MACHINE LEARNING APPLICATION IN NIGERIA HEALTHCARE SYSTEM, THE OPPORTUNITIES AND CHALLENGES.

A paper presented by:

OGBONNAYA VINCENT, NWENE

nwenevin@gmail.com (08066867988)

SHEYI JONATHAN, AYODELE

Sheyijoe.sj@gmail.com (07065026954)

And

ENGR. CYRIL N. OKWUEZE

cyokwueze@gmail.com (08066868222)

COMPUTER ENGINEERING DEPARTMENT, FEDERAL POLYTECHNIC OKO, ANAMBRA STATE

Abstract

The integration of machine learning (ML) into Nigeria's healthcare system presents a transformative opportunity to enhance patient care, streamline operations, and improve health outcomes. This paper explores the potential applications of ML in various aspects of healthcare, including disease prediction and diagnosis, personalized treatment plans, healthcare management, and predictive analytics for public health. The study highlights significant opportunities such as improved diagnostic accuracy, early detection of diseases, optimized resource allocation, and enhanced patient management. However, the implementation of ML in Nigeria faces substantial challenges, including limited digital infrastructure, insufficient data quality and availability, lack of skilled professionals, and regulatory and ethical concerns. By examining case studies and existing pilot projects within Nigeria and other comparable settings, this paper provides a comprehensive overview of the current state of ML in Nigerian healthcare, identifies key obstacles, and proposes strategic recommendations to overcome these barriers. Emphasizing the need for a multi-stakeholder approach, the paper advocates for collaborative efforts between government, private sector, and academic institutions to harness the full potential of ML, ultimately aiming to create a more efficient, equitable, and responsive healthcare system in Nigeria.

Keywords: Machine learning, Healthcare system, Diagnosis, Nigeria, Healthcare management

INTRODUCTION

Machine learning (ML) has emerged as a transformative technology across various sectors, and healthcare is no exception. In Nigeria, the integration of machine learning into the healthcare system holds significant potential to enhance the quality of care, improve patient outcomes, and address many of the challenges faced by the sector. The application of machine learning in healthcare involves using algorithms and statistical models to analyze vast amounts of medical data, enabling more accurate diagnoses, personalized treatment plans, and predictive analytics.

The essence of data science required to implement various ML models lies at the intersection of multiple disciplines (Deo, 2015), particularly mathematics, statistics, and computer science. Despite the impressive capabilities of human intelligence, humans are prone to errors due to limited short-term memory (Connor, 2019). With the rapid growth in data and the increasing computational power to process it, there is a substantial opportunity to leverage ML to aid human decision-making by incorporating extensive contextual information.

Despite the promise that machine learning holds for the Nigerian healthcare system, there are numerous challenges that must be addressed to realize its full potential. These include issues related to data quality and availability, infrastructure limitations, regulatory hurdles, and the need for skilled professionals who can develop and implement ML solutions. Additionally, there are ethical and privacy concerns associated with the use of patient data that must be carefully managed.

This paper explores the opportunities and challenges of implementing machine learning in Nigeria's healthcare system. It delves into the various ways ML can be applied to improve healthcare delivery, such as in predictive analytics, diagnostic accuracy, treatment personalization, and operational efficiency. Furthermore, it examines the barriers to successful implementation and suggests strategies to overcome these challenges. By understanding both the opportunities and obstacles, stakeholders can better navigate the integration of machine learning into the healthcare sector, ultimately enhancing the health and well-being of the Nigerian population.

TRADITIONAL PROGRAMMING VS MACHINE LEARNING

Traditional Programming

Traditional programming is a human-driven endeavour which requires meticulous problem analysis and the creation of code-based solutions. For example, developing an image recognition system to detect rats necessitates accounting for various lighting conditions, rat breeds, and poses. This manual approach is time-consuming and often falls short when confronted with diverse scenarios.

Machine Learning

Machine learning involves empowering algorithms to learn from data and adjust their behaviour accordingly. Instead of relying on explicit rule-based programming, machine learning algorithms discern complex patterns and relationships from the examples provided in the data.

