Review Article

Harnessing Machine Learning for Predictive Healthcare: A Path to Efficient Health Systems in Africa

Kabiru Gulma[1]*, Zainab Saidu[2], Kingsley Godfrey[2], Abubakar Wada[3], Zayyanu Shitu[4], Auwal Adam Bala[5], Safiya Bala Borodo[6], Sa'adatu M Julde[3] and Mustapha Mohammed[7]

[1]School of Global Health and Bioethics, Euclid University, The Gambia

[2]Clinton Health Access Initiative, Nigeria

[3]Faculty of Pharmaceutical Sciences, Bayero University, Kano, Nigeria

[4]Department of Clinical and Administrative Pharmacy, University of Abuja, Nigeria

[5]Department of Human Genetics, University of Texas Rio Grande Valley, Texas, USA

[6]Department of Pharmacology and Therapeutics, Bayero University, Kano, Nigeria

[7]Health Sector, Qatar University, Doha, Qatar

https://www.healthdisgroup.com/hiim

Check for updates

Abstract

Machine learning (ML) presents a transformative opportunity to strengthen African health systems through predictive healthcare. This paper explores the applications, benefits, and implementation challenges of ML in African health contexts, where resource limitations and infrastructure gaps often impede efficient healthcare delivery. By leveraging supervised and unsupervised ML models-such as decision trees, neural networks, and support vector machines-predictive healthcare can aid in early disease detection, improve patient outcomes, and optimize resource allocation. Real-world case studies across the continent, including malaria forecasting and telemedicine applications, illustrate the potential of ML to mitigate the burdens of delayed diagnosis, an underutilized workforce, and a fragmented health infrastructure. However, barriers such as limited access to high-quality, structured health data, privacy concerns, algorithmic bias, and ethical dilemmas related to fairness and transparency must be addressed. The manuscript critically examines data preprocessing techniques, data source diversity, and the necessity of ethical frameworks for AI integration. Future directions include leveraging wearable technologies, integrating interdisciplinary research, and contextualizing ML models within Africa's unique socio-political and epidemiological realities. The study argues for developing equitable, data-driven, and scalable ML solutions tailored to Africa's public health priorities, shifting from reactive to predictive health systems.

Introduction to machine learning in health-care

The advent of Machine Learning (ML) in the healthcare sector capabilities in both clinical practice and health research. It creates the opportunity for the efficient and effective use of health systems, including training, research, and service delivery [1]. Machine learning is an umbrella term that encompasses techniques, algorithms, and tools that derive knowledge and predictions from datasets. It includes regression, clustering, classification, dimensionality reduction, and recommender systems. It also includes tools such as decision trees, random forests, Naïve Bayes, support vector machines, and neural networks. Machine learning is constantly evolving, with new algorithms being proposed, adopted, and improved.

Disease prediction and diagnosis, personalized medicine, and healthcare management are aspects of healthcare advertising the potential of machine learning [2]. Healthcare

001

holds disease prediction and diagnosis, personalized medicine, and healthcare management, all of which can benefit from machine learning [2]. These healthcare advancements offer the potential for greater happiness, enhanced productivity, and significant financial savings. However, the challenge lies in cost-efficiently delivering these services, especially in the developing world. Africa, the second-largest and second-most populous continent on Earth, is expected to double its population by 2050. Despite being rich in natural resources, it is the poorest continent.

Poor health systems are a major burden in addressing the health needs of their population, particularly in the sub-Saharan region, which has the highest need for improvements in healthcare systems and service delivery. Smart disease predictive models built on robust decision trees have the potential to enhance the efficiency of healthcare delivery in this region. This potential relies on the availability and sensible use of historical disease datasets. However, implementing these smart biomedical systems faces numerous challenges, including the availability, accessibility, and usability of historical datasets. Additionally, the reluctance of healthcare providers to share their datasets and the quality of the datasets being incomplete, high-dimensional, and noisy are major obstacles. Given these challenges in Africa, it is important to shed light on the issues afflicting the health systems and explore potential solutions. The promise of happier lives, increased productivity, and financial savings. The challenge remains the cost-efficient delivery of these services, especially for the developing world. Africa is the second-largest and second-most populous continent on Earth, with its population expected to double by 2050. Despite being rich in natural resources, Africa is the poorest continent. Poor health systems are a major burden in handling their diseased population. Sub-Saharan Africa has the highest need for interventions to improve health systems and service delivery. Smart disease predictive models   built on robust decision trees have the potential to improve the efficiency of healthcare delivery. The potential of predictive models guarantees the availability and sensible use of historical disease datasets. However, the implementation of these smart biomedical systems is associated with numerous challenges, including the availability, accessibility, and usability of the historical datasets; the datasets being the intellectual property of healthcare providers who are reluctant to share them; and the quality of the datasets being incomplete, high-dimensional, and noisy, among others. Considering these challenges endemic to Africa, it is relevant to cast a light on the challenges afflicting the health systems and then explore them and how predictive models can address them.

Overview of machine learning

Machine learning offers an algorithmic solution to automated data analysis and models the numerical relationship between the outcome of a study and the data collected on relevant variables [1]. The trained ML algorithm learns from the data and can then predict the desired outcome based on the input data value with a declared uncertainty. Recent developments in this field have led ML algorithms to outperform human

specialists in analyzing complex datasets like medical images. Such models hold the promise of automating tedious routine clinical tasks and, in turn, improving healthcare access and quality, especially in developing economies. ML has become increasingly critical in the fields of perception, automation, and processing of well-structured data. It enables the efficient translation of natural text and voice, the categorizing of electronic pathology slides in minutes, the provision of infrastructure planning assistance, and the streamlining of supply chains [3]. Healthcare systems, analyzing discrete and continuous waveforms like electrocardiography (ECG), electroencephalography (EEG), and oximetry, could benefit greatly from the new possibilities of ML in signal processing. Still, before ML can process waveforms as accurately and reliably as text, there is a long way to go.

