. It has a good probabilistic interpretation of results and model parameters.	1. It cannot provide an accurate result for a complex data set 2. Not able to classify when a response has more than two categories.
Decision tree	1. The performance of the tree model is easy to understand and interpret. 2. Decision trees can be validated by using statistical tests and can create robust classifiers.	1. The Algorithm used in the decision tree depends on the order of the attributes or variables. 2. Sometimes cannot perform better as compared to other classifiers.
Artificial Neural network	1. Best algorithm to detect a complex pattern in the data. 2. It does not require any statistical assumptions or statistical training.	1. Result interpretation is very difficult in neural network algorithms. 2. It has the characteristic of a black box.

categorical data initially, but later on, it was also used to predict continuous variables. A decision tree is usually used to create a classification and regression trees (CART) model to predict continuous variables and for categorical variables, it uses the chi-square automatic interaction detection method (CHAID). It builds a tree by using the underlying algorithm. The construction of a decision tree is like a tree that provides logical outputs. The decision tree usually has multiple levels, the top level is called the root node. All internal nodes (nodes having at least one child) denote tests on input variables. A decision tree is easy to understand and is especially used for the diagnostic protocol. Malignant pleural mesothelioma (MPM) is rare cancer with a heterogeneous prognosis. Brims FJ [19] derived the classification and interaction effect of clinical parameters on the outcome. The author used the decision tree prediction model to classify the high and low risk of death using clinical parameters. This model was quite effective for malignant pleural mesothelioma (MPM) disease.

Artificial Neural Network

Neural network techniques are now the most popular and dominant predictive modeling technique almost in every field of clinical research [20, 21]. An artificial neural network is an interconnected group of nodes, similar to the connection of neurons in the brain, and is inspired by the animal's central nervous system. This prediction model is a comprehensive and preferable method of prediction of target variables. The main purpose of neural networks is to identify the complex relationship that exists in the data set. The structure of neural networks is alike neurons in the human brain. The structure of the neural contains 3 possible layers,i.e. input layer,the hidden layer, and the output layer. Based on the complexity and size of the data the hidden layer can be either one or two. The neural network evaluates the variable's importance based on the normalized dataset. This variable importance provided by the neural networks is produced by the relative influence of each variable. It provides a ROC curve for each analysis, which is based on sensitivity vs. specificity.Understanding the complex relationship between the data set is often difficult. The artificial neural network can recognize the complexity of the data set and provide the best prediction model.

Rau HH *et al*. [22] developed a web-based liver cancer prediction model for type II diabetes patients by using an artificial neural network. The author created 3 sub-models and compared all the sub-model with the logistic regression model. The authors considered gender, age group, hyperlipidemia, alcoholic fatty liver disease, viral hepatitis, alcoholic hepatitis, and alcoholic cirrhosis as predictors to predict the liver cancer of the patients. However, in the end, he considered a suitable prediction model for type II diabetes patients. Li H *et al*. [23] developed an artificial neural network prediction model for congenital heart disease based on risk factors.here authors included socio-demographic characteristics, pregnancy history, family history, environmental risk factors, and dietary and lifestyle behaviors to predict the heart disease of the patients. Then he divided the data into a training set and a testing set for a comparative study. However, he ended up creating a suitable predictive model.

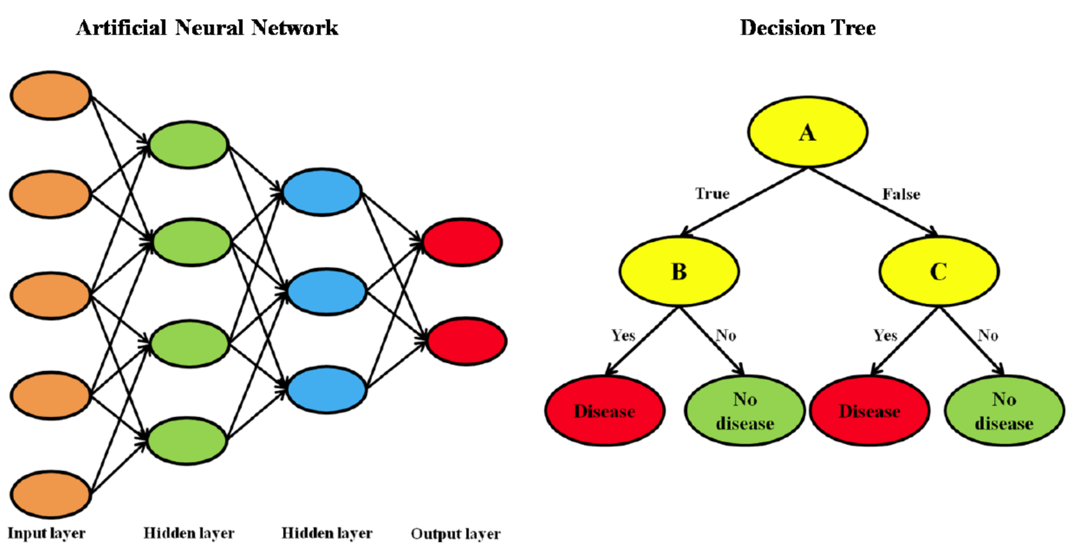

Figure 1: Graphical representation of the artificial neural network and decision tree.

Table 3: Examples of some Prediction Models Used in Different Diseases

Reference	Disease	Model	Data Type	Sample size	Performance
Priyanka *et al.* [14]	Diabetes	LR	Clinical data	768	Accuracy (LR = 0.74)
Rau HH *et al.* [22]	Liver Cancer	LR, ANN	Clinical data	2060	Accuracy (LR = 0.666ANN =0.757)
Rakesh *et al.* [16]	Preterm Birth	LR, DT	Hospital-based data	90	Accuracy (LR = 0.78, DT=0.67)
Berhie KA *et al.* [17]	Stillbirth Prediction	LR	Demographic and health survey data	12560	Accuracy (LR = 0.99)
Li H *et al.* [23]	Heart Disease in pregnant women	ANN	Hospital-based data	336	Accuracy (ANN = 0.91)
Bekesiene S *et al.* [24]	Stress prediction	ANN	Covid 19 armed forces data	111	Accuracy (ANN = 0.85)
Kwang-Sig *et al.* [25]	Spontaneous Preterm Labor and Birth	ANN, DT, LR	Hospital-based data	596	Accuracy (LR = 0.918) ANN =0.911, DT=0.832)
xing xiao *et al.* [26]	Chronic kidney disease	ANN, LR	Hospital-based data,Department of Nephrology,Huadong Hospital, Shanghai	551	Accuracy (LR = 0.82,ANN =0.80)
Shu-Ping *et al.* [27]	MICU Survival Prediction	ANN, LR	Hospital-based data	1496	Accuracy (LR = 0.60, ANN =0.69)
Shiva *et al.* [28]	Thyroid disorders	ANN, LR	Hospital-based data	310	Accuracy (LR = 0.91, ANN =0.96)
Zhou *et al.* [29]	Pancreatic cancer	ANN, LR	Clinical data from cancer patients	221	Accuracy (LR = 0.736, ANN =0.774)
Gyanendra *et al.* [30]	Thyroid Disease	Logistic regression, decision trees	Graven Institute in Sydney	215	Accuracy (LR=0.81, DT=0.87)
Yutaka *et al.* [31]	Lower Back Pain	LR,	Clinical data	96	Accuracy (LR=0.78
Guo li *et al.* [32]	Schistosomiasis Prognosis	LR, DT, ANN	A previous study based on a Hubepopulation sample	4136	Accuracy (LR = 0.799, ANN =0.803,DT=0.803)

CONCLUSION

Various models are now being used to predict outcomes; here we explore the advantages and disadvantages of linear regression, logistic regression, decision tree, and artificial neural network models in depth. Decision trees and artificial neural networks may be utilized for both categorical and continuous outcome variables. Clinical prediction models are highly important nowadays for the diagnosis or prognosis of any disease, allowing us to avoid diseases accordingly. It is because of the capacity to detect complicated correlations in data and deliver a more accurate estimate. However, picking the right clinical prediction model for diagnostic and prognostic investigations is crucial. Estimating the presence/occurrence or potential future course of a specific clinical outcome is crucial in clinical research and the public health field. The ability to handle all of these estimating techniques is quite beneficial for predictive models. The clinical prediction model may be used in numerous observational and diagnostic investigations, randomized control trials, and other medical research projects. Nevertheless, the primary goal of this review was to compile clinical prediction models, which are often employed in the fields of public health and medical research.