Fig 1 The Difference between Traditional Programing and Machine Learning Model

Traditional programming is ideal for tasks with clear, deterministic rules, while machine learning excels in handling complex, data-rich environments where rules are not easily defined or are constantly changing.

TYPE OF MACHINE LEARNING MODELS

Machine Learning models are divided into supervised learning, unsupervised learning, and reinforcement learning, their applications are dependent on the kind of problem you are trying to solve (Bi Q et al, 2019).

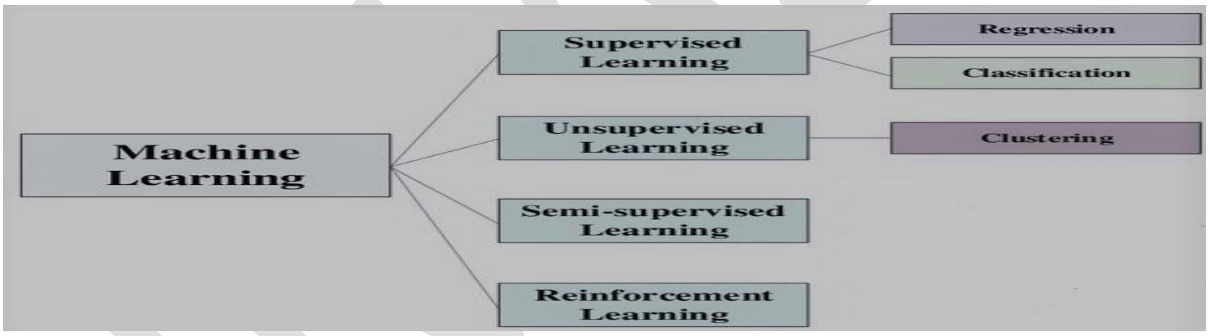

Fig 2 Types of Machine Learning Models

SUPERVISED LEARNING

Supervised machine learning (ML) models are applied when the outcome of interest is specified and the data is explicitly labelled with this outcome (Jiang et al., 2022). For instance, the outcome might be the presence or absence of a disease such as diabetes or hypertension. A critical requirement for these models is the accuracy of the outcome labels in the data, as any bias in outcome assessment can negatively impact the model's performance and reduce its generalizability to populations outside the training dataset.

The process of developing a supervised ML model is well-defined and mature, given the frequent use of these models. The initial step involves formulating a problem statement, followed by identifying the relevant data needed to address the problem (Kotsiantis, 2020). Data processing is a crucial step, which may involve data wrangling, limiting the data to necessary variables, and removing irrelevant variables (Endel & Piringer, 2015).

A common method for creating a supervised ML model includes selecting a random sample of the data, choosing an appropriate algorithm, training the model, and evaluating its performance. There are several types of algorithms to choose from, and one approach is to train the model using multiple relevant algorithms and evaluate their performance using metrics such as the confusion matrix and receiver operating characteristic (ROC) curve (Kendale et al, 2018). Developing the final model is an iterative process, where the algorithm with the best combination of parameters is selected to ensure maximum predictive ability.

Different algorithms used for supervised machine learning

The algorithms used in supervised learning can be categorically divided into regression and classification algorithms based on the prediction of a quantitative or categorical variable (Mahesh, 2019).

A. **Regression Algorithm**

Linear Regression

Linear regression is a widely used algorithm designed to predict a quantitative variable using one or more independent variables, assuming a linear relationship between the independent variables and the outcome variable (Gandhi, 2018). Due to its extensive use, linear regression is well-understood, making it easier for health professionals to interpret compared to other machine learning algorithms.

Support Vector Machine Regression

When a quantitative dependent variable does not exhibit a linear relationship with the predictors, support vector machine (SVM) regression can be an effective alternative for predicting the outcome variable. SVM regression offers advantages such as accurate predictions without sacrificing generalizability and robustness to outliers. However, this model has limitations; it is less suitable for large datasets and scenarios where the number of variables exceeds the sample size (Raj, 2020).

Decision Tree Regression

Decision Tree Regression is a model designed to predict a quantitative variable by sequentially asking true-or-false questions. The model makes predictions when it reaches a point where it can confidently provide an accurate answer. One of the key strengths of this algorithm is its intuitive nature, as it mimics the way humans think and predict outcomes by asking multiple questions. Additionally, it is a non-parametric method, meaning it does not rely on strict assumptions about the distribution of the outcome variable.

Classification algorithms

Logistic Regression

Logistic regression is a widely used classification algorithm for predicting a categorical outcome when the outcome has two levels (Brownlee, 2016). One disadvantage of logistic regression is the need to manually add interaction terms. For predicting categorical variables with more than two levels, multinomial logistic regression is used (Molnar, 2021).

K Nearest Neighbours

K Nearest Neighbors (KNN) is a straightforward prediction algorithm noted for its ease of implementation and interpretation (Choudhary, 2017). It requires minimal parameter tuning. However, the algorithm becomes slower as the number of predictor variables increases (Harrison, 2019). KNN categorizes results based on the majority vote of neighbouring data points near the point of interest, utilizing distances such as Euclidean, Manhattan, and Minkowski for this purpose.

Naïve Bayes Classifier

The Naïve Bayes classifier is based on Bayes' theorem and is well-suited for predicting outcomes with more than two levels (Gandhi, 2018). When the assumption of independence among predictor variables holds, it requires less training data compared to logistic regression. However, in real-world scenarios, the predictor variables are often not truly independent.

UNSUPERVISED LEARNING

In unsupervised machine learning (ML), the outcome is not specified, and the data is not labelled. The algorithm identifies and infers patterns from the data without relying on an outcome variable. Various methods can be employed for unsupervised learning, with clustering being a prominent technique where data is grouped into clusters by an algorithm. However, caution is needed when making decisions based on clusters, as cluster analysis can overestimate similarities between groups. Unsupervised ML techniques are useful for identifying unusual patterns in the data and detecting anomalies, such as potential fraud. When dealing with many features, dimensionality reduction can be applied to streamline the dataset.

SEMI-SUPERVISED LEARNING

Semi-supervised learning is a machine learning approach that combines both labelled and unlabelled data for training. This method leverages the large amounts of unlabelled data, which is often easier and cheaper to obtain, alongside a smaller set of labelled data. The primary objective is to improve learning accuracy over what would be achieved using only labelled data

REINFORCEMENT LEARNING

Reinforcement learning is a practical application of machine learning where a sequence of models is created to make decisions. The system takes actions and receives feedback based on those actions (Osinski & Budek, 2018). It then modifies its decisions and actions according to the feedback received. This process begins with trial and error, allowing the system to learn and improve over time. The system continues to adjust until it consistently receives favourable feedback. When various machine learning algorithms are creatively combined to make different decisions, receive regular feedback, and optimize final decisions based on outcomes, reinforcement learning is at play.

In healthcare, decision-making is complex and requires considering numerous factors. Reinforcement learning can be used to design decision support systems that provide treatment recommendations to physicians. However, several challenges arise when applying reinforcement learning in healthcare, such as evaluating the decisions made by the system and determining appropriate rewards to modify decisions and actions (Riachi et al, 2021). Additionally, understanding disease dynamics and establishing contextual causal relationships between relevant factors and outcomes is crucial.

APPLICATION OF MACHINE LEARNING IN DIFFERENT AREAS OF HEALTHCARE

Clinical Decision Support Systems in Healthcare

Machine learning (ML) methods and algorithms are increasingly used to develop clinical decision support systems (CDSS) that assist clinicians in various medical tasks. During COVID-19 era, support vector machine, neural network, gradient boosted decision tree and logistic regression were ensemble to distinguish the mortality risk of patients with the infection (Gao et al, 2020). On the other, CDSS has been designed to reduce prescribing errors by assisting in prioritizing prescription checks. (Corny et al., 2020). A support system can as well be developed to assist the pharmaceutical industries a specific molecule for her research that is most suitable to pass through regulatory process and reach the market as drugs (Onay & Onay, 2020). Maternal health initiatives can also use a CDSS to predict ectopic pregnancies (De Ramon Fern et al., 2019). CDSS have also designed the use of imaging data like positron emission tomography to detect early stages of cancer to enable its treatment at the early stages of development (Teramoto et al, 2020). With the presence of CDSS, a deep-learning algorithm has been used to detect diabetic retinopathy in diabetic patients (Gulshan et al.,2016).

One of the major concerns in implementing ML algorithms is their ability to match expert opinion but research has shown that ML algorithms can perform comparably to experts. For instance, a deep neural network algorithm mimicked the diagnostic skills of experienced dermatologists in detecting skin cancer (Esteva et al.,2017). Similarly, a nested neural network algorithm performed on par with an experienced radiologist in predicting diseases based on chest x-rays.

Use of Machine Learning in Public Health

The use of Machine learning (ML) algorithms can predict population-level healthcare outcomes by analyzing large datasets. These algorithms are particularly advantageous when dealing with extensive data and non-linear relationships between outcomes and independent variables. The reproducibility of several ML algorithms highlights their effectiveness, but their true value lies in the use of authentic and high-quality data.

Predictive models in healthcare are not new; for example, the Framingham health score was developed in 1967. While parametric statistics have traditionally been used to predict population-level outcomes, there is a growing interest in leveraging various ML algorithms for these predictions. Although the theoretical foundations for ML models have existed for some time, the recent surge in the generation of large, variable data at high velocity by public health organizations has expanded the use of ML in this field. Enhanced data processing capabilities through advanced hardware and cloud solutions have further facilitated the effective application of complex algorithms to Big Data. Examples of ML applications in population health include predicting childhood lead poisoning, suicidal ideation, diabetic retinopathy, managing public health emergencies, and tracking yellow fever incidence.

A review of population health prediction through ML shows that neural networks, support vector machines, and single tree-based models are the most commonly used algorithms for predicting public health outcomes (Morgenstern et al., 2020). These algorithms are frequently employed to predict both communicable and non-communicable diseases, with cardiovascular disease being the most predicted outcome. Additionally, ML models are used to predict behaviours such as healthcare utilization. ML algorithms for predicting population-level health outcomes typically use structured data from electronic medical records and similar sources. Although unstructured data from multiple publicly available sources could be valuable in predicting various health outcomes, this practice is not yet common among public health practitioners.

Benefits of Implementing Machine Learning in Healthcare System in Nigeria

Implementing machine learning (ML) in healthcare systems in Nigeria presents numerous opportunities to improve healthcare delivery, enhance patient outcomes, and equally optimize operational efficiency. Here are some key aspects where ML can be transformative:

1. **Improved Diagnostics and Disease Detection**:
    - ML algorithms can analyze medical images, genetic data, and patient records to detect diseases early, often with greater accuracy than traditional methods.
    - Examples include the detection of cancers, diabetic retinopathy, and cardiovascular diseases.
2. **Personalized Treatment**:
    - ML can help tailor treatment plans to individual patients based on their genetic makeup, lifestyle, and other factors, leading to more effective and personalized healthcare.
3. **Predictive Analytics**:
    - ML models can predict disease outbreaks, patient readmissions, and the progression of chronic diseases, enabling proactive healthcare measures and resource allocation.
4. **Enhanced Medical Research**:
    - ML facilitates the analysis of vast amounts of biomedical data, accelerating the discovery of new drugs, treatments, and understanding of diseases.
    - It can identify patterns and correlations in complex datasets that might be missed by human researchers.
5. **Risk Management**:
    - ML models can identify patients at high risk for certain conditions or adverse events, allowing for timely interventions and reducing complications.
6. **Patient Monitoring and Support**:

- Wearable devices and mobile health apps use ML to monitor patients' vital signs in real-time, providing alerts for potential health issues.
- ML-powered chatbots and virtual assistants can offer patients immediate support and information, improving patient engagement and adherence to treatment plans.

7. **Reduction of Human Error**:
   - By automating routine tasks and analyzing complex data, ML can reduce the likelihood of human errors in diagnosis, treatment, and administrative processes.
8. **Cost Reduction**:
   - By improving diagnostic accuracy, optimizing treatment plans, and enhancing operational efficiency, ML can contribute to significant cost savings in healthcare.

By harnessing the power of ML, Nigeria can address some of her critical healthcare challenges, improve the quality of care to her people, and enhance the overall efficiency of her healthcare system.

Challenges in Implementing ML Models in Healthcare

Implementing machine learning (ML) models in healthcare involves several challenges that can affect their effectiveness and adoption. Here are some of the key challenges that could possibly hinder the implementation of machine learning in Nigerian healthcare system:

1. Data Quality and Availability:

- Incomplete Data: Healthcare data is often incomplete or missing critical information.
- Data Standardization: Lack of standardized formats across different healthcare systems complicates data integration.
- Data Labelling: Labelling healthcare data for supervised learning is labour-intensive and requires expert knowledge.

2. Data Privacy and Security:

- Patient Privacy: Ensuring patient data privacy and complying with regulations like Health Insurance Portability and Accountability Act (HIPAA) is crucial.
- Data Security: Protecting sensitive health data from breaches and cyber-attacks is essential.

3. Integration with Existing Systems:

- Interoperability: ML models need to integrate seamlessly with existing electronic health record (EHR) systems and other healthcare IT infrastructure.
- Legacy Systems: The use of outdated technology could pose a strong integration challenges.

4. Regulatory and Ethical Issues:

- Regulatory Approval: ML models must undergo rigorous regulatory approval processes to ensure safety and efficacy.
- Ethical Concerns: Addressing ethical issues such as bias in algorithms and ensuring equitable access to ML benefits.

5. Model Interpretability:

- Black-box Models: Many ML models, especially deep learning, are complex and difficult to interpret, making it hard for clinicians to trust and adopt them.
- Explainability: Ensuring that ML models provide clear and understandable explanations for their predictions.

7. Resource and Cost Constraints:

- Implementation Costs: Developing, deploying, and maintaining ML models can be expensive.
- Training and Expertise: There is a need for specialized skills to develop and manage ML models, which may be lacking in many healthcare organizations.

Addressing these challenges requires a collaborative effort between data scientists, healthcare professionals, regulatory bodies, and technology developers to ensure that ML models can be effectively and safely implemented in healthcare settings.

Conclusion and Recommendations

Hypothesis-driven research utilizing epidemiologic and statistical knowledge has been a cornerstone of healthcare for many years. However, the recent advent of large datasets and enhanced computational power of high-speed physical and virtual machines has facilitated the development of various predictive machine learning (ML) algorithms. These algorithms have enabled the creation of clinical decision support systems and the prediction of population-based health parameters.

Given the growing demand for ML algorithms in health research and their application in clinical practice, it is crucial for health professionals to receive proper training to comprehend the relevant terminologies and methodologies. Likewise, data scientists need to understand the conceptual connections between data science and epidemiological principles.

Lastly, it is essential to uphold the ethical principle of "do no harm" when generalizing findings from these ML algorithms to ensure the responsible application of these technologies in healthcare.

REFRENCE

Deo RC. Machine learning in medicine. Circulation 2015;132(20):1920–30. https://doi.org/10.1161/CIRCULATIONAHA.115.001593.

Connor CW. Artificial intelligence and machine learning in anaesthesiology. Anaesthesiology 2019;131(6):1346–59. https://doi.org/10.1097/ ALN.0000000000002694.

Bi Q, Goodman KE, Kaminsky J, Lessler J. What is machine learning? A primer for the epidemiologist. Am J Epidemiol 2019;188(12):2222–39. https://doi.org/ 10.1093/aje/kwz189.

Jiang T, Gradus JL, Rosellini AJ. Supervised machine learning: a brief primer. Behav Ther 2020;51(5):675–87. https://doi.org/10.1016/j.beth.2020.05.002.

Osinski B, Budek K. What is reinforcement learning? The complete guide. ´ deepsense.ai. Published July 5, 2018. Accessed December 17, 2021. https://deepse nse.ai/what-is-reinforcement-learning-the-complete-guide/.

Kotsiantis SB. Supervised machine learning: a review of classification techniques. :2020.

Endel F, Piringer H. Data Wrangling: making data useful again. IFAC-Papers Online 2015;48(1):111–2. https://doi.org/10.1016/j.ifacol.2015.05.197.

Kendale S, Kulkarni P, Rosenberg AD, Wang J. Supervised machine-learning predictive analytics for prediction of post induction hypotension. Anaesthesiology 2018;129(4):675–88. https://doi.org/10.1097/ALN.0000000000002374.

Mahesh B. Mach Learn Algorithms - Rev. 2019. https://doi.org/10.21275/ ART20203995.

Choudhary R, Gianey HK. Comprehensive review on supervised machine learning algorithms. In: 2017 international conference on machine learning and data science. MLDS); 2017. p. 37–43. https://doi.org/10.1109/MLDS.2017.11.

Gandhi R. Introduction to machine learning algorithms: linear regression. Medium. Published May 28, 2018. Accessed December 15, 2021. https://towardsdatascience .com/introduction-to-machine-learning-algorithms-linear-regression-14c4e 325882a.

Raj A. Unlocking the true power of support vector regression. Medium. Published October 5, 2020. Accessed December 16, 2021. https://towardsdatascience.com/u nlocking-the-true-power-of-support-vector-regression-847fd123a4a0.

Brownlee J. Logistic regression for machine learning. Mach Learn Mastery. Published March 31, 2016. Accessed December 16, 2021. https://machinelearning mastery.com/logistic-regression-for-machine-learning/.

Molnar C. 5.2 logistic regression | interpretable machine learning. Accessed December 16, 2021. https://christophm.github.io/interpretable-ml-book/logistic.

Harrison O. Machine learning basics with the K-nearest neighbours algorithm. Medium. Published July 14, 2019. Accessed December 16, 2021. https://towa rdsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algori thm-6a6e71d01761.

Gandhi R. Naive Bayes classifier. Medium. Published May 17, 2018. Accessed December 16, 2021. https://towardsdatascience.com/naive-bayes-classifier-81d 512f50a7c.

Riachi E, Mamdani M, Fralick M, Rudzicz F. Challenges for Reinforcement Learning in Healthcare. arXiv:210305612 [cs]. Published online March 9, 2021. Accessed December 17, 2021. http://arxiv.org/abs/2103.05612.

Gao Y, Cai GY, Fang W, et al. Machine learning based early warning system enables accurate mortality risk prediction for COVID-19. Nat Commun 2020;11(1):5033. https://doi.org/10.1038/s41467-020-18684-2.

Corny J, Rajkumar A, Martin O, et al. A machine learning-based clinical decision support system to identify prescriptions with a high risk of medication error. J Am Med Inf Assoc 2020;27(11):1688–94. https://doi.org/10.1093/jamia/ocaa154.

Onay A, Onay M. A drug decision support system for developing a successful drug candidate using machine learning techniques. Curr Comput Aided Drug Des 2020; 16(4):407–19. https://doi.org/10.2174/1573409915666190716143601.

De Ramon Fern´ andez A, Ruiz Fern´ andez D, Prieto S´ anchez MT. A decision support ´ system for predicting the treatment of ectopic pregnancies. Int J Med Inf 2019;129:198–204. https://doi.org/10.1016/j.ijmedinf.2019.06.002.

Teramoto A, Yamada A, Tsukamoto T, et al. Decision support system for lung cancer using PET/CT and microscopic images. In: Lee G, Fujita H, editors. Deep Learning in medical image analysis: Challenges and applications. Advances in experimental medicine and biology. Springer International Publishing; 2020. p. 73–94. https://doi.org/10.1007/978-3-030-33128-3_5.

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(22):2402–10. https://doi.org/10.1001/jama.2016.17216.

Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017;542(7639):115–8. https://doi.org/ 10.1038/nature21056.

Morgenstern JD, Buajitti E, O'Neill M, et al. Predicting population health with machine learning: a scoping review. BMJ Open 2020;10(10): e037860. https://doi. org/10.1136/bmjopen-2020-037860.