iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VIII **Month of publication:** August 2021

DOI: https://doi.org/10.22214/ijraset.2021.37836

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VIII Aug 2021- Available at www.ijraset.com

Systematic Review of the Literature on Machine Learning Techniques Employed in Real-World Data Analysis for Patient-Provider Decision Making

Kushagra Singh Sisodiya[1], Ayushi Dubey[2]
[1]*Computer Science and Engineering, Dr. A.P.J. Abdul Kalam University, Indore*
[2]*Information Technology, Lakshmi Narain College of Technology and Science(LNCTS), Bhopal*

Abstract: ***The Industrial Revolution 4.0 has flooded the virtual world with data, which includes Internet of Things (IoT) data, mobile data, cybersecurity data, business data, social networks, including health data. To analyse this data efficiently and create related efficient and streamlined applications, expertise in artificial intelligence specifically machine learning (ML), is required. This field makes use of a variety of machine learning methods, including supervised, unsupervised, semi-supervised, and reinforcement. Additionally, deep learning, which is a subset of a larger range of machine learning techniques, is capable of effectively analysing vast amounts of data. Machine learning is a broad term that encompasses a number of methods used to extract information from data. These methods may allow the rapid translation of massive real-world information into applications that assist patients and providers in making decisions. The objective of this literature review was to find observational studies that utilised machine learning to enhance patient-provider decision-making utilising secondary data.***
Keywords: Machine Learning, Real World, Patient, Population, Artificial Intelligence

I. INTRODUCTION

Formerly, methods for assessing massive real-world sets of data (big data) as well as other systematic reviews have been centred on results that might be utilised to educate the public. Whereas real-world results generally applicable to entire communities, the capacity to forecast or offer credible evidence there at patient level is much less proven, owing to the complexity of clinical practice and the diversity of variables considered by the healthcare professional. It is very impossible to forecast the prognosis of any particular patient correctly using conventional techniques, that give population estimates and measurements of variability, even more so when subgroup studies are included. Patient care is complicated, and decision-making should take a number of non-linear, linked factors into account. When only population-level details are collected, health care decision-makers are left in the dark about the optimal course of therapy for a certain patient.

II. DISCUSSION

Systems for clinical forecasting include a technique for using patient-level data to help healthcare stakeholders in determining patient care choices. For decades, health care practitioners have relied on these models, sometimes known as prediction rules and prognostic models. Originally, such models incorporated demographics, medical, as well as therapy data of patients into a graphical and numerical model, most frequently regression, classification, or neural networks, although they have just a few predictor variables. The Framingham Heart Research set the precedent for using longitudinal data to develop a framework of conventional decision-making. Risk calculators as well as estimators are being created to aid doctors assess a sufferer's risk of having CVDs such as atrial fibrillation or cardiovascular disease. A multivariable regression model is typically used in this research in order to analyse risk variables that have been identified in the literature. In order to forecast the probability of an unfavourable result based on the sufferer's evaluation across all risk variables assessed, these data are used to construct a scoring system for each element.

Patients receiving regular clinical treatment now have access to data sets that are easier to gather and analyse prospective predictors (such as genomic data) could really surpass the thousands and thousands, requiring the development of new strategies for quickly data - intensive applications. A growing number of clinical researchers are turning to AI, especially machine learning techniques (a subtype of AI), to create predictive model, pattern matching methods, including deep learning techniques in order to improve patient care for integrating complex data, such as genetic and clinical data. These methods are utilised in the health care professions to perform tasks which would otherwise require significant time and expertise and virtually definitely result in a possible error. The underlying concept is that a machine would learn to make predictions without a predetermined body of norms via trial and error using just the data itself. Simply stated, machine learning is the process of "gathering and analysing data" for the goal of decision-making.

A. Data-Driven Learning May Be Approached In Two Ways

1) Unsupervised

2) Supervised

To form opinions from datasets that contains input data without labels. The technique of unsupervised learning is a kind of machine learning. In unsupervised learning, cluster analysis is by far the most often used method of learning. Use it to discover hidden patterns or groups in your data using exploratory data analysis (EDA). A predetermined collection of inputs and outputs is used to generate predictions in supervised learning. For supervised learning, a wide range of statistical methods are accessible. A regression model, such as "regression splines", "projection pursuit regression "(PPR), or penalised regression (PR), necessitates adapting a framework to the information and estimating parameters that are subsequently employed in the predicted values.

Another option is a method that partitions a data set sequentially based on the correlations between predictor and result variables (e.g., classification and regression trees [CART]). In addition, neural networks, discriminant functions, including linear classifiers, and also support vector classifiers or machines, are examples of artificial intelligence techniques.

