Systematic Methods on Machine Learning Techniques for Clinical Predictive Modelling

Radhika A, Priya G

Abstract: ***Predictive modelling is a mathematical technique which uses Statistics for prediction, due to the rapid growth of data over the cloud system, data mining plays a significant role. Here, the term data mining is a way of extracting knowledge from huge data sources where it's increasing the attention in the field of medical application. Specifically, to analyse and extract the knowledge from both known and unknown patterns for effective medical diagnosis, treatment, management, prognosis, monitoring and screening process. But the historical medical data might include noisy, missing, inconsistent, imbalanced and high dimensional data.. This kind of data inconvenience lead to severe bias in predictive modelling and decreased the data mining approach performances. The various pre-processing and machine learning methods and models such as Supervised Learning, Unsupervised Learning and Reinforcement Learning in recent literature has been proposed. Hence the present research focuses on review and analyses the various model, algorithm and machine learning technique for clinical predictive modelling to obtain high performance results from numerous medical data which relates to the patients of multiple diseases.***

Keywords: Machine learning, clinical predictive modelling, healthcare application, disease prediction methods.

I. INTRODUCTION

Nowadays, Data Mining (DM) and Knowledge Discovery Process (KDP) plays a vital role due to the growing size of the data over the cloud platform specifically on medical application with respect to the symptoms of diseases and how to predict and diagnose it correctly (Wang et al. 2019; Vaidya et al. 2014; Eissa, Elmogy, and Hashem 2016). In clinical predictions model, KDP pursues toward gathering knowledge through finding a relationship amongst numerous data attributes (Ma et al. 2019). On the other hand, the knowledge discovery process supports medical experts towards preventing false prediction results, diagnostic, identified the relationships amongst different medical indicators, detection of diseases as well as enhance the process of treatment decision making (Eissa, Elmogy, and Hashem 2016). In order to analysing these kind of data, researchers, governments and most of the organizations offered enhanced services as well as improve value to their relations with their customers, users, patients and so on (Parikh, Kakad, and Bates 2016; Moloud Abdar et al. 2019; Darcy, Louie, and Roberts 2016).

Revised Manuscript Received on March 30, 2020.
***** Correspondence Author
Dr. A. Radhika*, Assistant Professor, Department of Statistics, Periyar University, Salem, TamilNadu, India, email: radhisaran2004@gmail.com
G. Priya, Research Scholar Department of Statistics, Periyar University, Salem, TamilNadu, India, email: gpriya30@yahoo.co.in

In clinical predictive modeling machine learning technique plays a significant role for effective decision making, optimizing the cost and classification of disease, which will enhance the learning and system service performance.

Most of the researcher suggested various recognition and classification method like artificial neural network, Naïve Bayesian, support vector machine, logistic regression and so on (Kaur and Kumari 2018). But till date, we are facing the problem during the extraction of hidden and useful data from huge set of multi-dimensional medical repositories. Hence there is a need to study and review the various machine learning technique for clinical predictive modelling as well as recommend the solution method for future research.

The remaining study was structured as follows: Section 2 discusses the summary of machine learning technique and role in the clinical prediction process. Section 3 presents the previous studies related to machine learning method with respect to diagnosis and various disease prediction models in medical application. Section 4 defines the materials and search methods in this paper. Section 5 discusses the available study results and compares the performance in terms of disease prediction and classification approach. Section 6 concludes this study and recommends guidelines for future work.

II. OVERVIEW OF MACHINE LEARNING

The Machine Learning is a powerful procedure that permits computers toward learning the information from both unstructured and structured data (Kaur and Kumari 2018). On the other hand, it permits the system toward learning and gains intelligence based on the earlier input data. In addition, it can identify and understand the information that makes predictions and decisions (Sun 2013; Kaur, Lechman, and Marszk 2017; Kaur and Kumari 2018). Furthermore, it generalize the data patterns and decision rules from a labelled set of input as well as consumes that information toward produce classifications or predictions on data (Mateos-Pérez et al. 2018). The clinical predictive modelling process applies to different health care sectors which aims to optimize the resource, improve clinical outcomes, accurately diagnose the disease and enhancement of patient care (Nithya and Ilango 2017). Generally, ML is a kind of artificial intelligence (AI) which is successfully applicable toward solving the biological issue (Zamani Esfahlani et al. 2018; Ker et al. 2018; Walczak and Velanovich 2017).

Particularly, ML concern to the design and implementation of computational methods which uses the data toward enhancing the system performance in terms of forecasting the accuracy (Mohri, Rostamizadeh, and Talwalkar 2012).

The Machine Learning technique offers an effective solution toward enhance the drug discovery process and prediction of drug–tissue relations (Turki and Taguchi 2019). The machine learning technique comprises five stages during at implementation such as preparing the input data, estimator selection, estimator evaluation, regularization and parameter learning (Doupe, Faghmous, and Basu 2019). First, the dataset will be prepared by considering the details of input-output pair specifically the output pair will contain the details of human had a hospitalization with respect to the binary outcome variables and patient covariates as considered as an input. After that, the original data will be divided into three subsets: test data, training data and validation dataset as a ratio of 15%, 70% and 15% respectively (Hastie, Tibshirani, and Friedman 2009).

Second, we will select the suitable class of estimators, which mapping both input and outputs. In most of the medical industries, supervised and unsupervised learning approach is utilized. Third, we will estimate the learning features via analysing the iterated information. Fourth, we will shorten the estimator through fining the system/ data complexity. At last, validate the performance of each function toward testing the generalized features of learned estimator as well as will calculate the predictor and actual observed results in testing phase.

Based on the above stages, research by Ma et al. (2019), applied ML technique in the prediction of disease for extraction of valuable knowledge and detecting hidden data. A study suggested that the random forest is an effective data mining framework, that has been extensively used in different real-time medical predictor scenarios (Szűcs 2013). Additionally, the RF method is widely applicable to various applications, particularly, ability to handle parallel task, data usability for numerous DM process, and measurements of strong disease prediction model. A research by (Ma et al. 2019) presented a RF model for clinical disease prediction model. Also, a research by (Nithya and Ilango 2017) presented an analytical model which is automatically analyse and predict the result. In order to find the hidden insights without analysing the source code, the ML technique has been utilized by the most of the researcher. The commonly used ML techniques are summarized in table 1.

Table 1: Merits and demerits of Machine Learning techniques applied in healthcare which is reported in Doupe et al. (2019).

S.No	Method	Intuition	Merits	Demerits
1.	**Unsupervised learning:** K-means clustering, principle component analysis (PCA), factor analysis, hierarchical clustering, Neural network (NN)	Based on the description of correlated, the input data are clustered into original dimensions	Based on the measurement of original commonalities, we can simplify the noisy or complex information Able to classify institutions or people into data groups.	Cannot determine the accuracy of the results in precise manner.
2	The ensemble of decision trees: gradient boosting machines	The input data weighted subsets are fitter into multiple decision trees whereas the prediction errors in the first tree notify like how to enhance the next tree performance during the implementation	Able to offer highest performance with respect to the error rate amongst predicted and observed outcomes when compared to the other ML techniques	It needs more effort for tuning the parameter to obtain the optimal performance. this research work does not describe the some specific mechanism.

3	**Deep learning:** Neural networks	A sequence of data transformation, where the results from one sequence of transformations data and the inputs to the next sequence of transformations, repetitively via multiple layers of transformations, ultimately provides the abstractions/ generalization from the data.	It helps to forecast the result for high complex interactions and non-linear relationships. It can offer the better identification result for extremely large and noisy, and non-tabular database.	It requires more influence for computation process.
4	Regularization	To minimize the over fitting, penalize the estimators which contains multi-collinearity and covariates	It yields the simple estimators and also enhances the generalizability. It offers more accurate results for even small changes in estimator.	Sometimes it selects false data during prediction process of highly correlated predictors. It contains more computational complexity.
5	Decision trees	Based on the value of specific features to categorized the data.	It has good interpretability	It contains over fitting issues.
6	An collection of decision trees: random forest	Fit numerous decision trees to bootstrap-resampled methods of the data; a) result of average regression result process b) for the process of classification).	Requires little researcher effort to “tune” the estimators to ensure the optimal performance; rapid movement to implement	It doesn’t offered the detailed explanation of mechanism for corresponding outcomes so it provides enhanced prediction than conclusion
7	Machine learning meta-learners	It integrates the numerous tools of machine learning to attain the prediction results	The ensembles can yield an excellent approximation of that function	Time- and computing-power intensive; also it doesn’t explain any mechanism for result, hence it provides enhanced results prediction than inference.

A. Machine Learning's Role in Predictions

Machine learning offered the efficient algorithms, and this model achieves the high accuracy rate, and it utilized some real-time applications, the value of enterprises dataset and contributing to various approaches succeeding (Nithya and Ilango 2017). Based on predictive value, the optimize decisions are found by, machine learning approaches in large scale dataset. The predicative tasks are effectively handled by the machine learning algorithm, and it includes the behaviour that has maximum tendency to perform preferred results. The general structure for clinical prediction modelling is shown in figure 1.

B. Variables in Prediction Models

In common, the prediction models are implemented with some specific both input (medical data and baseline demographic data) and output variables that are evaluated or performed by specific techniques. In the machine learning process, it can be classified into supervised learning, unsupervised learning and semi-supervised learning. In supervised learning process, the resultant model is applied to forecast the result of original datasets (medical data). In unsupervised learning process used to find the unknown pattern of the data. Moreover, the semi-supervised learning process, the small portion of input data are derived based on their respective output variables. Selecting those variables is to consider while implementing a prediction model is one of the great significance, and it can influence how the model used in longitudinal studies.

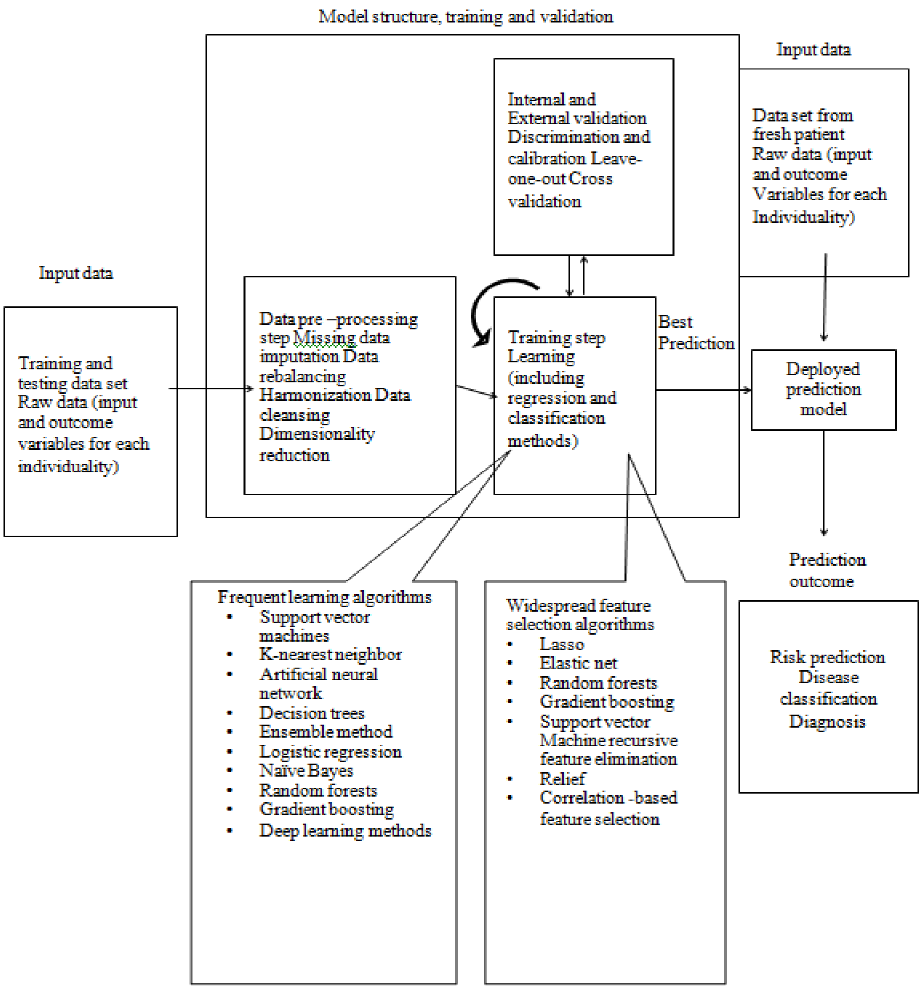

Figure 1: The General Structure for Clinical Prediction Modelling

Source Adopted From Jamshidi et al. (2019)

In the development of the prediction model, the ML method can also be utilized at various stages. Training dataset (original medical data from patients dataset with knee osteoarthritis) is applied to generate a model for both techniques (Jamshidi, Pelletier, and Martel-Pelletier 2019). Initially, the data are preprocessed to eliminate the unwanted data which are not used in the process (clean the data, remove the missing data (imputation of data), reduce the dimensionality, harmonization and rebalancing of data). Then, the best variables are selected by the feature selection technique, and this technique used the machine learning algorithm (Elastic net, Lasso, and random forest algorithm) to improve their efficiency.

Further, the training dataset can be integrated to the suggested model i.e., the model can learns the data from their training stage to find the patterns and match the input and their corresponding output. ML algorithms normally used for this measure includes support vector machines, k-nearest neighbours, artificial neural network, decision trees and ensemble methods. The model can be internally validated or externally validated. The resulting forecast model is then used to forecast the outcome (such as the risk, disease classification or diagnosis) of new input data (such as a new patient). Finally, the machine learning algorithms k-nearest neighbours, decision trees, support vector machines, artificial neural network, and ensemble methods) are applied to classify the test data as disease or diagnosis, risk or normal data. Then, the model is validated both internally and externally.

III.BRIEF OVERVIEW OF THE LITERATURE

The previous studies related to machine learning technique predictive modelling are discussed as follows:

A research by Abdar et al. (2019) applied ten conventional ML technique and performed three suitable SVM methods in medical application. Furthermore, they have combined the particle swarm optimization and genetic method for parallel selection of features and optimization of classifier parameters. The result of simulation shows that the developed model of N2Genetic-nuSVM offered the 93.08% of accuracy rate and 91.51% of F1-score. At the final stage of the research, the suggested technique is compared with some traditional methods such as Artificial Bee Colony (ABC), PSO and GA for feature selection and optimisation. From that, the suggested model offers better solutions than other convolutional techniques. Also test the algorithms with other technique like decision trees (e.g., C4.5, RF, and C5.0) and multilayer perceptron .

A research work by Bur et al. (2019) implemented and evaluated the ML algorithms to find the disease of pathologic lymph node metastasis with clinically node-negative oral cancer. The algorithms were utilized five clinical variables and pathologic variables. Further, these kinds of variables were applied to forecast the presence of disease (occult lymph node metastasis) and this suggested model trained with 782 patients data from NCDB. Based on tumour DOI, the machine learning approaches are consistently outperformed for respective predictive models. Furthermore, they have suggested the development of Machine Learning by high-quality multi-institutional data. It is required to develop the algorithm that can be applied clinically to prevent that patients with the occult nodal disease are effectively treated while eliminating the cost and the morbidity of neck dissection in patients without pathologic nodal disease.

A research by Shashikant and Chetankumar, (2019) implemented an ML technique, and the corresponding data were gathered from the research group of data science MITU Skillogies Pune, India. Moreover, they implement three predictive models with Heart Rate Variability (HRV) indices of 19 input features and two output classes. The developed model was evaluated with some performance metrics such as sensitivity, accuracy, F1 score, precision, Area under the curve (AUC) and specificity. Logistic regression of this model was attained 88.50% of accuracy, 91.79% of sensitivity, 83.11% of precision, 86.03% of specificity and 0.88 of AUC and 0.87 of F1-score. Similarly, the model of decision tree was achieved 92.59% of accuracy, 97.31% of specificity, 97.29% of precision, 90.11% of sensitivity, 0.93 of F1 score and 0.94 of AUC, the random forest model was also achieved 94.59% of precision, 95.03% of specificity, 93.61% of accuracy, 92.11% of sensitivity, 0.95 of AUC and 0.93 of F1 score. The logistic regression and decision tree were offered the less accurate result than RF model. In future, they have planned to enhance the accuracy by considering deep learning approach.

The disease prediction using traditional disease risk models typically includes an ML technique and particularly, the model was trained by supervised learning algorithm with training data and their corresponding labels(Lin, Luo, et al. 2016; Lin, Chen, et al. 2016). This kind of model was used in clinical environments (Marcoon et al. 2013). But still, the defects and characteristics were followed by these methods. The characteristics are selected by experience when the data set (patient health data) is small with some specific conditions (Bandyopadhyay et al. 2015). But, the characteristics of preselected may not satisfy their influencing factor and changes in diseases.

Table 2: Summary of Previous Studies

Author (Year)	Title	Medical condition	Machine learning algorithm used	The aim of the research	Important findings	Conclusion	Recommendation
Al Maadeed et al. (2017)	Multispectral imaging and images are learning automated cancer diagnosis.	Cancer	Round - Robin	Integrating image capturing techniques to help the ML	ML can only be improved by examining the largest population of image capturing	ML is today in nascent stages in the research world	Complex data acquisition

Matsumoto et al. (2015)	Forecasting the Radioprotectors Targeting p53 for Suppression of Acute Effect of Cancer Radiotherapy using ML	cancer	Random Forest and SVM	Examined the inhibitory activity of some zinc(II) chelators against radiation-induced apoptosis of MOLT-4 cells using machine learning techniques	Radioprotectors targeting p53 were predicted A method for separately predicting radiation protection function and toxicity for each compound using ML was performed	RF has high AVe score than SVM	If the quantity of data is increased, then the insights would be better which is a premise for future researchers
Leung et al. (2016)	Genomic medicine	Cancer	Learning algorithm	The important problems were addressed in genomic medicine.	ML can help to model the relationship between DNA and the quantities of key molecules in the cell.	Genomic medicine is a research area still in its nascent stages	Development of computational problems and data sets
Sankararaman (2016)	Introduction to genomics CM226: ML for Bioinformatics	Genetic disease	Needleman-Wunsch, Smith-Waterman, and alignment heuristics.	Answering biological questions using tools from computer science, Statistics and mathematics.	Studies in Genomics can help cure genetically inherited diseases.	ML can improve the understanding of Genomics	Study of a large population of data available on genomic diseases
Jordan and Radhakrishnan (2014)	ML Predictions of Cancer driver Mutations	Cancer	SIFT	Developing a method to forecast the activation status of kinase domain mutations	Support Vector mechanisms in cancer cure	A method to forecast whether cancer driver mutations are driver mutation.	The use of SV mechanisms is a fairly reliable mechanism for forecasting the effect of kinase domain mutations
Kumar and Ramakrishnan (2014)	Binary classification of cancer microarray gene expression data using extreme learning machines	Cancer	Moore Penrose ELM algorithm	Usage of ELM for cancer microarray gene expression	ELM evaluated the binary classification of cancer microarray gene expression on five bench-marked datasets	An assessment of classification accuracies of ELM and other conventional classifiers are presented.	Further research on ELM is Recommended by practically time consuming

Matsumoto et al. (2015)	Forecasing the Radio Protectors Targeting p53 for Suppression of Acute Effect of Cancer Radiotherapy using ML	Cancer	SVM	Examined the inhibitory activity of zinc(II) chela tors against radiation	protected mice from acute lethality due to the hematopoietic syndrome, indicating that pharmacologically temporary suppression of p53 effectively minimize the radiation smash up	predicted radio protectors targeting p53	We expect to become a better result if the number of data is increased
Papachristou et al. (2016)	Comparing M L Clustering with Latent Class Analysis on Cancer Symptoms Data	Cancer	Clustering	reported a secondary analysis on cancer symptoms data, and compared the performance of five ML clustering algorithms	Identified analytical method to process all the available data	Investigated the performance of readily ML algorithms in relation to common practice statistical methods, such as LCA	Research will benefit from exploring all the multivariate versions of these datasets, taking the domain a step forward

IV. MATERIALS AND METHODS

A search of literature papers was performed in databases such as IEEE Explore, Google Scholar, and other renowned international journals on medical applications of machine learning. The year range for the acquisition of research articles was limited within 2012-2018. Moreover, we have manually searched reference section of specific article which is mentioned in the review articles via researchers. Few of the researcher reported the study with ML technique in detailed manner on clinical predictive modelling and diagnosis of different disease were incorporated. In this manner we have selected 43 appropriate articles that offered suitable results with related to the clinical predictive modelling of Machine Learning, with particular reference to different Cancerous malignancies.

Inclusion and Exclusion Criteria

The study inclusion and exclusion criteria have been segregated based on the below standards:

- ✓ We have considered the study which comprises the development of a predictive or diagnostic model based on ML or artificial intelligence.
- ✓ A study which contains a measure of disease severity, relation amongst immune biomarker and different bronchiolitis severity groups then we have included.
- ✓ Papers with Algorithms and empirical data to be given preference Peer-reviewed articles which includes editorials, tutorial summaries, panel discussions, keynotes, reviews, and position papers were excluded since they are more presentational in nature
- ✓ Studies related to finance and market data by different methods like review and analysis, case study and experimental research are excluded.
- ✓ Augmented articles- If two articles from comparable investigation on a related subject were distributed in different data presentation scenarios (e.g., conference, journal- international or national level), where we added only the journal article, since conference data are not very data specific and more informational in nature.
- ✓ All copied reviews initiated from various resources were identified and neglected

Search Strategy and Selection Criteria

The search was conducted using articles in science direct, IEEE Explore, web of science and Scopus Digital library returned 188 outcomes. By evaluating the research title, aim and appropriate data present in the study, we have separated the articles. At the point when there were a few articles may possibly not choose by scrutinizing the themes and conceptual, and thus these articles were held for the accompanying round of evaluation. We banned the articles that stayed random, or whose full substance was not assessable or available on the web. Since we were enthusiastic about exploratory surveys, we rejected papers that were not populated with observational information. We encircled our pursuit string by means of the guidelines, as demonstrated as follows. We consolidated a condition for picking careful audits in our pursuit string in the midst of the study method. "Clinical predictive modelling" AND "knowledge discovery process" AND "data management technique" OR "machine learning technique" OR "medical application" OR "disease predication" OR "diagnosis & treatment". But, the search was restricted mainly to IEEE Explore and science direct, both of which are recognized worldwide for comprehensive public access to research papers.

V. DISCUSSION

From the review of literature, different ML techniques have been applied to study various diseases, such as liver disease (Moloud Abdar, Yen, and Hung 2017), breast cancer (Shukla et al. 2018), Parkinson's disease (M Abdar 2018), lung cancer (Yang and Chen 2015) and heart disease (Rajesh and Dhuli 2018).

A research by Abdar et al.

(2019) applied ML technique and performed three suitable SVM methods in medical application. Also, they suggested, in future, adding ABC algorithm for parameter optimization and feature selection could provide better solution. A work by Bur et al. (2019) suggested a ML algorithms toward predicting the pathologic lymph node metastasis in patients with clinically node-negative oral cancer. But it required to develop the algorithm, that can be applied clinically to prevent that patients with the occult nodal disease are effectively treated while eliminating the cost and the morbidity of neck dissection in patients without pathologic nodal disease. A research by Shashikant and Chetankumar, (2019) implemented an ML technique and the corresponding data were gathered from research group of data science MITU Skillogies Pune, India.. At the end of end research they suggested as to enhance the accuracy by considering deep learning approach. But still, the defects and characteristics were followed by these methods. The characteristics are selected by experience, when the data set (patient health data) is small with some specific conditions (Bandyopadhyay et al. 2015). But, the characteristics of preselected may not be satisfy their influencing factor and changes in diseases.

In overall, the earlier developed techniques were offered the sufficient outcome for identifying the corresponding diseases. Appropriate analysis of these diseases, including cancer with slightest probable error, is the primary prospect of patients (Moloud Abdar et al. 2019). The experience and knowledge are demanded by the detection process. The integration of data mining and ML algorithms are enhancing the accuracy rate, reduce the number of diagnosis error and also provide the quality of service to the patients. However, the rapid growth of computing technologies and computational improvements in hardware have made AI strategies progressively open to medicinal services results in experts and medical organization. This article requires the code to implement the entire process for illustration, the majority of estimators portrayed with 100 lines, and it takes 30 minutes to run the coding part. Here the 100000 individual data are used with standard system (PC).

Since AI or ML techniques are progressively embraced for human services results investigate, we offer three of exhortation, following rules depicted from the literature (Luo et al. 2016). Firstly, an estimator ought to give an answer for a prespecified issue as opposed to just recognizing relationship in a huge dataset. Prespecifying the issue being tended to, including thoughts of accomplishment, may help minimize the ratio of false-positive rate (Luo et al. 2016). Furthermore, prespecifying the measurements for correlation of estimators can support to avoid the false ratio of data classification. Secondly, the group of observers planned to use the estimator should be considered. Regardless whether the group of spectators needs to comprehend what highlights create the estimator's expectations, or basically have the option to apply it to future datasets through a real-time or automated application, for instance the backend of a medicinal record ought to be resolved. It will affect the estimator's choices due to its complexity of understanding with the technical terminologies for instances, and the deep learners made estimation from the available information than the others. Some accuracy metrics are utilized to estimate the performance of the system model and it will be conflict with various goals like disease prediction of human understanding. There is no rule of thumb for how much performance enhancement its satisfactory to justify using less interpretable estimators. Regularly, a superior performing model as far as some precision metric will be in strife with different objectives, for example, the human comprehension of forecasts. There is no standard guideline for how much execution development is adequate to legitimize utilizing less interpretable estimators. We recommend that specialists have clear support for picking an estimator. It is essential to have "information compassion," which denotes to the possibility that regardless of how complex the strategy, a dataset that is poor in quality or inadequately instructive for a known enquiry won't be valuable, regardless of whether enormous in size (El-Sayed and Galea 2017). Henceforth, estimation and choice inclinations apply to AI techniques as much as to some other types of auxiliary information investigation (Crown 2015). At last, rising AI techniques are conceivably helpful for the medicinal services results scientist if prediction is a significant and important undertaking. Prediction can be joined with causal research to progress our understanding (Fernandez 2019) .

A. Conclusion:

From this article, we mean to bring down the hindrances to actualizing Artificial Intelligence techniques. These assets will enable analysts to find out about extra models like Kernel approach, Naïve Bayes, Deep Learning and so on. As Machine Learning techniques advances, and the newer area of ML uses multilayered ANN to deliver high accuracy in tasks. we contend that the standards for good practice audited in this preliminary will probably serve wellbeing model in the field of healthcare services (Doupe, Faghmous, and Basu 2019).

VI.FINDINGS AND CONCLUSION

This review study seeks to serve as a reference guide by provided an overview of previous studies as regards the application of machine learning in medical application. The findings of our study are summarized as follows:

- The clinical predictive modelling and knowledge discovery process healthcare has been investigated by researchers, as observed by the large number of studies published in the past decade.

The various pre-processing and machine learning methods and models such as Supervised Learning, Unsupervised Learning and Reinforcement Learning were investigated then recommends the solution for future study.

Also discussed the significance of predictive modelling using ML technique in healthcare industry, it will also evaluate a choice of scenarios in the use of predictive analytics with a particular focus on provision liberation within health care.

The recommendations for the future study are:

- ✓ To extend the machine learning approaches, this will comprise multiple attributes information that come from multiple sources, including domain adaptation.
- ✓ To authenticate the use of ML tools as a clinical prediction model to forecast drug candidates for different diseases. Also, will improve the prediction accuracy in a target task
- ✓ To identify the effective solution for missing data, data cleansing, data imputation, dimensionality reduction, rebalancing and harmonization of the data.
- ✓ To suggest an effective deep learning approach for solving the clinical prediction issues while combining the structured and unstructured data in the healthcare field to assess the risk of disease.

REFERENCES

1. Abdar, M. 2018. Impact of Patients ' Gender on Parkinson ' S Disease Using Classification Algorithms. Journal of AI and Data Mining 6, no. 2: 277–285.
2. Abdar, M., W. Książek, U.R. Acharya, R.-S. Tan, V. Makarenkov, and P. Pławiak. 2019. A New Machine Learning Technique for an Accurate Diagnosis of Coronary Artery Disease. Computer Methods and Programs in Biomedicine 179 .
3. Abdar, M., N.Y. Yen, and J.C.-S. Hung. 2017. Improving the Diagnosis of Liver Disease Using Multilayer Perceptron Neural Network and Boosted Decision Trees. Journal of Medical and Biological Engineering 38, no. 6 (December 7): 953–965.
4. Bandyopadhyay, S., J. Wolfson, D.M. Vock, G. Vazquez-Benitez, G. Adomavicius, M. Elidrisi, P.E. Johnson, and P.J. O'Connor. 2015. Data Mining for Censored Time-to-Event Data: A Bayesian Network Model for Predicting Cardiovascular Risk from Electronic Health Record Data. Data Mining and Knowledge Discovery 29, no. 4 (July 4): 1033–1069.
5. Bur, A.M., A. Holcomb, S. Goodwin, J. Woodroof, O. Karadaghy, Y. Shnayder, K. Kakarala, J. Brant, and M. Shew. 2019. Machine Learning to Predict Occult Nodal Metastasis in Early Oral Squamous Cell Carcinoma. Oral Oncology 92 (May): 20–25.
6. Crown, W.H. 2015. Potential Application of Machine Learning in Health Outcomes Research and Some Statistical Cautions. Value in Health 18, no. 2 (March): 137–140.
7. Darcy, A.M., A.K. Louie, and L.W. Roberts. 2016. Machine Learning and the Profession of Medicine. JAMA 315, no. 6 (February 9): 551.
8. Doupe, P., J. Faghmous, and S. Basu. 2019. Machine Learning for Health Services Researchers. Value in Health 22, no. 7 (July): 808–815.
9. Eissa, M.M., M. Elmogy, and M. Hashem. 2016. Rough – Granular Computing Knowledge Discovery Models for Medical Classification. Egyptian Informatics Journal 17, no. 3 (November): 265–272.
10. El-Sayed, A.M., and S. Galea. 2017. Systems Science and Population Health. Ed. Abdulrahman M. El-Sayed and Sandro Galea. Oxford University Press.
11. Fernandez, M.A.L. 2019. Targeted Maximum Likelihood Estimation for a Binary Outcome: Tutorial and Guided Implementation. Migariane. https://migariane.github.io/TMLE.nb.html.
12. Hastie, T., R. Tibshirani, and J. Friedman. 2009. The Elements of Statistical Learning. Revista Espanola de Las Enfermedades Del Aparato Digestivo. II. Vol. 26. Springer Series in Statistics. New York, NY: Springer New York.
13. Jamshidi, A., J.-P. Pelletier, and J. Martel-Pelletier. 2019. Machine-Learning-Based Patient-Specific Prediction Models for Knee Osteoarthritis. Nature Reviews Rheumatology 15, no. 1 (January 6): 49–60.
14. Jordan, E.J., and R. Radhakrishnan. 2014. Machine Learning Predictions of Cancer Driver Mutations. In Proceedings of the 2014 6th International Advanced Research Workshop on In Silico Oncology and Cancer Investigation - The CHIC Project Workshop (IARWISOCI), 1–4. IEEE.
15. Kaur, H., and V. Kumari. 2018. Predictive Modelling and Analytics for Diabetes Using a Machine Learning Approach. Applied Computing and Informatics.
16. Kaur, H., E. Lechman, and A. Marszk. 2017. Catalyzing Development through ICT Adoption. Ed. Harleen Kaur, Ewa Lechman, and Adam Marszk. Cham: Springer International Publishing.
17. Ker, J., L. Wang, J. Rao, and T. Lim. 2018. Deep Learning Applications in Medical Image Analysis. IEEE Access 6: 9375–9389.
18. Kumar, C.A., and S. Ramakrishnan. 2014. Binary Classification of Cancer Microarray Gene Expression Data Using Extreme Learning Machines. In 2014 IEEE International Conference on Computational Intelligence and Computing Research, 1–4. IEEE.
19. Leung, M.K.K., A. Delong, B. Alipanahi, and B.J. Frey. 2016. Machine Learning in Genomic Medicine: A Review of Computational Problems and Data Sets. Proceedings of the IEEE 104, no. 1 (January): 176–197.
20. Lin, K., M. Chen, J. Deng, M.M. Hassan, and G. Fortino. 2016. Enhanced Fingerprinting and Trajectory Prediction for IoT Localization in Smart Buildings. IEEE Transactions on Automation Science and Engineering 13, no. 3 (July): 1294–1307.
21. Lin, K., J. Luo, L. Hu, M.S. Hossain, and A. Ghoneim. 2016. Localization Based on Social Big Data Analysis in the Vehicular Networks. IEEE Transactions on Industrial Informatics 13, no. 4 (August): 1932–1940.
22. Luo, W., D. Phung, T. Tran, S. Gupta, S. Rana, C. Karmakar, A. Shilton, et al. 2016. Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research: A Multidisciplinary View. Journal of Medical Internet Research 18, no. 12 (December 16): e323.
23. Ma, Z., J. Ma, Y. Miao, and X. Liu. 2019. Privacy-Preserving and High-Accurate Outsourced Disease Predictor on Random Forest. Information Sciences 496 (September): 225–241.
24. Al Maadeed, S., S. Kunhoth, A. Bouridane, and R. Peyret. 2017. Multispectral Imaging and Machine Learning for Automated Cancer Diagnosis. In 2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC), 1740–1744.
25. Marcoon, S., A.M. Chang, B. Lee, R. Salhi, and J.E. Hollander. 2013. HEART Score to Further Risk Stratify Patients With Low TIMI Scores. Critical Pathways in Cardiology 12, no. 1 (March): 1–5.
26. Mateos-Pérez, J.M., M. Dadar, M. Lacalle-Aurioles, Y. Iturria-Medina, Y. Zeighami, and A.C. Evans. 2018. Structural Neuroimaging as Clinical Predictor: A Review of Machine Learning Applications. NeuroImage: Clinical 20: 506–522. +
27. Matsumoto, A., T. Ito, Y. Nishi, T. Teraoka, S. Aoki, and H. Ohwada. 2015. Prediction of Radioprotectors Targeting p53 for Suppression of Acute Effect of Cancer Radiotherapy Using Machine Learning. In 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 1725–1727. IEEE. http://ieeexplore.ieee.org/document/7359941/.
28. Mohri, M., A. Rostamizadeh, and A. Talwalkar. 2012. Foundations of Machine Learning. https://dl.acm.org/citation.cfm?id=2371238.
29. Nithya, B., and V. Ilango. 2017. Predictive Analytics in Health Care Using Machine Learning Tools and Techniques. In 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), 492–499. IEEE. http://ieeexplore.ieee.org/document/8250771/.
30. Papachristou, N., C. Miaskowski, P. Barnaghi, R. Maguire, N. Farajidavar, B. Cooper, and X. Hu. 2016. Comparing Machine Learning Clustering with Latent Class Analysis on Cancer Symptoms' Data. In 2016 IEEE Healthcare Innovation Point-Of-Care Technologies Conference (HI-POCT), 162–166. IEEE. http://ieeexplore.ieee.org/document/7797722/.
31. Parikh, R.B., M. Kakad, and D.W. Bates. 2016. Integrating Predictive Analytics Into High-Value Care. JAMA 315, no. 7 (February 16): 651. http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.2015.19417.
32. Rajesh, K.N.V.P.S., and R. Dhuli. 2018. Classification of Imbalanced ECG Beats Using Re-Sampling Techniques and AdaBoost Ensemble Classifier. Biomedical Signal Processing and Control 41 (March): 242–254. https://linkinghub.elsevier.com/retrieve/pii/S1746809417302872.
33. Sankararaman, S. 2016. CM226: Machine Learning for Bioinformatics. Introduction to Genomics. http://web.cs.ucla.edu/~sriram/courses/cm226.fall-2016/slides/intro.pdf.
34. Shashikant, R., and P. Chetankumar. 2019. Predictive Model of Cardiac Arrest in Smokers Using Machine Learning Technique Based on Heart Rate Variability Parameter. Applied Computing and Informatics (June).

https://linkinghub.elsevier.com/retrieve/pii/S2210832719301048.

35. Shukla, N., M. Hagenbuchner, K.T. Win, and J. Yang. 2018. Breast Cancer Data Analysis for Survivability Studies and Prediction. Computer Methods and Programs in Biomedicine 155 (March): 199–208.
36. Sun, S. 2013. A Survey of Multi-View Machine Learning. Neural Computing and Applications 23, no. 7–8 (December 17): 2031–2038. http://link.springer.com/10.1007/s00521-013-1362-6.
37. Szűcs, G. 2013. Random Response Forest for Privacy-Preserving Classification. Journal of Computational Engineering 2013: 1–6.
38. Turki, T., and Y. Taguchi. 2019. Machine Learning Algorithms for Predicting Drugs–tissues Relationships. Expert Systems with Applications 127 (August): 167–186.
39. Vaidya, J., B. Shafiq, W. Fan, D. Mehmood, and D. Lorenzi. 2014. A Random Decision Tree Framework for Privacy-Preserving Data Mining. IEEE Transactions on Dependable and Secure Computing 11, no. 5 (September): 399–411. http://ieeexplore.ieee.org/document/6616536/.
40. Walczak, S., and V. Velanovich. 2017. An Evaluation of Artificial Neural Networks in Predicting Pancreatic Cancer Survival. Gastroenterology 152, no. 5 (April): S1248. https://linkinghub.elsevier.com/retrieve/pii/S0016508517341574.
41. Wang, X.A., F. Xhafa, J. Ma, and Z. Zheng. 2019. Controlled Secure Social Cloud Data Sharing Based on a Novel Identity Based Proxy Re-Encryption plus Scheme. Journal of Parallel and Distributed Computing 130 (August): 153–165.
42. Yang, H., and Y.-P.P. Chen. 2015. Data Mining in Lung Cancer Pathologic Staging Diagnosis: Correlation between Clinical and Pathology Information. Expert Systems with Applications 42, no. 15–16 (September): 6168–6176. https://linkinghub.elsevier.com/retrieve/pii/S0957417415002134.
43. Zamani Esfahlani, F., K. Visser, G.P. Strauss, and H. Sayama. 2018. A Network-Based Classification Framework for Predicting Treatment Response of Schizophrenia Patients. Expert Systems with Applications 109 (November): 152–161. https://linkinghub.elsevier.com/retrieve/pii/S0957417418302823.

AUTHORS PROFILE

Dr. A. Radhika, Assistant Professor from the Department of Statistics, Periyar University, Salem. Area of Specialization is Biostatistics, Clinical Trials, Design of Experiments.

G. Priya, Research Scholar from the Department of Statistics, Periyar University, Salem