Applications in healthcare

In recent years, machine learning has become a cutting-edge technology across multiple sectors, including healthcare, where it has numerous applications and the potential to transform the health system from being less reactive to being more predictive. With the advancement of this technology, healthcare systems will be optimized so that diagnosis, treatment, and patient outcomes will be improved [2]. The healthcare systems can be optimized to improve the service provided to patients through maximizing the usage of available resources in the best way possible, in such a manner that ensures the service quality does not drop below a required threshold [4]. The objective is to alleviate specific challenges and needs in the African healthcare system by using machine learning technology. Challenges like diagnosing a disease, recommending medication, or predicting the emergence of a disease outbreak can be tackled. Some healthcare systems' needs, like resource allocation and emergency response, will also be covered. Thus, enabling the health system optimization and an increase in the quality of health services. Having an efficient health system can impact the patients' lives and outcomes significantly, specifically across timeliness, quality, and accuracy of treatment decisions.

Challenges in African healthcare systems

Healthcare is an ever-growing challenge worldwide, especially in Africa, where the healthcare systems often lack medical resources, infrastructure, trained professional healthcare personnel, incident reporting systems, or proper data retention protocols. This list of challenges poses several barriers preventing the increasing implementation of ML technologies in predictive healthcare. Due to these challenges, individuals cannot or are restricted in getting health services or providing the population with efficient health services [2].

A lack of access to medical resources results in healthcare systems' inability to provide high-quality medical services to a significant portion of the population. This may be due to patients' inability to seek medical attention at all due to geography restrictions or the lack of proper healthcare facilities and personnel. Patients who can access healthcare are prone to be examined and treated at a late stage of any medical

incident due to the necessity of traveling great distances or due to the healthcare service being free to the public. The late stage of medical incidents significantly hampers the ability for physicians to prescribe preventive or quick treatment options, resulting in the potential worsening of the medical incidents and greater complications [5].

Resource constraints

Healthcare systems in Africa largely suffer from resource constraints. Funding is limited, and low budgets per capita often lead governments to devote the vast majority of financial allocations to preventive health programs or, on a selective basis, to chronic diseases such as HIV [2]. Solo hospitals often have to raise funds to maintain their staff and infrastructure. Fortunately, some specialized hospitals and private clinics have better financial access depending on the structure of the local health system. Such limitations cannot allow for the view of how well-off healthcare is. Moreover, budget constraints tend to restrict improvements in the efficiency of existing health services, especially when resources are limited. Not having access to a general view of what constitutes a good health system makes it difficult to implement the more advanced machine learning methods on public databases whenever some level of aggregate records is available. Resource constraints are not just in terms of budgetary allocations. Many healthcare systems in Africa are poorly enforced. There exist at least fifteen different cadres of health personnel (doctors, medical officers, health assistants, health officers, etc.) with different powers and functions. Such a complicated structure in most of the industrialized Western countries and the Far East would not be acceptable. A clear hierarchy of medical personnel hardly exists, and laymen are often involved in critical decision-making. Health personnel with high qualifications often refuse to work in rural regions and would prefer careers outside of medicine. For instance, a health officer in Kenya has to supervise a catchment area of 50,000 people, and day-to-day requests often exceed his/her workload [6]. Such difficulties cannot go unnoticed in improving and implementing machine learning for predictive healthcare effectively. A valid data-driven healthcare model has to be based on knowledge of the health system and the rules of propagation between different competitive equilibria.

Infrastructure gaps

The aim of this section is to examine infrastructure gaps that currently exist in healthcare and healthcare delivery on the African continent. Infrastructure is described as the systems or facilities that are essential for a country or organization to function effectively. In the context of healthcare, infrastructure includes essential facilities and manpower systems to deliver efficient healthcare services. The provision of infrastructure in African healthcare has long existed, including biomedical technology, e-health, communication technology, telemedicine, Information, Education, and Communication (IEC), training manuals/materials, and transport systems. However, the infrastructure still has challenges, making it difficult to implement the proposed machine learning technology for predictive healthcare. The discussion on current gaps in healthcare infrastructure in the African continent is summarized below.

Africa still has the lowest healthcare infrastructure, paving the way to slow investment in science and health Research & Development (R&D), which hinders the growth of cutting-edge bioinformatics and biosystem analysis capabilities [5]. This is in addition to the large population that lacks access to basic and affordable health services. The World Health Organization (WHO) tracks country-based measures of health service coverage and access to essential medicines and commodities. The availability of Primary Health Care (PHC) services is crucial for achieving universal health coverage. Africa has the highest number of countries where fewer than 90% of the population lives within reach of such services [2]. Affordable and quality health services for all remain a challenge, although fast-growing countries are closing their gaps faster than war-torn or have gone through other shocks. There is a wide gap to bridge as some countries exhibit population coverage greater than 0.95, while others are below 0.5.

Beyond the general opportunities, specific healthcare challenges in Africa that hinder efficient health delivery include the absence of centralized health information systems, low Electronic Health Record (EHR) penetration, and weak surveillance infrastructures. Machine learning offers innovative ways to mitigate these issues. For instance, ML models have been deployed to optimize patient triage in overstretched clinics using historical symptoms and disease spread data. In South Africa, a neural network-based classifier was trained on chest X-rays to detect early signs of tuberculosis with over 87% sensitivity [5]. Likewise, in Ethiopia, a supervised learning model integrated antenatal data to predict maternal mortality risk, aiding frontline health workers in prioritizing high-risk pregnancies [2]. These examples highlight ML's potential to alleviate constraints, even when systemic gaps persist.