ABBREVIATIONS

VIF = Variance influence factor

MLR = Multiple linear regression

CART = Classification and regression tree

CHAID = Chai square automatic interaction detection

DT = Decision tree

ANN = Artificial neural network:

ROC = Receiver operating curve

CONFLICT OF INTEREST

No potential conflict of interest relevant to this article was reported.

FUNDING

NA.

REFERENCES

[1] van Leeuwen J. Approaches in Machine Learning, in Algorithms in Ambient Intelligence, Verhaegh WFJ, Aarts E, Korst J, Eds., Springer Netherlands: Dordrecht 2004; pp. 151-166. https://doi.org/10.1007/978-94-017-0703-9_8

[2] Aleskerov E, Freisleben B, Rao RB. CARDWATCH: a neural network based database mining system for credit card fraud detection. Proceedings of the IEEE/IAFE 1997 Computational Intelligence for Financial Engineering (CIFEr), 1997; pp. 220-226.

[3] Kim E, Kim W, Lee Y. Combination of multiple classifiers for the customer's purchase behavior prediction. Decision Support Systems 2003; 34(2): 167-175. https://doi.org/10.1016/S0167-9236(02)00079-9

[4] Collins GS, *et al.* Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting. BMC Medicine 2011; 9(1): 103. https://doi.org/10.1186/1741-7015-9-103

[5] Mallett S, *et al.* Reporting methods in studies developing prognostic models in cancer: a review. BMC Medicine 2010; 8(1): 20. https://doi.org/10.1186/1741-7015-8-20

[6] Han K, Song K, Choi BW. How to Develop, Validate, and Compare Clinical Prediction Models Involving Radiological Parameters: Study Design and Statistical Methods. Korean J Radiol 2016; 17(3): 339-50. https://doi.org/10.3348/kjr.2016.17.3.339

[7] Collins GS, *et al.* External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. BMC Med Res Methodol 2014; 14: 40. https://doi.org/10.1186/1471-2288-14-40

[8] Chidambaram AG, Josephson M. Clinical research study designs: The essentials. Pediatr Investig 2019; 3(4): 245-252. https://doi.org/10.1002/ped4.12166

[9] Riley RD, Ensor J, Snell KIE. Calculating the sample size required for developing a clinical prediction model 2020; 368: m441. https://doi.org/10.1136/bmj.m441

[10] Janssen KJ, *et al.* Dealing with missing predictor values when applying clinical prediction models. Clin Chem 2009; 55(5): 994-1001. https://doi.org/10.1373/clinchem.2008.115345

[11] Rath S, Tripathy A, Tripathy AR. Prediction of new active cases of coronavirus disease (COVID-19) pandemic using multiple linear regression model. Diabetes Metab Syndr 2020; 14(5): 1467-1474. https://doi.org/10.1016/j.dsx.2020.07.045

[12] Panda NR. A Review on Logistic Regression in Medical Research. National Journal of Community Medicine 2022; 13(04): 265-270. https://doi.org/10.55489/njcm.134202222

[13] Panda NR, Pati JK, Bhuyan R. Role of Predictive Modeling in Healthcare Research: A Scoping Review. International Journal of Statistics in Medical Research 2022; 11: 77-81. https://doi.org/10.6000/1929-6029.2022.11.09

[14] Rajendra P, Latifi S. Prediction of diabetes using logistic regression and ensemble techniques. Computer Methods and Programs in Biomedicine Update 2021; 1: 100032. https://doi.org/10.1016/j.cmpbup.2021.100032

[15] Liu L, *et al.* Dental Caries Prediction Based on a Survey of the Oral Health Epidemiology among the Geriatric Residents of Liaoning, China 2020; 2020: 5348730. https://doi.org/10.1155/2020/5348730