There are several different kinds of models aggregation (or ensemble learning) that are used to create prediction tools. It is possible to utilise model averaging to fit a variety of different types of patterns to the exact same data.

Statisticians and also the science community that utilises them are well-versed in conventional statistical regression methods. However, they tend to overlook complex relationships and are restricted in flexibility when examining a large set of samples. When using traditional regression modelling, choosing the "right" model may be a difficult process as well. Traditional regression models in the era of big data have certain limitations that non-traditional machine learning algorithms & techniques can overcome, but they don't give a complete solution since the algorithms and methods must be viewed in relation to the data that was used in the research

It's important to note that although machine learning techniques are applied towards both population models & informed patient-provider decision - making process, it's important to emphasise that maybe the information, model, or outcomes used to inform a person nursing needs should maintain the highest level research quality standards, since as a decision made will almost certainly have an impact on both long-term and short-term patient outcomes. Population-based estimations are subject to some ambiguity, but patient-level models should be kept as accurate as possible in order to provide high-quality patient treatment.

III. APPROACHES TO MACHINE LEARNING IN GENERAL

There were 12 studies that incorporated decision trees/random forest analyses with neural networks. Additionally, we examined the latent growth mixture model; support vector machine classifiers; LASSO regression; and new Bayesian techniques The Akaike Information, the Bayesian Information Criterion, as well as the Lo-Mendel-Rubin likelihood ratio test are just a few of the many techniques for evaluating model fit that have been employed in analytical approaches to support machine learning. In addition to the AUC of receiver-operator characteristic curves (ROC), sensitivity assessments were also conducted. The geometric mean, the Matthews correlation coefficient (which runs from -1.0, indicating totally incorrect information, to +1.0, indicating flawless prediction), the usage of a confusion matrix to identify true/false negatives/positives, determining the root mean square error between the anticipated and original result profiles, or finding the classifier a priori information.

IV. PACKAGES OF STATISTICAL SOFTWARE

Machine learning methods used in these studies varied considerably, and no consistency was discovered. As previously stated, one research that utilized decision tree analysis utilized Quinlan's C5.0 decision tree method, while another utilized an older version of same software (C4.5). Various versions of R were used in other decision tree studies. IBM's SPSS Software Sciences (SPSS), Microsoft's Azure Machine Learning Platform, or Python were used to design the model. Artificial neural network studies were conducted using Neural Designer or Statistical V10. Six research withheld information on the software used to conduct the analyses.

V. THE SCHOLARS CONSIDER BOTH WEAKNESSES AND STRENGTHS

Numerous articles assessed the relative merits and demerits of a machine learning methods employed. Machine-based learning techniques have been lauded for their simplicity and low complexity. The use of machine learning methods to large datasets was both successful and efficient. It was noted that variables that were significant only at patient level were included in this study, even if they would not be significant at the population level using traditional regression analysis model building. According to one publication, machine learning's effectiveness is highly dependent on the model selection technique & parameter optimization, and therefore that machine learning alone would not result in better predictions unless these steps are done properly.

Even when properly constructed, machine learning approaches may have limitations which should be considered in future research utilising these techniques. Among the qualifying papers, model overfitting and an excessive amount of information were identified as weaknesses. Additionally, limitations imposed by the data sources to use for machine learning, like the lack of all necessary variables and incomplete information, may hamper the performance and development of these models.

VI. COMMENT

A thorough assessment and review of the situation was the goal of this research regarding machine learning approaches sources of secondary methods, data, and methodologies that might be used to aid decision-making between the patient and the provider ,as a result of the application of machine learning techniques to individual decision making much more often than to observational studies, and hence the explanation of this research doesn't really apply equally to any and all machine learning-based research. Numerous articles address the disadvantages of making individual decisions using population-based forecasts. To be more precise, a populace summary statistic does not really relate to any individual of that cohort. Population projections are a point on a possibly widespread reach, as well as any individual patient can fall along any that allocation and be significantly different from the median estimate value.

Throughout the papers found, a consistent modelling approach was often utilised. It has long been established that using a single estimate technique may introduce significant uncertainty. Multiple methods and replications of the produced models are needed to overcome this limitation. This, along with the recent progress of more complex analytics, new guideline for choosing and creating ML algorithms has been established In certain instances, a single model could be able to match the data and give an appropriate response, new techniques such as model averaging may be employed to improve the model's confidence. Numerous research considered in this review used iterative modelling with multiple families of modelling techniques. For future research utilising ML-based models, this should be a standard practise.

