APPLICATION OF MACHINE LEARNING IN DATA ANALYSIS IN HOSPITALS FOR DECISION MAKING: A SYSTEMATIC LITERATURE REVIEW

Bundi D.
School of Science and Technology, Department of Computing, United States International University Africa, Kenya.

International Academic Journal of Information Sciences and Project Management (IAJISPM) | ISSN 2519-7711

Received: 22nd January 2023

Published: 6th February 2023

Full Length Research

Available Online at: https://iajournals.org/articles/iajispm_v3_i7_120_130.pdf

Citation: Bundi, D. (2023). Application of machine learning in data analysis in hospitals for decision making: A systematic literature review. *International Academic Journal of Information Sciences and Project Management, 3*(6), 120-130.

ABSTRACT

There are several ways to learn from data using machine learning, which is a wide word. Using these technologies, big real-world datasets might be quickly converted into apps that help patients and healthcare providers make better decisions. Patient-provider-level decision-making was intended to be informed by published observational studies on the use of machine learning. Two reviewers independently assessed papers that met the qualifying criteria after implementing the search technique. Different statistical programs and procedures were employed in the selected investigations. A decision tree and a random forest were the two most frequent techniques. Less than 1% of the research used external validation, while the majority used internal validation. Only eight research used more than one machine learning method to analyze the data. In the application of machine learning techniques to patient-provider decision making, a broad range of methodologies, algorithms, statistical software, and validation procedures were used. Multiple machine learning methodologies must be employed, the model selection process must be clearly specified, and both internal and external validation is required to guarantee that choices for patient care are based on the most accurate evidence possible.

Key words: Machine learning, Decision making, Decision tree, Random Forest, automated neural network.

INTRODUCTION

The smart healthcare system has gotten a lot of attention recently, because to advancements in medical infrastructure (A. Grewal, M. Kaur, and J. H. Park, 2019). 'Smart healthcare' is a relatively new idea that refers to a system of guidelines that encompasses all four facets of patient care: prevention, diagnosis, treatment, and management. Smart medical systems can connect and share information at any time and from any location, unlike conventional medical systems (M. Kang, et al., 201). Smart healthcare differs from conventional medicine in that it is based on the principles of preventability, speed, and interconnectedness of data. Through the use of wireless networks and mobile devices, medical professionals may continually monitor, process, and evaluate critical medical occurrences (preventability). Every patient's medical history is readily available to doctors at any time, allowing them to construct an accurate diagnosis and treatment plan without delay (immediacy). To query about medical pictures and advice, medical workers may enter into the medical system from any location and get patient referral information via the medical network (interconnection of information).

In order to perform these activities, new digital technologies are needed. For transactions, BC adheres to strict privacy guidelines. For the administration of information systems, it is primarily utilized to ensure safe storage, transactions, process automation, and other applications (M. Tahir, et al., 2020). In healthcare, ML is the premier technology for complicated analysis, intelligent judgment, and innovative problem solution (W. Li, et al., 2021). Prior research on the use of digital technology in the smart healthcare sector has been focused on a single area or nation. This is changing. No research has been done to determine where these two medical innovations are right now. The association between authors, affiliations, keywords, and research hot spots has not been examined in any relevant study. It has become more important to combine the perspectives of researchers from a variety of disciplines in the study of smart healthcare during the last five years, in order to get a fuller understanding of its current condition. The goal of this study was to use bibliometric visualization to show how ML is being used in smart healthcare investigations. ML approache may be used in the healthcare industry in a thorough evaluation offered in this research. Researchers, journals, sponsors, and funding sources are all taken into account when we assess the current state of research in various nations, institutions, and fields. As an added bonus, this study presents a breakdown of the many ways the art has been used in the medical profession. Our findings will help medical professionals make the most of ML as a treatment option. Lastly, we look at the most recent research trends based on ML in order to provide researchers a research path to follow in the future.

BACKGROUND AND RELATED WORK

Clinical imaging

When it came to clinical data, the earliest applications of deep learning to image processing came from studies of brain MRI scans in order to forecast Alzheimer disease and its variants (Liu S, et al., 2014). A hierarchical representation of low-field knee MRI data was inferred using CNNs to automatically partition cartilage and predict the likelihood of osteoarthritis in different medical domains (Prasoon A, et al., 2013). Despite the fact that this technique relied on 2D photos, the results were superior to those of a more advanced strategy that used manually picked 3D multi-scale features. Multiple sclerosis lesions and breast nodules were also segmented using deep learning in multi-channel 3D MRI (Yoo Y, et al., 2013) as well as in ultrasound images. With the help of professional ophthalmologist annotations, Gulshan et al. (2016) employed CNNs to diagnose diabetic retinopathy in retinal fundus photos, achieving excellent sensitivity and specificity across about 10,000 test images. A large data set of 130 000 pictures (1942 biopsy-labeled test images) was used to evaluate the CNNs' ability to correctly categorize biopsy-proven clinical images of distinct forms of skin cancer (keratinocyte carcinomas vs benign seborrheic keratoses and malignant melanomas against benign nevi) (Esteva A, et al., 2011).

Genomics

In high-throughput biology, machine learning is used to capture the inherent structure of increasingly massive and high-dimensional datasets (e.g. DNA sequencing, RNA measurements). Improved performance over standard models, more interpretable data and new insights into the structure of biological data may be gained by using deep models. Neuronal networks in genomics were the first to replace traditional machine learning with deep structures without altering the input information. The splicing activity of individual exons was predicted by Xiong et al. (2015) using a fully connected feed-forward neural network. In order to train the algorithm, more than 1000 specified characteristics were retrieved from the candidate exon and its immediate surroundings. As opposed to simpler methods, our strategy was able to predict splicing activity more accurately and detect uncommon mutations associated with splicing misregulation using machine learning and artificial intelligence.

Mobile monitoring

A wide range of mobile applications, including health monitoring, are being transformed by sensor-equipped smartphones and wearables (Shameer K, et al., 2017). One wearable gadget can now keep track of a variety of medical risk indicators as the line between consumer health wearables and medical devices blurs. Patients might have direct access to personal analytics via these devices, which could improve their health, promote preventative treatment, and assist in the management of chronic illnesses (Piwek L, et al., 2016). The analysis of this new sort of data relies heavily on machine learning. However, because of hardware restrictions, only a few recent publications have exploited deep models in the health care sensing area. It is still a challenge to operate an efficient and reliable deep architecture on a mobile device to analyse noisy and complicated sensor data (Ravi D, et al., 2017). Hardware restrictions were the subject of many research. Lane and Georgiev (2015) suggested a low-power deep neural network inference engine that used both the CPU and DSP of the mobile device without causing any severe encumbrance on the hardware. Their DeepX software accelerator should reduce the device resources needed for deep learning, which is now a major roadblock to mobile adoption

RESEARCH METHODOLOGY

Research questions

The following questions were developed and used to analyse the existing research on machine learning in data analysis in hospitals for decision making.
RQ1: What is the application of machine learning in healthcare?
RQ2: what are the challenges of applying machine learning in healthcare?

Locating studies

Data analytics, big data, data mining, machine learning, and healthcare were the initial targets of the search strategy protocol, which included five single search terms, three platforms (EBSCOhost, ProQuest, and Scopus), the use of Boolean operators (AND/OR), a search of all fields, and two main exclusion criteria—only papers published by academic journals in English were considered. In order to discover the final list of relevant papers, this search approach was tried and adjusted several times. An initial step toward improving relevancy was adding more precise terms (such as "artificial intelligence," "business intelligence," and "internet of things"), more specific concepts (such as "healthcare," "business intelligence," and "machine learning," as well as the term "healthcare," due to the lack of standardization between the terms "healthcare" and "healthcare" in publications and academic texts). The Boolean expression was also used to abstractions instead of all fields or text to regulate the scope, which helped boost sensitivity. Finally, the search for publications took conference materials into account.

Study selection and evaluation

8,529 articles were discovered and reviewed for the following reasons: duplicates (16.4 percent), not linked to data science disciplines (29.4%), not entirely focused on healthcare systems (47.4%), and without an electronic file were excluded (0.4 percent). Only 576 papers (6.8 percent) from the original collection were approved as the final publication set for this study. These 576 papers were categorized into two sets: theoretical and applied publications, for the sake of this study. Over a period of five years, a total of 105 theoretical articles were included in this research, with the majority of them focusing on the advantages and disadvantages of applying data analytics and machine learning to healthcare systems as a whole. A different collection of 471 papers was devoted to the use of data science, data analytics, and machine learning algorithms to healthcare systems, each of which dealt with a distinct issue, illness, medical condition, or ailment.

Analysis and synthesis

The included papers were subjected to a quantitative and qualitative examination. An extensive review was conducted in accordance with the PICO Framework. This means that each article was evaluated in terms of journal topic, surgical domain, number and composition of cohorts, study timing, whether it was conducted retro- or prospectively; outcome focused on; ML technique applied; number of included predictor variables; method to compare ML with; results of comparison; strengths; weaknesses; predicted CDM effect. With 95 percent confidence intervals, reported AUROC values for ML were extracted and then compared to traditional techniques if appropriate. The best ML and traditional techniques were utilized in order to get the greatest overall results. R (R Foundation for Statistical Computing) and RStudio (RStudio, Vienna, Austria) were used to perform the statistical analyses (RStudio, Inc., Boston, USA).

RESULTS

Descriptive statistics

The search strategy was run and identified a total of 34 publications that utilized machine learning methods for individual patient-level decision making. The most common reason for study exclusion, as expected, was due to the study not meeting the patient-level decision making criterion. Most of the real-world data sources included retrospective databases or designs (79.4%), primarily utilizing electronic health records.

Results for research question 1

What is the application of machine learning in healthcare?

A patient's condition may be interpreted using ML techniques since they can detect related instances, establish normalcy statistics, and locate outliers in large datasets. They might help improve healthcare, regardless of whether they are used to diagnose or estimate risk. Some ML applications for clinical data interpretation are theoretically solid, but they do not address genuine clinical demands and concentrate on binary categorization of normal vs. abnormal (Madani A, et al., 2018), which severely restricts their utility in everyday clinical practice. Furthermore, studies that demonstrate an effect on objective clinical outcomes rather than on surrogate measures are still required. The way ahead is via the development of consensus suggestions on how to deal with a clinical requirement using ML and the deeper integration of technical and clinical contributions.

It is up to the clinician to decide whether to watch the patient and wait until an incident prompts the need for a decision, gather additional data to increase the probability of making the appropriate choice, undertake an intervention and monitor the result. Physicians may use cost-effective machine learning algorithms to determine the best course of action (Funkner AA, et al., 2017). The prediction power of ML approaches based on imaging has been evaluated in many research. When compared to standard prediction models, an echocardiography-based DL model proved to be more accurate in predicting in-hospital mortality in patients with coronary heart disease and heart failure (Kwon J, et al., 2019). It was found that an ensemble machine learning (ML) approach that examined SPECT myocardial perfusion studies performed better than an experienced reader in patients with suspected coronary artery disease, or in combination of the reading physicians at predicting major adverse cardiovascular events. One last study showed that automated systems may better predict cardiovascular events when fed CT images from asymptomatic as well as stable and acute chest pain cohorts into a DL implementation. On top of ECG measurements and cardiopulmonary exercise tests, a large population of people with congenital heart disease was assessed for prognosis and treatment based on machine learning, rather than imaging (Diller G-P, et al., 2019).

ML has recently found the interaction between previously unrelated imaging tests via the discovery of previously overlooked correlations. Utilizing mammography, a ML model was used to identify individuals with calcifications in their breast arteries and those at high risk of cardiovascular disease. It has also been utilized to anticipate anomalies in the macro vasculature based on the eye's microvascular properties using ML and the non-invasiveness of retinal scanning. Cardiovascular risk factors can be predicted using retinal fundus photographs, resulting in an easier and more cost-effective cardiovascular risk stratification (Poplin R, et al., 2018), or the ML implementation that inferred CAC scores from retinal photographs, which proved to be as accurate as CT scan-measured CAC in predicting cardiovascular events.

Results for research question 2

What are the challenges of applying machine learning in healthcare?
Judgments derived from low-level data acquisition and feature extraction activities, such as interpreting and making decisions about a patient's state, carry a far greater risk than decisions generated from low-level data acquisition and feature extraction tasks. For this reason, ML results must be interpreted by the specific experts in different medical fields and verified in a much more extensive fashion (e.g, class IIa or class IIb pathways for marketability), culminating in randomized, prospective studies.

The extraction of meaningful ideas from raw data is a major difficulty for ML methods to status interpretation. This problem is only one of several that arises from the nature of the data itself. Training and outcome data's trustworthiness and representativeness are at the heart of the use of machine learning in healthcare. A credible measure is needed to compare diverse datasets, which is not easy to discover. In addition, data collecting techniques should be structured to address changes in gender, ethnicity, and age, as well as unusual outliers, for effective interpretation (Bild DE, et al., 2002). Additional consideration should be given to longitudinal data, such as during a stress protocol or course of a disease (Lee G, et al., 2019). Three types of data are used to train ML models: randomized clinical trials, cohorts, and clinical routine real-world data. The quality and completeness of this data varies from high to poor. Data from randomized clinical trials, for example, may not be generalizable to other types of data (e.g., surveys). This makes the sharing of information across various kinds of data difficult.

Bias, or the lack of representativeness of the training sample, is another major issue with the data presently accessible. There have been concerns about the generalizability of multicentre studies, which may be skewed toward the techniques of treatment utilized in the investigated centres, according to a recent evaluation of risk prediction models (Wynants L, et al., 2019). As an example, procedures for cardiac MRI differ widely across institutions and equipment vendors. The health disparity between the dominant social group, whose data is used to train algorithms, and people of colour might be exacerbated as a result of this prejudice.

When evaluating a trained model in new clinical centers, vigilance is required. ML users can witness to the fact that there will always be a trade-off between boosting the system's local performance and ensuring that the system generalizes effectively (Futoma J, et al., 2020). It's possible that the human predisposition to accept a computer-generated answer without seeking for contrary facts may potentially impair clinical interpretation and decision-making. Goddard et al. (Goddard K, et al., 2012) found that when the ML solution is trustworthy, it enhances human performance, but when the answer is erroneous, human mistakes rises. A diagnostic algorithm's failure may be traced back to a number of different factors. Access to training data and learning systems equipped with tools that enable recreating the logic behind a choice might alleviate some of these difficulties.

Discussion

To arrive at a clinical judgment, a physician must look at all the information at their disposal and compare it to previous cases or patterns they've learned to identify. In the framework of assumed normalcy and typical instances, prior information on therapy effects is employed to manage this unique patient. Only extremely experienced therapists may use this 'eminence-based' strategy. Diagnostic recommendations based on large cohorts or clinical trials (Ponikowski P, et al., 2016) are often used by professional organizations. While recommendations have greatly improved medical treatment, they do not take into account all of the available data. This justifies the usage of ML in this context. Routinely gathered data is generally significantly noisier, diverse and incomplete, which is why most machine learning models are trained on data from randomized clinical trials (Kalscheur MM, et al., 2018).

By doing imputation or adopting formulations that expressly reflect that the data is partial, ML approaches must cope with incompleteness. Aside from the narrow selection criteria of cardiology trials (including co-morbidities like ethnicity and gender and age and lifestyle), patients may have been treated differently prior to the investigation or at a different stage of the disease, and they may undergo different decision pathways during the study. Obtaining hard outcomes to train an algorithm is generally challenging, for example, to record death the research would need to be performed until everyone dies, which is unrealistic both for time and financial restrictions. Patients may have diverse reasons for suffering side effects, even if they are documented (Oladapo OT, et al., 2015).

Unsupervised predictive ML/DL algorithms, which may fail to comprehend the context from which data have been gathered, may generate undesirable findings that might damage patients as a consequence of their inability to correlate input descriptors to outcomes. If input descriptors are used to place people according to how similar they are, and this similarity is paired with prior information, it may be used to infer a diagnosis or predict a treatment response, an unsupervised dimensionality reduction technique seems more promising.

Limitations of the review

Real-world data needed to create, test, and validate algorithms was never evaluated in this research. This review does not explicitly evaluate researchers, but they should be aware that real-world data sources have limits no matter what approach is used. It is important for researchers to be aware of the degree to which machine learning techniques are dependent on the data format and availability, and to examine a proposed dataset to verify that it is adequate for the machine learning project when employing observational datasets for research. Databases must be thoroughly analysed to determine which factors are included, as well as which variables may have predictive or prognostic significance but are not included. Retrospective databases' omission of critical information raises questions about their suitability for machine learning. Constraints such as unmeasured confounding, bias, and patient selection criteria must also be considered. Aside from the obvious considerations, there are things that need to be taken into account when adopting these techniques. It is crucial to note that the Luo checklist is a great tool for ensuring that any machine-learning study fulfils high research standards for patient care, and particularly involves the examination of missing or possibly wrong data (i.e. outliers). In addition, prior to putting the model into action, it is recommended that all relevant data be thoroughly evaluated and that a variety of modelling approaches be used.

Concluding remarks

Automated machine learning (ML) methods enable computers to detect patterns in data and improve over time. Clinical decision-making is about to undergo a paradigm change because to these algorithms and the massive amounts of data generated by the digitization of healthcare systems. In order for ML models to be seamlessly integrated into clinical workflows, it is necessary to understand the processes that physicians use to make choices, which in turn helps identify the areas in which ML models may be most effective. ML might change several parts of healthcare, especially cardiovascular medicine, if the challenges and dangers discussed in this research can be overcome. Engineers and doctors must work together to design and test particular ML-enabled clinical applications in order for the promise to be realized.

REFERENCES

A. Grewal, M. Kaur, and J. H. Park, "A unified framework for behaviour monitoring and abnormality detection for smart home," *Wireless Communications and Mobile Computing*, vol. 2019, Article ID 1734615, 16 pages, 2019.

Brosch T, Tam R. Manifold learning of brain MRIs by deep learning. Med Image Comput Comput Assist Interv 2013;16:633–40.

Cummings ML. "Automation bias in intelligent time critical decision supportsystems," In:AIAA 3rd Intelligent Systems conference. (2004), p. 6313.

Diller G-P, Kempny A, Babu-Narayan S V, HenrichsM, BridaM, Uebing A, et al. Machine learning algorithms estimating prognosis and guiding therapy in adult congenital heart disease: data from a single tertiary centre including 10019 patients. Eur Heart J. (2019) 40:1069–77. doi: 10.1093/eurheartj/ehy915

Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017;542:115–18.

Funkner AA, Yakovlev AN, Kovalchuk S V. Data-driven modeling of clinical pathways using electronic health records. Procedia Comput Sci.(2017)121:8342.doi:10.1016/j.procs.2017.11.108

Futoma J, Simons M, Panch T, Doshi-Velez F, Celi LA. The myth of generalisability in clinical research andmachinelearninginhealthcare.LancetDigitHealth(2020)2:e48992.doi:10.1016/S2589-7500(20)30186-2

Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review offrequency, effect mediators, and mitigators. J Am Med Inform Assoc : JAMIA.(2012) 19:121–7. doi: 10.1136/amiajnl-2011-000089

Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 2016;316:2402–10.

Kalscheur MM, Kipp RT, Tattersall MC, Mei C, Buhr KA, DeMets DL, et al. Machine Learning Algorithm Predicts Cardiac Resynchronization Therapy Outcomes. Circulation: Arrhythm Electrophysiol. (2018) 11:499. doi: 10.1161/CIRCEP.117. 005499

Kwon J, Kim K-H, Jeon K-H, Park J. Deep learning for predicting in hospital mortality among heart disease patients based on echocardiography.Echocardiography. (2019) 36:213–8. doi: 10.1111/echo.14220

Lane ND, Georgiev P. Can deep learning revolutionize mobile sensing? In International Workshop on Mobile Computing Systems and Applications, Santa Fe, NM, USA, 2015, 117–22.

Lee G, Kang B, Nho K, Sohn K-A, Kim D. MildInt: deep learning-based multimodal longitudinal data integrationframework.FrontGenet.(2019)10:617.doi:10.3389/fgene.2019.00617

Liu S, Liu S, Cai W, et al. Early diagnosis of Alzheimer's disease with deep learning. In: International Symposium on Biomedical Imaging, Beijing, China 2014, 1015–18.

M. Kang, E. Park, B. H. Cho, and K. S. Lee, "Recent patient health monitoring platforms incorporating internet of things-enabled smart devices," *International Neurourology Journal*, vol. 22, no. 4, p. 313, 2018.

M. Tahir, M. Sardaraz, S. Muhammad et al., "A lightweight authentication and authorization framework for blockchain-enabled IoT network in health-informatics," *Sustainability*, vol. 12, no. 17, p. 6960, 2020.

Madani A, Ong JR, Tibrewal A, Mofrad MRK. Deep echocardiography:data-efficient supervised and semi-supervised deep learning towards automated diagnosis of cardiac disease. npj Digital Med. (2018)1:59. doi: 10.1038/s41746-018-0065-x

Oladapo OT, Souza JP, Bohren MA, Tunçalp Ö, Vogel JP, Fawole B, et al. Better Outcomes in Labour Difficulty (BOLD) project: innovating to improve quality of care around the time of childbirth. Reprod Health. (2015)12:48. doi: 10.1186/s12978-015-0027-6

Pathak J, Bailey KR, Beebe CE, Bethard S, Carrell DS, Chen PJ, et al.Normalization and standardization of electronic health records for highthroughput phenotyping: the SHARPn consortium. J Am Med Inform Assoc.(2013) 20:e341–8. doi: 10.1136/amiajnl-2013-001939

Piwek L, Ellis DA, Andrews S, et al. The rise of consumer health wearables: promises and barriers. PLoS Med 2016; 13:e1001953.

Ponikowski P, Voors AA, Anker SD, Bueno H, González-Juanatey JR, Harjola V-P, et al. ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. Eur Heart J. (2016) 37:2129–200. doi: 10.1093/eurheartj/ehw128

Poplin R, Varadarajan A V., Blumer K, Liu Y, McConnell M V,Corrado GS, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. Nat Biomed Eng. (2018) 2:158–64. doi: 10.1038/s41551-018-0195-0

Prasoon A, Petersen K, Igel C, et al. Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. Med Image Comput Comput Assist Interv 2013;16:246–53.

Ravi D, Wong C, Lo B, et al. A deep learning approach to on-node sensor data analytics for mobile or wearable devices. IEEE J Biomed Health Inform 2017;21:56–64.

Shameer K, Badgeley MA, Miotto R, et al. Translational bioinformatics in the era of real-time biomedical, healthcare and wellness data streams. Brief Bioinform 2017; 18:1105–124.

W. Li, Y. Chai, F. Khan et al., "A comprehensive survey on machine learning-based big data analytics for IoT-enabled smart healthcare system," *Mobile Networks and Applications*, vol. 26, no. 1, pp. 234–252, 2021.

Wang J, Ding H, Bidgoli FA, Zhou B, Iribarren C, Molloi S, et al. Detecting cardiovascular disease from mammograms with deep learning. IEEE TransMed Imaging. (2017) 36:1172–81. doi: 10.1109/TMI.2017.2655486

Wynants L, Kent DM, Timmerman D, Lundquist CM, Calster B. Untapped potential of multicenter studies: a review of cardiovascular risk prediction models revealed inappropriate analyses and wide variation in reporting. Diagnostic Prognostic Res. (2019) 3:9. doi: 10.1186/s41512-019-0046-9

Xiong HY, Alipanahi B, Lee LJ, et al. The human splicing code reveals new insights into the genetic determinants of disease. Science 2015;347:1254806.

Yoo Y, Brosch T, Traboulsee A, et al. Deep learning of image features from unlabeled data for multiple sclerosis lesion segmentation. In: International Workshop on Machine Learning in Medical Imaging, Boston, MA, USA, 2014, 117–24.