[16] Saroj RK, Anand M. Environmental factors prediction in preterm birth using comparison between logistic regression and decision tree methods: An exploratory analysis. Social Sciences & Humanities Open 2021; 4(1): 100216. https://doi.org/10.1016/j.ssaho.2021.100216

[17] Berhie KA, Gebresilassie HG. Logistic regression analysis on the determinants of stillbirth in Ethiopia. Maternal Health, Neonatology and Perinatology 2016; 2(1): 10. https://doi.org/10.1186/s40748-016-0038-5

[18] Aleem IS, Schemitsch EH, Hanson BP. What is a clinical decision analysis study? Indian J Orthop 2008; 42(2): 137-9. https://doi.org/10.4103/0019-5413.40248

[19] Brims FJ, *et al.* A Novel Clinical Prediction Model for Prognosis in Malignant Pleural Mesothelioma Using Decision Tree Analysis. J Thorac Oncol 2016; 11(4): 573-82. https://doi.org/10.1016/j.jtho.2015.12.108

[20] Shahid N, Rappon T, Berta W. Applications of artificial neural networks in health care organizational decision-making: A scoping review. PLOS ONE 2019; 14(2): e0212356. https://doi.org/10.1371/journal.pone.0212356

[21] Patel JL, Goyal RK. Applications of artificial neural networks in medical science. Curr Clin Pharmacol 2007; 2(3): 217-26. https://doi.org/10.2174/157488407781668811

[22] Rau HH, *et al.* Development of a web-based liver cancer prediction model for type II diabetes patients by using an artificial neural network. Comput Methods Programs Biomed 2016; 125: 58-65. https://doi.org/10.1016/j.cmpb.2015.11.009

[23] Li H, *et al.* An artificial neural network prediction model of congenital heart disease based on risk factors: A hospital-based case-control study. Medicine 2017; 96(6). https://doi.org/10.1097/MD.0000000000006090

[24] Bekesiene S, Smaliukiene R, Vaicaitiene R. Using Artificial Neural Networks in Predicting the Level of Stress among Military Conscripts. Mathematics 2021; 9(6): 626. https://doi.org/10.3390/math9060626

[25] Lee KS, Ahn KH. Artificial Neural Network Analysis of Spontaneous Preterm Labor and Birth and Its Major Determinants 2019; 34(16): e128. https://doi.org/10.3346/jkms.2019.34.e128

[26] Xiao J, *et al.* Comparison and development of machine learning tools in the prediction of chronic kidney disease progression. Journal of Translational Medicine 2019; 17(1): 119. https://doi.org/10.1186/s12967-019-1860-0

[27] Lin SP, *et al.* A comparison of MICU survival prediction using the logistic regression model and artificial neural network model. J Nurs Res 2006; 14(4): 306-14. https://doi.org/10.1097/01.JNR.0000387590.19963.8e

[28] Borzouei S, *et al.* Diagnosing thyroid disorders: Comparison of logistic regression and neural network models. J Family Med Prim Care 2020; 9(3): 1470-1476. https://doi.org/10.4103/jfmpc.jfmpc_910_19

[29] Tong Z, *et al.* Development, Validation and Comparison of Artificial Neural Network Models and Logistic Regression Models Predicting Survival of Unresectable Pancreatic Cancer. Frontiers in Bioengineering and Biotechnology 2020; 8. https://doi.org/10.3389/fbioe.2020.00196

[30] Chaubey G, *et al.* Thyroid Disease Prediction Using Machine Learning Approaches. National Academy Science Letters 2021; 44(3): 233-238. https://doi.org/10.1007/s40009-020-00979-z

[31] Owari Y, Miyatake N. Prediction of Chronic Lower Back Pain Using the Hierarchical Neural Network: Comparison with Logistic Regression-A Pilot Study. Medicina (Kaunas) 2019; 55(6). https://doi.org/10.3390/medicina55060259

[32] Li G, *et al.* Comparison of three data mining models for prediction of advanced schistosomiasis prognosis in the Hubei province. PLOS Neglected Tropical Diseases 2018; 12(2): e0006262. https://doi.org/10.1371/journal.pntd.0006262

Received on 05-01-2023 Accepted on 14-02-2023 Published on 08-03-2023

https://doi.org/10.6000/1929-6029.2023.12.02