To guarantee model correctness, external validation is necessary; however it was rarely performed in the papers that are included in the study. This may be due to a number of factors, including a scarcity of appropriate datasets or even a lack of knowledge about the essential importance of external validation. External validation testing models is needed prior to their usage in any patient-provider situation as using machine learning in model development increases. Without this data, the generalisation of models cannot be determined. External validation also was omitted from publications that did not contain it, because generalisation was addressed in just a few studies, including one that had an external validation element.

When combined with appropriate response variable stratification, k-fold testing may be utilised as part of model selection process, according to a single research. The majority of the research included in this evaluation improved the confirmability by five or 10 fold. The research did not assess the validity of real-world data utilised to design, test, and verify the algorithms. Researchers should be aware of the following, despite the fact that it is not specifically addressed in this study:

Regardless of the method employed, the limitations inherent in real-world data sources persist. However, while using observation - based sets of data for machine learning approaches, the investigator ought to be aware of the consequences of the methods used being dependent on the data structure as well as accessibility, and therefore should consider carefully a proposed source of data to ensure it is appropriate for said machine learning project.

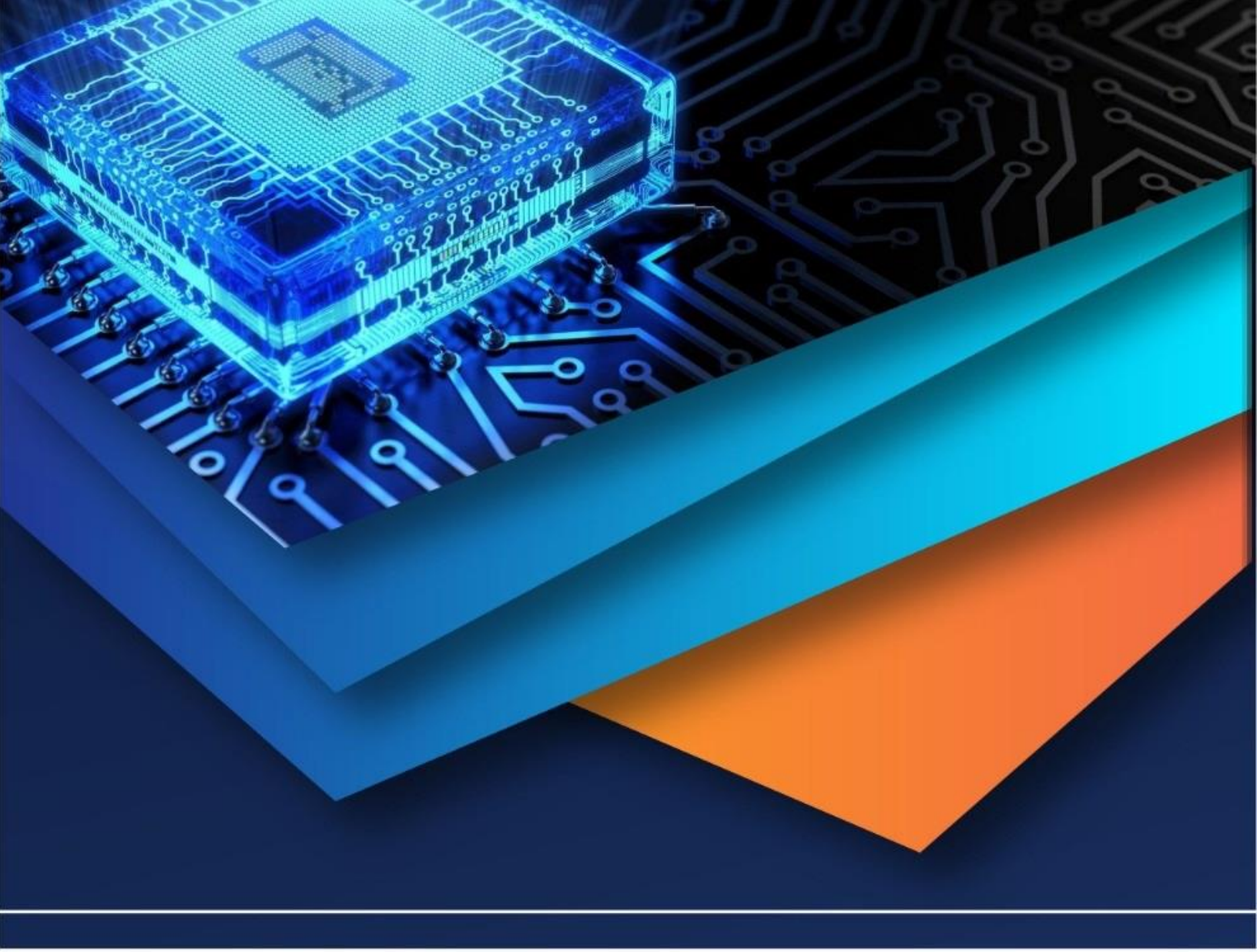
VII. CONCLUSIONS

This study discovered a diverse range of methodologies, methods, statistical software, including validation procedures used in the use of ML approaches to teach patient-provider decision-making through secondary resources. According to some resources, it is critical to incorporate a wide range of modelling techniques so if developing machine teaching models for patient care. These models must also adhere to high research rules in order to reliably enable able to share proof decision making by many health care providers. To be used to guide patient care, models must first be assessed against well-defined selection criteria and then validated both internally and externally. Just a few researches have reported the level of evidence required to help patients and providers in making healthcare decisions.

REFERENCES

[1] Steyerberg EW, Claggett B. Towards personalized therapy for multiple sclerosis: limitations of observational data. Brain. 2018;141(5):e38-e.
[2] Fröhlich H, Balling R, Beerenwinkel N, Kohlbacher O, Kumar S, Lengauer T, et al. From hype to reality: data science enabling personalized medicine. BMC Med. 2018;16(1):150.
[3] Steyerberg EW. Clinical prediction models. Berlin: Springer; 2019.