Benefits of predictive healthcare

Predictive healthcare is essential for efficient health systems, especially as the burden of diseases rises. Predictive health encompasses the frameworks and technologies for forecasting health issues and outcomes and making recommendations. It focuses on proactive measures to stop the occurrence of particular health conditions. Current healthcare systems are reactionary, with diagnostic and treatment processes initiated only during emergencies or when symptoms appear.

The onset and prognostic risk of diseases can be measured and predicted using collected health-related datasets from citizens. Adaptive machine learning algorithms can predict health issues based on historical datasets from health institutions. This allows for informed decision-making in preventive healthcare and the timely diagnosis and treatment of emerging diseases. Predictive analytics in public health refers to using collected datasets to track and forecast the spread and incidence of diseases, thereby enabling actionable measures before emergence [2].

The rise of ML capabilities has yielded numerous benefits in predictive healthcare. These benefits include predictive analytics in healthcare, predicting diseases before emergence, reduced healthcare costs, improved health outcomes, efficient utilization of resources, and overall efficiency in the healthcare system and institutions [7]. Predictive healthcare allows public health institutions to predict disease spread based on historical data on the spread and epidemiology of diseases, thereby anticipating potential outbreaks with proper action plans.

Early disease detection

Patients with various diseases, including cancer, diabetes, and illnesses triggered by infectious agents such as COVID-19, tuberculosis, and malaria, often undergo protracted and costly treatments for stage II and successive stages. Since they go unnoticed during their early stages, they are challenging to treat and control. The objective of a predictive healthcare system is to detect such diseases in advance and take counteractions at the desired moment, preferably before the higher stages from a personal and societal point of view. Machine learning efforts, especially supervised ML schemes, are feasible to detect an illness in the beginning stages, provided there exists sufficient data documenting the initial as well as subsequent phases [5].

For cancers of various organs, diabetes, and infectious diseases such as HIV, tuberculosis, malaria, etc., which currently have no cure, there are respective earlier stages of these diseases, which, if detected at a suitable time, can be treated with drugs or diagnostics, delaying progression to more advanced stages. For example, hospitals have machinery to detect a cancerous tumor in the brain and breast, as well as diagnostic apparatus for diabetes, tuberculosis, etc. Still, these devices are either utilized at advanced stages of the disease or require the operator to have expertise in pathological consultation. Generally, the screening process using such machines sends the information anonymously to an expert, who not only takes longer to report, but also involves extensive monetary expenditure and points out where the apparition develops. Given the above context, there is a need for a standalone device that can be individually customized to screen for conditions such as cancer, diabetes, and tuberculosis—without requiring clinical expertise or significant financial resources [2].

Resource optimization

Resource optimization is the process of ensuring that existing resources are utilized more efficiently. Healthcare systems in Africa face numerous issues regarding resource allocation and utilization. Patient numbers often exceed the capacity of healthcare facilities. For instance, there is projected to be a shortage of healthcare workers worldwide, with an estimated shortage of 18 million physicians, nurses, and midwives by 2030. Other than healthcare workers, there are various healthcare supplies, such as medicines, equipment, and transportation, that, if not utilized correctly, could lead to preventable loss of lives.

Machine learning has been used to combat these problems by taking a data-driven approach. ML-based strategies can be used to understand the influx of patients and their needs based on historical data, such as estimating the number of patients likely to visit a particular hospital daily or monthly and thus planning properly for the staff and medicines required. ML can also assist in designing an efficient staff schedule that considers the required working hours of individual workers while ensuring that enough workers are present at a particular time of the day. Moreover, ML can assist in maximizing the use of vehicles and even predicting time delays, such as in the distribution of medical supplies and patient transportations [8].

Machine learning models for healthcare prediction

There is an opportunity to utilize ML models to predict healthcare outcomes in African health systems. Machine learning has grown in interest over the years due to its importance in predictive analytics. There has been an effort in working with healthcare datasets to predict and assess heart disease mortality [9], and to predict other diseases, such as diabetes and stroke. If the prediction is accurate and successful, it can be incorporated into an expert system to advise doctors and improve patient outcomes. Two possible models that can be developed are identifying whether patients have heart diseases based on their characteristics and predicting whether a patient's health will deteriorate within the next 30 days based on their clinical attributes. The predicted outcome can support referring patients to specialists for further examination. With Africa facing healthcare shortages, ML can be used to optimize patient allocation within health systems [2]. The goal is to develop machine learning models to predict healthcare and to investigate and suggest ideal and attainable models for healthcare prediction in health systems in Africa. To meet this goal, there is a need to discuss the following: i) relevant datasets to predict healthcare, ii) modelling and evaluating healthcare prediction models, iii) potential healthcare prediction models and their application, and iv) the impact of healthcare prediction models on African health systems. African health systems refer to those in countries such as, such as Rwanda, Uganda, Zambia, South Sudan, and Tanzania.

Different machine learning models offer distinct advantages and trade-offs, particularly in African medical contexts. Decision trees are favored for their transparency and ease of interpretation by health workers without technical expertise, making them suitable for rural deployments. Random forests, a tree ensemble method, enhance predictive accuracy while still preserving interpretability. Support Vector Machines (SVMs), on the other hand, are effective for small to medium-sized health datasets, particularly in binary classification tasks such as disease presence or absence. Neural networks offer powerful non-linear modeling capabilities, useful for tasks such as interpreting chest X-rays or segmenting medical images; however, they often require large datasets and computational resources, which may not always be available. Hence, model selection must be guided by context, balancing accuracy, interpretability, and resource availability. In Malawi, for example, random forests have been used to predict maternal

health complications using minimal features with over 80% accuracy, while in Nigeria, SVMs trained on malaria diagnosis data achieved high sensitivity in community settings [2].

Supervised learning

In supervised learning, an ML model is trained to make predictions when given labeled training data. A model is provided with training examples consisting of input-output pairs. The inputs in the training set typically feature vectors containing multiple numeric or nominal attributes (independent variables). The output values (the dependent variables) are prediction targets. The training set teaches the model to predict output values when given input feature vectors. The goal is to construct a model that generalizes well to previously unseen examples. In this context, unseen examples refer to input feature vectors not included in the training set. In supervised learning, the input feature vectors correspond to a fixed-size data matrix. In contrast, the output values correspond to a vector with the same number of components as rows (i.e., examples) in the data matrix. Depending on the nature of the output values, supervised learning problems can be classified into two categories: regression and classification problems [3].

Application of Supervised Learning in African Healthcare Contexts: Disease prediction has been proven beneficial for patient health and wellness regarding disease prevention and timely treatment. Existing methods require expertise and deal with an extensive collection of data, making it challenging to predict diseases. Technology advancements enable the analysis of large amounts of medical and experimental data associated with patient health. The rapid growth in patient health records motivates the development of advanced technologies for effective chronic disease prediction. Machine learning methods can be applied to electronic health records to improve patient health awareness and assist in the prognosis and treatment of chronic diseases. African healthcare systems face challenges preventing computer-based disease prediction systems from being deployed, used, monitored, and maintained. However, disease prediction methods based on supervised machine learning classification algorithms require fewer resources and could drastically improve patient health and wellness [2].

Unsupervised learning

Unsupervised learning techniques enable the algorithm to draw inferences from datasets without labeled training examples. The goal is to model the underlying structure or distribution in the data to learn more about the data itself. A common application of unsupervised learning is clustering, which attempts to group similar datapoints or find a way to separate dissimilar datapoints [1]. Other examples include dimensionality reduction or density estimation. As this topic is tightly coupled with discriminative supervised learning, explaining the difference between supervised and unsupervised learning in more detail is beneficial since healthcare datasets are often not labelled [3]. Using unsupervised techniques, patterns or relationships within the data may be discovered that

would otherwise be hidden (e.g., relationships across patient time-series, or between clinically distinct disease patterns).

Data collection and preprocessing in health-care

Gathering data from various sources is a crucial first step in all predictive analytics applications [10]. It is well known that machine learning algorithms benefit from larger training datasets. Hospitals and healthcare organizations are increasingly careful to track the courses of their patients' diseases over time. Therefore, many epidemiological data points can generally be accumulated and stored over the years. However, in the case of African hospitals, some data types may be missing entirely, for instance, certain laboratory tests may not be performed locally, or health records may exist only in unaided notebooks. Data previously gathered outside the hospitals may not contain the required granularity or dimensions. Consequently, determining whether a sufficient number and variety of relevant data points exist may be rather involved and requires familiarity with the hospital's information systems and recording standards [11].

Preprocessing refers to the procedures required to make the acquired data tenable concerning the chosen predictive model and to ensure its reliability. This includes the assessment of the completeness and consistency of the data concerning the intended use. It is essential to ensure that sufficient data points are available and that clinically meaningful target labels can be extracted for supervised learning. Afterward, all raw data must be transformed from its original format into a uniform representation, which is, in general, a tabular database. Furthermore, post-hoc procedures aimed at dealing with missing values, outlier detection and processing, noise removal, aggregation, and temporal alignment may be required. Without these preprocessing steps, Poor-quality data leads to unreliable predictions.

Despite these challenges, it would be short-sighted to consider ML applications in Africa entirely premature. While poor digitalization hampers large-scale deployments, incremental solutions are emerging. Mobile health platforms (e.g., CommCare, OpenSRP) are enabling real-time patient data collection even in rural settings, while national initiatives like the DHIS2 expansion in Nigeria and Tanzania have strengthened health data infrastructure. Cloud-based repositories, supported by lightweight mobile interfaces, are bypassing traditional hardware dependencies. Furthermore, pilot programs in Kenya and Rwanda are training community health workers to serve as "digital stewards," collecting structured health data via tablets, thereby gradually bridging the digital divide. These localized strategies suggest a phased but realistic pathway toward ML readiness in African healthcare systems.

Data sources

The application of ML in predictive healthcare requires access to relevant data sources. These data sources can take many forms, and each category will be described in detail to highlight its importance.

Electronic Health Records (EHRs): EHRs provide a rich data source that can be used for machine learning in predictive healthcare. They contain a wealth of information about patients' medical history, diagnoses, medications, and treatment plans. However, EHRs can be complex and unstructured, making it challenging to extract meaningful insights. Despite this challenge, EHRs have been widely used in predictive health initiatives [2].

Medical Imaging: Medical imaging data, such as X-rays, MRIs, and CT scans, can be used for ML in predictive healthcare. These images can be analyzed to detect abnormalities and diagnose diseases. However, medical imaging data can be large and complex, requiring specialized knowledge and tools to analyze effectively [10].

Wearable devices: Wearable devices like fitness trackers and smartwatches can provide real-time data about individuals' health and well-being. This data can be used to develop machine learning algorithms that predict health issues before they become serious problems. However, the use of wearable devices in predictive health initiatives is still in its infancy and requires further research and development.

Other sources of data: Other data sources include social media, online health communities, and health-related apps. This data can provide insights into public health trends, disease outbreaks, and individuals' health risks. However, using these data sources in predictive health initiatives raises ethical and privacy concerns.

Data cleaning and transformation

Data cleaning, the procedure of removing or handling nonstandard elements from a dataset (e.g., outliers, duplicates, and null data instances), and feature engineering, the process of creating new variables from existing ones (e.g., merging or processing existing features into more informative ones), are sometimes underestimated by data science beginners [12]. In Section 5 of the guidelines - Data preparation, healthcare analytics aims to provide advice and detailed approaches on the various steps necessary to carry out the data cleaning and feature engineering processes in practice. It covers both general aspects of data cleaning, such as null data imputation, variable coding, and the detection of anomalies, and healthcare-specific applications, such as treating numerous biomedical signal datasets, free-text notes, and Vitamin D records. Thus far, it has focused almost exclusively on the application of ML to the health domain, which is an important emerging question with a growing number of influential publications and projects. Nonetheless, the rapidly growing health data sets pose major challenges in terms of data integration and interoperability [11]. The variety and complexity of health data require new solutions to enable its future use in various applications and to create new information and knowledge elicited from the current data assets. Multiple initiatives have recently been launched globally in various countries and global organizations to address these issues. Here, attention is focused on aspects related to data quality, ensuring that healthcare data is consistent, accurate, complete, and up to date.

To overcome the challenge of data scarcity and incompleteness in African healthcare contexts, several advanced strategies can be employed. Transfer learning enables model training using pre-trained weights from larger, global datasets and fine-tuning them on smaller, local datasets, reducing the need for vast training examples. This approach has shown success in radiology and dermatology applications in low-resource settings. Data augmentation techniques, such as introducing controlled noise, feature scaling, or image rotation (in medical imaging), can also expand the training set artificially. Additionally, synthetic data generation through Generative Adversarial Networks (GANs) or simulation-based modeling can help enrich datasets while preserving patient privacy. These techniques allow for robust model development even in environments with limited, fragmented, or sensitive data. The key steps involved in preparing health data for ML applications are summarized in Figure 1 below.

Ethical considerations in healthcare predictive modeling

The application of ML algorithms in healthcare predictive modeling holds immense promise for improving health outcomes and resource allocation in low- and middle-income countries, particularly in Africa. However, the emergence of complex ethical considerations and associated challenges with addressing these ethical dilemmas within the health systems of African nations is concerning. Although advances in ML approaches to predictive modeling of health indicators exist, current models for Africa do not reflect the continent's socio-economic and health indicators context [13]. Many such models lack transparency and interpretability in generating predictions or allocating health resources. Moreover, ethical issues may arise for African nations as priorities shift toward using ML systems for predictive healthcare. Such ethical dilemmas require careful consideration before implementing predictive models in any health system.
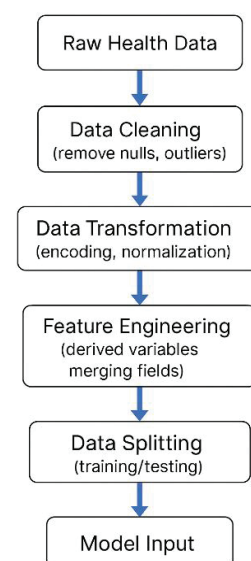

Figure 1: Data preprocessing workflow for predictive healthcare ML models, illustrating key steps from raw data cleaning to model-ready input.

Concerns around privacy and confidentiality persist, as machines need to access health data to predict events. What health data will be shared, and how? Who will have access to it? Many African countries have poor data sharing regimes, privacy laws, and regulatory frameworks. Will countries with such gaps in regulation be left behind in the new health data economy, or will their citizens bear the brunt of unhealthy and disadvantageous deals? In Africa, health data are often poorly digitized and cleaned, and curatorial decisions favor the more affluent with less regard for the poor, who may lack free access to the Internet and smartphones.

Additionally, it is difficult to ascertain what prior care decisions a model was trained on, and how this might shape predictions and influence future decisions. If a model is trained on biased data, it may neglect the health needs of underrepresented groups of rural population, instead predicting risks among wealthier urban patients. Who will take responsibility if health resources are not allocated to the poor in this target setting? Despite the obligation of the African state to care for all its citizens, implementing predictive models in any African health system would likely unintentionally perpetuate disparities against such patients [14].

Privacy and confidentiality

With advancements in machine learning-based predictive healthcare, it is vital to safeguard personal health information by observing privacy and confidentiality regulations [15]. Predictive healthcare analysis in a highly heterogeneous health data environment across institutions should comply with the local rules and regulations for patient privacy, confidentiality, and data handling. Data protection acts in respective countries, states, and organizations should be thoroughly reviewed at the feasibility stage of a use-case implementation to avoid legal drawbacks and breaches of patient trust [16]. While data security is more inclined towards technical issues, data privacy and confidentiality cover more ethical and legal dimensions. Data privacy includes retrieving, storing, and reusing data solely with the owner's consent. Data confidentiality denotes obligations to guarantee that the exchanged data does not reach unauthorized personnel. As such, it is essential to define the data protection rights at the feasibility stage of developing an ML model to analyze patient health data, especially in developing countries, where privacy infrastructure may be limited. If predicting diseases using health data from other organizations, the consent from the data owners needs to be taken with a clear understanding of what consequences it will have on the identified individuals. This includes permissions on data retention timelines, data re-exchanging with third parties, published outcomes, and any profitable utilization of the health data by the organizations involved. Even though the proposed solution complies with the above guidelines, the security of health data cannot be guaranteed.

Bias and fairness

Equity and fairness commitments are closely related concepts in discourses on the ethical dimensions of health predictive modeling fueled by machine learning. In part, this relates to how they both reference collectives rather than individuals: equity relates to collectives and fairness to interactions among members of those collectives [13]. Accordingly, a health predictive modeling approach might be biased but not unfair or equitable, such as an algorithm erroneously predicting individuals from a certain group as low-risk because of nonmedical characteristics, such as zip code (e.g., poverty/high mortality areas). Conversely, it is possible to have an equitable health predictive modeling approach that is not fair. For example, an algorithm reflecting and upholding existing health inequities might correctly predict individuals from certain populations as high-risk based on nonmedical proxies, thus ensuring that these individuals are unjustly disadvantaged. Although initially more concerned with fairness (e.g., model performance and algorithmic auditing), there is growing recognition that cooperation among stakeholders is needed to seed the initial conditions for fairness and equity to flourish [17]. This cooperation often involves de-privileging decision rights over data, algorithm training, model development, deployment, and the information triage powered by these activities in favor of disadvantaged groups. More generally, it is crucial to ensure the healthy coevolution of fairness and equity in health predictive modeling and machine learning. There may be more immediate violations of equity than concerns with fairness in this context; the latter might be utopian given the socio-historical trajectories of racial/ethnic and another disadvantage. While some privacy regulations exist worldwide, data governance in health predictive modeling broadly, and in algorithmic health predictive modeling in particular, remains largely unregulated. The absence of data and algorithms regulation undermines any attempt to establish fairness and equity in health predictive modeling. The initial conditions for any interaction in this fast-moving transnational landscape are uneven. Health predictive modeling represents a new platform for race, class, and other structural discrimination. Algorithmic bias is likely when data used to train models reproduces existing patterns of healthcare inequities, for example, data reflects socio-geographical determinants of access to care rather than the need for care.

Case studies of machine learning in African healthcare

Various computational algorithms have been developed to support healthcare functions over the years, such as Expert Systems (ES), machine learning, and Deep Learning (DL) [2]. Expert systems can help African healthcare systems diagnose patients and select treatment plans without extensively trained medical personnel. An example has been where an expert system has been incorporated with fuzzy logic systems to improve the diagnosis of chronic conditions in South Africa. The fuzzy inference system comprises 9 IF-THEN fuzzy rules that consider risk level variables: blood pressure, cholesterol level, body mass index, and age. Natural Language Processing (NLP) has also been used to develop a medical chatbot to diagnose patients in their early stages of disease. The chatbot is designed to narrow down a list of potential diseases based on the input of symptoms into the chat. Likita, a chatbot,

could diagnose common ailments and improve healthcare delivery in Africa. It uses knowledge-based and statistical-based approaches to tree-based and NLP parsing models.

Deep Learning (DL) is a subset of machine learning (ML) that can process large amounts of data and could aid medical workers in decision-making. For instance, X-ray images have been used to classify pneumonia and diagnose early-stage tuberculosis. The results found that a faster R-CNN is more accurate than the other two models for diagnosing malaria parasites. ML models have been developed to predict and classify chronic diseases in Africa. The models were developed using 6 classifiers with 8 RF tuners, then validated with 7 different metrics on 2 datasets, achieving up to 99.06% accuracy with 3 features and RF tuners on artificial neural networks [3]. African healthcare systems suffer from a lack of data access due to inadequate resources, which is often a key ingredient in the development and training of AI systems. AI in African healthcare has generally been used in disease mapping, such as HIV. Several ML techniques identify HIV predictors for screening and malaria predictions. Although African healthcare systems strive to implement AI, it is still difficult to respond to public health emergencies, resulting in increased mortality and morbidity.

Malaria prediction in Sub-Saharan Africa

Malaria, a mosquito-borne disease, is a major cause of morbidity and mortality in Sub-Saharan Africa and other parts of the world. Current malaria control and prevention efforts focus on diagnostic testing, curative treatment, and vector control, assuming that the disease occurs only in individuals who exhibit clinical symptoms. However, these efforts are undermined by late diagnosis, unavailability of health services, and shortages of antimalarial drugs. In this regard, the absence of a proactive approach and risk-based planning increases the burden of controlling the distribution of the disease [18].

Predictive healthcare systems have recently emerged as a new approach to improve health systems. These systems enable individual-level predictions about future patients based on their clinical history, health-related behaviors, and environmental factors. Predictive healthcare technologies are only at the initial stages of development, and attempts are being made to deploy systems for certain diseases [19]. In Africa, malaria is a potentially preventable disease, and predictive healthcare systems could control and prevent the spread of malaria by predicting risk levels based on environmental, demographic, and climate factors. A two-stage predictive healthcare system would enable sharing accumulated environmental, demographic, and climate data, allowing healthcare providers to make accurate predictions and allocate budgets efficiently. An overview of the malaria prediction framework using ML is illustrated in Figure 2 below.

Telemedicine applications

Telemedicine is among the applications that provide the benefit of distance service. The healthcare field has telemedicine applications that allow individuals to consult during medical emergencies, conduct medical diagnoses,
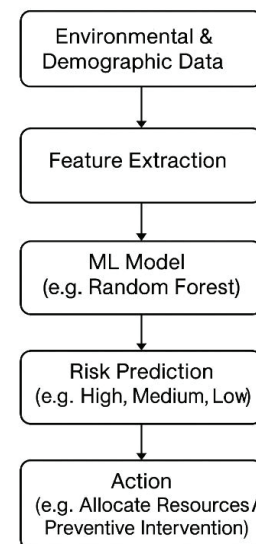

Figure 2: An illustration of a machine-learning-based malaria prediction framework in sub-Saharan Africa. The model uses environmental and demographic inputs to generate risk-level predictions that support public health interventions.

and seek advice on treatment and precautions to be taken without having to visit a hospital or clinic in person. Delivery of telemedicine services, in a wide range, to distant and out-of-reach areas is an open concern. Deploying resource-efficient telemedicine applications for healthcare according to the needs of patients and context is required, with predictive concepts utilizing machine learning algorithms. Analysis of data logs on telemedicine services delivered to patients can turn the problem of retargeting patient groups for telemedicine services into a classification problem. Predictive algorithms using telemedicine log data can be built to identify patients requiring telemedicine services, assuring superior patient care. Essentially, predictive telemedicine applications in remote areas can enable the health system to be resource efficient, and machine learning models can address the need to propose the telemedicine application appropriately [20].

In many African countries, access to healthcare services remains limited. In rural areas, the lack of transportation facilities and the cost incurred prevent individuals from seeking medical advice. In many communities, where specific diseases are prevalent, a proportion of the population remains without treatment. To narrow the gap in healthcare accessibility and achieve the goal of Universal Health Coverage effectively, a targeted healthcare delivery system can be best suited. The need for healthcare services or preventive measures of a disease can be predicted based on the past profile of patient treatment and diagnosis of a specific disease for a large, grouped population. The examination of individuals and recommending the diseases that should be considered further can be done before they consult with a physician. For this purpose, telemedicine applications can be opted for [2].

Future directions and emerging trends

Numerous obstacles impede the widespread adoption of health monitoring systems in Africa and other developing nations, including *i)* not readily available solutions, as many

systems and solutions are either crafted from scratch or specifically designed for Western countries. *ii)* a vicious cycle of poor deployment and utilization leading to negative perception of systems among political leaders and decision makers who therefore avoid future implementation, and *iii)* the risk of historical, fraudulent, and trust issues associated with potential foreign aid from government or private companies, which may impact local organizations trying to offer aid to improve health.

Nevertheless, with the rush toward the metaverse and the greater efforts to deploy the latest generation of satellite in orbit, there exist new opportunities for richer comparison of health metrics from around the world, including urban heat islands, access to transportation via roads and traffic, greenness and vegetation index, and many more. Many of these exist also as publicly available datasets which could be further exploited for rich comparison, also unlocking new health metrics which could be explored further for causation and creation of population exposure scores. The COVID-19 pandemic has also highlighted how the dynamics between behavior, transportation, and outbreaks could unlock novel signature characteristics to explore general questions in health and epidemiology further. Similar successful models explicitly accounting for behavior are emerging quickly in social media, online, and travel history domains, which can be used to develop further alternative and richer metrics for health monitoring systems and issues comparison.

With the instability engendered by the Ukraine crisis, new technologies aimed at exploiting new electric sources or boosting agriculture production are reaching the market. They could be deployed directly for health monitoring. Specifically, after the conflict, many countries are emptying the storage capacities built for the natural gas imports to appease the market price spike, and thus, proposals are emerging to convert these facilities into green hydrogen manufacturing plants. The idea is that through existing gas tubes, the hydrogen could be transported through partly pumped through pipes, and the rest could be mandated for fuel conversion of existing gas plants, which could be converted to burn hydrogen almost seamlessly. If these concepts materialize, within a couple of years, huge amounts of electricity will be required either in African countries or EU countries closer to the Sahara from solar panels and wind farms.

Interdisciplinary research opportunities

Advances in computing and internet connectivity, has decreased the cost of storing vast amounts of data. The availability of large and multidimensional datasets, along with sophisticated algorithms (e.g., machine learning methods), has enabled computers to evaluate data without the need for explicit settings and/or criteria on the part of a researcher. Using ML methods, computers can infer, improve, and evolve their decision-making processes based on the evaluated data. Machine learning methods have been successfully demonstrated in a variety of domains and, therefore, can be utilized for predictive healthcare [21]. Predictive healthcare is defined here as employing ML methods to utilize available

health-related data and improve health systems. Therefore, there exist inter-disciplinary research opportunities in establishing the appropriate meaning and design of predictive healthcare systems [2].

ML focuses on developing algorithms that allow computers to learn from previously acquired data, rather than following a preset of an explicit decision-making scheme. ML methods can, under appropriate model complexity and dataset size, discover implicit patterns/trends from data, enabling computers to assess, predict, and/or control the behavior of complex systems. The availability of large multidimensional datasets, particularly biomedical and chemical data, along with the increase in computer processing power, has triggered the increasing use of machine learning algorithms to evaluate data containing relevant information that is beyond human capabilities. Recently, machine learning methods have been applied in the biomedical and environmental research fields, including predictive modeling of various diseases (e.g., lung, liver, obesity, cancer, etc.) based on complex mRNA, microRNA, and metabolomic data.

Integration of wearable technologies

The rapid growth of health-related wearables has triggered research on their reliability and accuracy, which are issues to consider when implementing them in mobile Health (mHealth) environments [22]. Most technologies currently used in health-related wearables were originally developed for other purposes, such as marketing studies in retail (e.g., accelerometers). Thus, the algorithms and usage conditions are often proprietary, limiting transparency, and there are concerns about proprietary companies controlling massive volumes of sensitive health information. Risky situations can arise from a lack of transparency about how data is processed, sold, or used. At present, none of the technology-based companies commercially involved in health wearables provide the conditions for using these technologies. This can be highly problematic and unacceptable if proprietary technologies influence diagnosis or critical health decisions and generating influence critical decisions regarding patients' health care [23].

References

1. Marcinkevičs R, Ozkan E, Vogt JE. Introduction to Machine learning for physicians: A survival guide for data deluge. ETH Zurich; 2022. Available from: https://www.research-collection.ethz.ch/bitstream/handle/20.500.11850/588804/1/Intro_to_ML_for_Physicians_Preprint.pdf

2. Etori N, Temesgen E, Gini M. What we know so far: Artificial intelligence in African healthcare. arXiv; 2023. Available from: https://arxiv.org/pdf/2305.18302

3. Habehh H, Gohel S. Machine learning in healthcare. Curr Genomics. 2021;22(4):291–300. Available from: https://pubmed.ncbi.nlm.nih.gov/35273459/

4. Mashraqi AM, Allehyani B. Current trends on the application of artificial intelligence in medical sciences. J Family Med Prim Care. 2022;11(8):4567–72. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10484692/

5. Williams D, Hornung H, Nadimpalli A, Peery A. Deep learning and its application for healthcare delivery in low and middle-income countries. Front

Artif Intell. 2021;4:645447. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8117675/

6. Sendra-Balcells C, Campello VM, Torrents-Barrena J, Ahmed YA, Elattar M, Ohene-Botwe B, et al. Generalisability of fetal ultrasound deep learning models to low-resource imaging settings in five African countries. Sci Rep. 2023;13:2728. Available from: https://europepmc.org/article/pmc/pmc9932015

7. Dixon D, Sattar H, Moros N, Kesireddy SR, Ahsan H, Lakkimsetti M, et al. Unveiling the influence of AI predictive analytics on patient outcomes: A comprehensive narrative review. Cureus. 2024;16(5):e12345. Available from: https://doi.org/10.7759/cureus.59954

8. Chung TH, Rostami V, Bastani H, Bastani O. Decision-aware learning for optimizing health supply chains. arXiv; 2022. Available from: https://arxiv.org/abs/2211.08507

9. Mpanya D, Celik T, Klug E, Ntsinjana H. Predicting mortality and hospitalization in heart failure using Machine learning: A systematic literature review. Int J Cardiol Heart Vasc. 2021;34:100773. Available from: https://pubmed.ncbi.nlm.nih.gov/33912652/

10. Taoussi C, Hafidi I, Metrane A. Prediction of medical pathologies: A systematic review and proposed approach. Int J Online Biomed Eng. 2023;21(2):121–36. Available from: https://online-journals.org/index.php/i-joe/article/view/52639

11. Maletzky A, Böck C, Tschoellitsch T, Roland T, Ludwig H, Thumfart S, et al. Lifting hospital electronic health record data treasures: Challenges and opportunities. JMIR Med Inform. 2022;10(10):e38557. Available from: https://medinform.jmir.org/2022/10/e38557/

12. Chicco D, Oneto L, Tavazzi E. Eleven quick tips for data cleaning and feature engineering. PLoS Comput Biol. 2022;18(12):e1010718. Available from: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1010718

13. Paulus JK, Kent DM. Predictably unequal: Understanding and addressing concerns that algorithmic clinical prediction may increase health disparities. NPJ Digit Med. 2020;3:99. Available from: https://www.nature.com/articles/s41746-020-0304-9

14. Chen IY, Pierson E, Rose S, Joshi S, Ferryman K, Ghassemi M. Ethical Machine learning in health care. Annu Rev Biomed Data Sci. 2020;3:123–44.

Available from: https://www.annualreviews.org/doi/10.1146/annurev-biodatasci-092820-114757

15. Thapa C, Camtepe S. Precision health data: Requirements, challenges and existing techniques for data security and privacy. Comput Biol Med. 2020;129:104130. Available from: https://www.sciencedirect.com/science/article/pii/S0010482520304613

16. Murdoch B. Privacy and artificial intelligence: Challenges for protecting health information in a new era. BMC Med Ethics. 2021;22:122. Available from: https://bmcmedethics.biomedcentral.com/articles/10.1186/s12910-021-00687-3

17. Raza S. Connecting fairness in Machine learning with public health equity. arXiv; 2023. Available from: https://arxiv.org/abs/2304.04761

18. Modu B, Polovina N, Lan Y, Konur S, Asyhari A, Peng Y. Towards a predictive analytics-based intelligent malaria outbreak warning system. Appl Sci. 2017;7(8):836. Available from: https://www.mdpi.com/2076-3417/7/8/836

19. Oluwagbemi O, Clarence Y. Computational predictive framework towards the control and reduction of malaria incidences in Africa. Egypt Comput Sci J. 2012;36(2):1–17. Available from: http://ecsjournal.org/Archive/Volume36/Issue2/1.pdf

20. Khan R, Kumar Srivastava A, Gupta M, Kumari P, Kumar S. Medicolite—Machine learning-based patient care model. J Healthc Eng. 2022;2022:8109147. Available from: https://www.hindawi.com/journals/jhe/2022/8109147/

21. Beyene J, Harrar SW, Altaye M, Astatkie T, Awoke T, Shkedy Z, et al. A roadmap for building data science capacity for health discovery and innovation in Africa. Front Public Health. 2021;9:710961. Available from: https://www.frontiersin.org/articles/10.3389/fpubh.2021.710961/full

22. Wall C, Hetherington V, Godfrey A. Beyond the clinic: The rise of wearables and smartphones in decentralising healthcare. NPJ Digit Med. 2023;6:37. Available from: https://www.nature.com/articles/s41746-023-00971-z

23. Naseri Jahfari A, Tax D, Reinders M, van der Bilt I. Machine learning for cardiovascular outcomes from wearable data: Systematic review from a technology readiness level point of view. JMIR Med Inform. 2022;10(1):e29434. Available from: https://medinform.jmir.org/2022/1/e2943