[4] Schnabel RB, Sullivan LM, Levy D, Pencina MJ, Massaro JM, D'Agostino RB Sr, et al. Development of a risk score for atrial fibrillation (Framing- ham Heart Study): a community-based cohort study. Lancet. 2009;373(9665):739–45.

[5] D'Agostino RB, Wolf PA, Belanger AJ, Kannel WB. Stroke risk profile: adjustment for antihypertensive medication. Framingham Study Stroke. 1994;25(1):40–3.
[6] Framingham Heart Study: Risk Functions 2020. https://www.framingham heartstudy.org/ .
[7] Gawehn E, Hiss JA, Schneider G. Deep learning in drug discovery. Mol Inf. 2016;35:3–14.
[8] Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, et al. Applications of machine learning in drug discovery and development. Nat Rev Drug Discov. 2019;18(6):463–77.
[9] Marcus G. Deep learning: A critical appraisal. arXiv preprint arXiv:180100631. 2018.
[10] Grote T, Berens P. On the ethics of algorithmic decision-making in health- care. J Med Ethics. 2020;46(3):205–11.
[11] Brnabic A, Hess L, Carter GC, Robinson R, Araujo A, Swindle R. Methods used for the applicability of real-world data sources to individual patient decision making. Value Health. 2018;21:S102.
[12] Fu H, Zhou J, Faries DE. Estimating optimal treatment regimens via subgroup identification in randomized control trials and observational studies. Stat Med. 2016;35(19):3285–302.
[13] Liang M, Ye T, Fu H. Estimating individualized optimal combination thera- pies through outcome weighted deep learning algorithms. Stat Med. 2018;37(27):3869–86.
[14] Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. J Med Internet Res. 2016;18(12):e323.
[15] Toussi M, Lamy J-B, Le Toumelin P, Venot A. Using data mining techniques to explore physicians' therapeutic decisions when clinical guidelines do not provide recommendations: methods and example for type 2 diabe- tes. BMC Med Inform Decis Mak. 2009;9(1):28.
[16] Ramezankhani A, Hadavandi E, Pournik O, Shahrabi J, Azizi F, Hadaegh F. Decision tree-based modelling for identification of potential interactions between type 2 diabetes risk factors: a decade follow-up in a Middle East prospective cohort study. BMJ Open. 2016;6(12):e013336.
[17] Pei D, Zhang C, Quan Y, Guo Q. Identification of potential type II diabetes in a Chinese population with a sensitive decision tree approach. J Diabe- tes Res. 2019;2019:4248218.
[18] Neefjes EC, van der Vorst MJ, Verdegaal BA, Beekman AT, Berkhof J, Verheul HM. Identification of patients with cancer with a high risk to develop delirium. Cancer Med. 2017;6(8):1861–70.
[19] Mubeen AM, Asaei A, Bachman AH, Sidtis JJ, Ardekani BA, Initiative AsDN. A six-month longitudinal evaluation significantly improves accuracy of predicting incipient Alzheimer's disease in mild cognitive impairment. J Neuroradiol. 2017;44(6):381–7.
[20] Hische M, Luis-Dominguez O, Pfeiffer AF, Schwarz PE, Selbig J, Spranger J. Decision trees as a simple-to-use and reliable tool to identify individu- als with impaired glucose metabolism or type 2 diabetes mellitus. Eur J Endocrinol. 2010;163(4):565.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY