Data science training needs in sub-Saharan Africa: Implications for biomedical research and therapeutics capacity

Article *in* Open Research Africa · November 2023

DOI: 10.12688/openresafrica.14109.1

14 authors, including:

Modupe Coker
Rutgers, The State University of New Jersey

Olayinka Kotila
University of Ibadan

Temitayo Lawal
University of Ibadan

Adeniyi Fagbamigbe
University of Ibadan

REVIEW

Data science training needs in sub-Saharan Africa: Implications for biomedical research and therapeutics capacity [version 1; peer review: awaiting peer review]

Modupe O. Coker[1-3], Olayinka A. Kotila[4,5], Taiwo O. Adigun[6,7], Temitayo V. Lawal[3,8], Adeniyi F. Fagbamigbe[8], Angela U. Makolo[6,7,9], Mobolaji M. Salawu[8], David T. Ajayi[8,10], Paul E. Oluniyi[11,12], Judith U. Oguzie[11,12], Anise N. Happi[11], Anne G. Hoen[2], Chinedum P. Babalola[4,5,10,13], IkeOluwapo O. Ajayi[8,14]

[1]Department of Oral Biology, School of Dental Medicine, The State University of New Jersey, Rutgers, New Jersey, USA
[2]Department of Epidemiology, Geisel School of Medicine, Dartmouth College, Hanover, New Hampshire, USA
[3]Institute of Human Virology Nigeria, Abuja, Federal Capital Territory, Nigeria
[4]Centre for Drug Discovery, Development and Production, University of Ibadan, Ibadan, Nigeria
[5]Department of Pharmaceutical Chemistry, Faculty of Pharmacy, University of Ibadan, Ibadan, Nigeria
[6]Department of Software Engineering, School of Computing and Engineering Sciences, Babcock University, Ilishan-Remo, Ogun State, Nigeria
[7]University of Ibadan Bioinformatics Group (UIBG), University of Ibadan, Ibadan, Nigeria
[8]Department of Epidemiology and Medical Statistics, Faculty of Public Health, College of Medicine, University of Ibadan, Ibadan, Nigeria
[9]Department of Computer Science, Faculty of Science, University of Ibadan, Ibadan, Nigeria
[10]College of Basic Medical Sciences, Chrisland University, Abeokuta, Ogun State, Nigeria
[11]African Centre of Excellence for Genomics of Infectious Diseases (ACEGID), Redeemer's University, Ede, Osun, Nigeria
[12]Department of Biological Sciences, Faculty of Natural Sciences, Redeemer's University, Ede, Osun, Nigeria
[13]Genetic and Bioethics Unit, Institute of Advanced Training and Medical Research & Training (IMRAT), College of Medicine, University of Ibadan, Ibadan, Oyo, Nigeria
[14]Epidemiology and Biostatistics Research Unit, Institute for Advanced Medical Research and Training (IMRAT), College of Medicine, University of Ibadan, Ibadan, Oyo, Nigeria

v1 **First published:** 17 Nov 2023, **6**:21
https://doi.org/10.12688/openresafrica.14109.1
Latest published: 17 Nov 2023, **6**:21
https://doi.org/10.12688/openresafrica.14109.1

Abstract
Data use is becoming increasingly valuable worldwide and has positively improved decision-making in various sectors. Data science has revolutionized the study of epidemiology and disease control. Despite the unlimited potential of data science, Africa lags in its innovation and technological advancements. Although sub-Saharan Africa (SSA) has not taken full advantage of data science in healthcare delivery and biomedical research, attempts have been made to harness this opportunity. This article reviews the current level,

Open Peer Review

Approval Status *AWAITING PEER REVIEW*

Any reports and responses or comments on the article can be found at the end of the article.

potentials, and gaps in data science, genomics, and bioinformatics in SSA. We also identified needs associated with developing tools and building capacity in the region, while also acknowledging advances made in these areas to date. We identified opportunities in health data science in SSA, including benefits in combating diseases that burden this region. Applying data science has the potential to create comprehensive healthcare reports and convert them into relevant critical insights that can then be used to provide better care, reduce costs of treatment, predict outbreaks of epidemics, avoid preventable diseases, and improve quality of life in general. In order to achieve this, artificial intelligence and machine learning to build and apply models are some of the common mechanisms applied in all domains of medical research, alongside computational tools and models to simulate systems or expedite biomedical research like bioinformatics and computational biology. Data science, a body of knowledge that uses modern tools and techniques for data management and utilization of big data for innovative interventions and biomedical research is upcoming in Sub-Saharan Africa. Its importance in understanding epidemiology of diseases, public health, and surveillance as well as other domains in allied medical disciplines were highlighted in this review. The dearth in data science skills, expertise, and institutions engaged in training data science needs urgent intervention to optimize the potential uses to improve medicine and healthcare delivery in SSA.

Keywords
Data science, sub-Saharan Africa, Genomics, Epidemiology, Microbiome (microbiota), Pharmagenomics, Biomedical research

Corresponding author: Modupe O. Coker (dupe.coker@rutgers.edu)

Author roles: Coker MO: Conceptualization, Project Administration, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Kotila OA**: Conceptualization, Data Curation, Writing – Original Draft Preparation, Writing – Review & Editing; **Adigun TO**: Writing – Original Draft Preparation, Writing – Review & Editing; **Lawal TV**: Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; **Fagbamigbe AF**: Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Makolo AU**: Writing – Original Draft Preparation, Writing – Review & Editing; **Salawu MM**: Writing – Original Draft Preparation, Writing – Review & Editing; **Ajayi DT**: Writing – Original Draft Preparation; **Oluniyi PE**: Writing – Original Draft Preparation; **Oguzie JU**: Writing – Original Draft Preparation; **Happi AN**: Supervision, Validation, Writing – Review & Editing; **Hoen AG**: Supervision, Validation, Writing – Review & Editing; **Babalola CP**: Supervision, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Ajayi IO**: Supervision, Validation, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

How to cite this article: Coker MO, Kotila OA, Adigun TO *et al.* **Data science training needs in sub-Saharan Africa: Implications for biomedical research and therapeutics capacity [version 1; peer review: awaiting peer review]** Open Research Africa 2023, **6**:21 https://doi.org/10.12688/openresafrica.14109.1

First published: 17 Nov 2023, **6**:21 https://doi.org/10.12688/openresafrica.14109.1

Introduction

Data science is one of the world's fastest-growing inter-disciplinary fields that uses analytical and scientific methods, processes, algorithms and systems to extract information, knowledge and insights from both complex structured and unstructured data. The field of data science was birthed due to broad interest across many fields and applications. It is a field that acts as an umbrella, bringing together a varying number of disciplines using a common set of computational approaches and techniques[1–3]. Data science incorporates data architecture, data mining, data analytics, big data analytic tools as well as machine learning (ML) and artificial intelligence (AI)[1]. In particular, data science in health provides a unique opportunity to combine biomedical sciences, epidemiology, public health, computer sciences, bioinformatics, statistics and mathematics to transform data into meaningful and useful information. Data science is also applied to various other fields like education, urban development, and transportation among others.

The world is currently experiencing a rapid evolution of technologies, and this has brought about increased interconnectivity between researchers, sharing of data and generation of large amount of information daily[4,5]. Data is fast becoming the world's most valuable resource, and has positively impacted human thoughts and approach to the world, enabling better decision making in all sectors[5,6]. Many industries and sectors have embraced data science to replace or compliment traditional methods, with disciplines such as engineering, medicine, social science, business, science, and the humanities[4]. Data science has brought about powerful approaches to the study of epidemiology and control of diseases including the most recent COVID-19 epidemic[7].

In Africa, data science has an unlimited potential due to the availability of huge data generated daily through fast rising populations, a high level of genetic diversity and disproportionate burden of infectious disease as well as increasing prevalence of non-communicable diseases such as cancers and cardiovascular diseases[1]. However, in Africa, innovation and technological advancements have lagged[1,6]. Although some countries such as South Africa, through multi-country organizations like H3Africa, have a very strong capability in data science, this has not been replicated across African regions, as many countries have developed little or no capacity, knowledge and expertise in data science; especially in the sub-Saharan region[1,8].

Sub-Saharan Africa (SSA), home to over 46 countries and 1.1 billion people[9], has struggled with limitations in developing, implementing and sustaining innovative technologies and ideas that are central to the general development of the region[1,10]. Compared to other developed nations, it has been left behind in the data revolution[11] mainly due to poor infrastructure, limited training opportunities and lack of expertise in the field/discipline[1,11–13].

In order to harness the advances in bioinformatics and data science research, it is crucial to train the next generation of African scientists on its use and benefits[14,15]. One of the major challenges to the advancement of bioinformatics in Africa, as noted by Bishop *et al*. in 2015, is the shortage of trained bioinformaticians[16,17]. Doumbo and Krogstad also noted that availability of doctoral training on advanced topics in data science is necessary for African countries to establish and implement their own health priorities[18,19]

In order to ensure the effectiveness of bioinformatics and data science in SSA and benefit from advances in research, the potentials in technological advancement must be harnessed and modern applications of analytics must be implemented. It is also imperative to train the next generation of African scientists on the use of data science and bioinformatics while encouraging doctoral training on advanced topics to define and implement their own health priorities[16,18,20].

Data science has the potential to address many of the continent's challenges, and for sustainability, it is pertinent that African scientists themselves lead this drive. This will ensure the implementation of data science across all educational and societal strata. The need to strengthen capacity for data science in SSA cannot be overemphasized. This narrative review assesses the current level of data science training capabilities in Africa, and the potentials, and highlights the gaps in data science, genomics and bioinformatics in the sub-Saharan region of Africa as compared to the advances made in other developed nations. In addition, it identifies the needs of the region in terms of tools and capacity building, while simultaneously identifying the opportunities available and advances that the region has made in this regard. Finally, we suggest a sustainable approach, and posit recommendations geared towards the development of data science capabilities in SSA.

Data science and medicine

Opportunities for biomedical data science in SSA

Infectious diseases. Emerging and re-emerging infectious diseases remain a global public health threat causing high rate of illnesses and fifteen million deaths in 2015[21]. This has a profound economic impact on countries, in terms of healthcare costs and loss of productivity especially with the recent wave of pandemics which is ever present and global[22]. The SSA region suffers an enormous burden of infectious diseases such as malaria, HIV, tuberculosis and epidemic-prone disease such as cholera; with significant morbidity and mortality[23] and associated high economic impact. Infectious/communicable diseases account for 45% of the death in SSA and are the leading causes of death in most African countries[24–26]. Nigeria is the most populous country in Africa (with a population of 200 million) with a disproportionately high burden of communicable diseases among countries in SSA[27]. Such diseases of importance include the core diseases of poverty (malaria, HIV/AIDS and tuberculosis) and other poverty-related diseases including acute respiratory infections (ARI), measles, and diarrhea[28]. Very prevalent in the country, especially in the rural areas, are the neglected tropical diseases such as filariasis, onchocerciasis, trachoma, worm infestation, schistosomiasis, and leprosy[29].

Viral hemorrhagic fevers such as Lassa fever, Yellow fever, Ebola virus disease, COVID-19 and others known to cause epidemics have also resulted in significant morbidity and mortality[23,30]. The risk of dying from these infectious diseases is highest in low and middle-income countries (LMIC), mostly of WHO African Region and South-East Asia Region[31]. The United Nations - Sustainable Development Goals, target three (SDG-3) aims to end the epidemic of AIDS, tuberculosis, malaria and neglected tropical diseases and combat hepatitis, water borne disease and other communicable diseases by 2030[29]. These diseases accounted for an estimated 5.3 million in 2000 (2.2 million female and 3.1 million male) and went down to 4.3 million deaths in 2016 (1.7 million female and 2.7 million male)[22].

In 2021, the incidence of HIV infection declined from 2.2 million reported in 2010 to 1.5 million new incidences globally[32]. However, SSA remains the most heavily affected region, with women being the most infected[32]. There were estimated ten million cases of tuberculosis in 2017, with deaths increasing with age[33]. Malaria remains the leading cause of death and disability in SSA with young children and pregnant women being the most vulnerable[34]. Other causes of death among children are vaccine-preventable diseases (VPDs) such as measles, poliomyelitis, pneumonia, and meningitis[35]. Yearly reports show that over 30 million African children under the age of five suffer from VPDs and more than half a million die due to limited access to immunization services[35]. About 1.58 billion people (85%) required interventions against neglected tropical diseases in 2017[22].

The advent of the COVID-19 pandemic is certainly an unwelcome event in LMIC in SSA whose health systems are already deteriorating from poor resources in terms of financing, access, and severe shortage of skilled health manpower. The impact of this is obvious as prevalence of diseases increase, especially with the current double burden of diseases being experienced in Africa, and the associated high morbidity and mortality. It is therefore essential to emphasize the importance of real-time response to disease outbreaks in curtailing the spread and magnitude of epidemics/pandemics. The effectiveness of use and impact of real time outbreak response is dependent on how quick information/data can be collected, the frequency, completeness, with prompt analysis, and interpretation to inform a public health action. This in turn depends on the availability and functionality of modern analytical resources which depend on the technological capacity of the country. Data science, is vital for prompt and deep understanding of the epidemiology of diseases; prediction of future occurrences of epidemics and disease burden. It is the authors' opinion that this will enhance accurate and appropriate channeling of scarce resources in the region.

Non-communicable diseases. There is an increasing prevalence of non-communicable diseases (NCDs) and their risk factors in SSA owing to the ongoing demographic and epidemiological transition[36]. These diseases include diabetes mellitus, chronic kidney disease, chronic respiratory disease, cardiovascular diseases, cancers, road traffic injuries, substance use and mental disorders[37]. NCDs are the leading drivers of morbidity and mortality worldwide and account for 23% of deaths in Africa[37–39]. The WHO has projected that SSA is to experience the largest increase in mortality of about 42% from cardiovascular disease, cancer, respiratory disease and diabetes by 2030[40]. This projected increase is as a result of the NCDs risk factor surveillance report which indicated that most adults in SSA are exposed to at least one risk factor for NCDs including harmful alcohol use, tobacco consumption, unhealthy diet, physical inactivity, and obesity[38].

The health systems of African nations are unable to effectively combat this high burden of NCDs as a result of challenges with resources in terms of lack of infrastructure, shortage in capacity as well as skill and poor record keeping[41,42]. Meanwhile, there are data from various sources such as health facilities, researchers and health industry which are not curated or utilized[12,38]. These challenges constitute a huge obstacle in understanding the epidemiology of NCDs and consequently proper management and preventive measures.

Data science through data curation, mining, analytics, and big data tools as well as ML, provides a means to better understand the epidemiology, management and prognosis of NCDs. This will enable the health care provider make timely and accurate, evidence-based decisions for patient care. Prompt diagnosis and treatment of these diseases will significantly reduce the cost of care and disease burden in patients[43]. Systematic reviews reported that applications of big data analytics have proved useful in NCDs management, especially in terms of risk prediction and modelling, diagnostic accuracy, cost, hospital re-admission reduction, and patient-outcome improvement[44].

Public health surveillance. Public health surveillance is an ongoing, systematic collection, analysis and interpretation of health-related data essential to the planning, implementation, and evaluation of public health practice. Surveillance involves prompt collection of data on disease occurrence and public health related events closely integrated with the timely dissemination of these data to those responsible for preventing and controlling disease and injury[45,46]. The key objective is to provide scientific and factual database information essential to informed decision making and appropriate public health action.

The World Health Organization (WHO), African region, adopted the Integrated Disease Surveillance and Response (IDSR) strategy as a regional strategy for disease control in 1998 following the emergence and re-emergence of meningitis, cholera, yellow fever and measles[23]. The IDSR is a strategy and a tool to strengthen public health surveillance and response by promoting rational use of resources through integrating and streamlining common disease surveillance activities[47]. Due to the weak uncoordinated system of disease reporting and surveillance at the time, the government of member states of the WHO African region, with the assistance of the WHO secretariat supported countries to improve their disease surveillance and response capabilities so that they could detect and timely respond to communicable disease threats[47,48].

Yearly, over 100 infectious disease outbreaks and other public health emergencies occur in the WHO African region and generally speaking this affects all countries[45,49,50]. Since surveillance in Africa lacks adequate and resilient systems and suffers from systemic failure due to lack of political will, poor health infrastructure, service delivery and health management information systems with associated lack of human resource, the authors opine that infectious disease outbreaks and other public health emergencies are expected to continue unpredictably. Consequently, there is limited preparedness for controlling the devastating health events. SSA countries have limited resources to support data management systems[51]; however, with multiple epidemics documented over the past decades and growing burden of NCDs, there is a need for sustained, reliable data surveillance for targeting resources and evaluating programs. Data science is crucial in gathering large health information/data on health trend for predictive analysis to inform decision-making especially in the face of unprecedented disease outbreaks and increasing NCDs.

Application of data science in medicine

Modern technologies employed in medicine and healthcare generate massive data, which pose new challenges to today's treatment delivery methods. For example, pharmacogenomic data can be predictive of an individual's disposition to treatment outcome. This capacity has therefore revolutionalised medical treatment from being generalized to being personalized. Applying data science to the massive amount of data generated on a daily basis can help to create comprehensive healthcare reports and convert them into relevant critical insights that can then be used to provide better care, reduce costs of treatment, predict outbreaks of epidemics, avoid preventable diseases, and improve the quality of life in general. Healthcare delivery is now data-driven, making healthcare industry a major beneficiary of the emergence of data science; hence, medical diagnostics are becoming more efficient and accessible, medical treatment more personalized, and medical research more data-driven[1,52]. Data science is also driving improvements to/in the drug discovery process, personalized medicine and accurately interpret diagnostic images without human guidance, by designing algorithms and developing tools capable of ingesting large amounts of medical data to these processes[12]. Clinical and research data involve various types of datasets which could be used separately or collectively depending on the problem at hand. Data used in medicine and healthcare research include and are not limited to the following: i) Electronic health data: Demographic data are important for patients' identification, determining the distribution of age and gender, and communication with patients as well as other non-clinical data[53,54]. This data includes electronic health and medical records often termed EHR and EMR respectively. The health and medical records constitute the clinical data which are either collected during ongoing patient care or as part of a formal clinical trial program; ii) Genomic data: This describes the genetic content of the patients in form of DNA and RNA sequences, protein structure, genetic variant, and gene expression among others; iii) Medical imaging data: Features of interest can be extracted for ML or AI algorithms in a specific type of diagnostic image acquired with scanning devices, such as X-ray, computed tomography (CT) and magnetic resonance imaging (MRI)[55].

Benefits of data science in healthcare delivery and medical research

The numerous challenges faced in health care delivery and medical research in the continent can be mitigated with investment in building data science capacity[19]. If adequately implemented, data science improves efficiency and effectiveness of health care delivery[53,56]. It is very useful in managing multi-centre and multi-stage health intervention. Data science enhances precision, accuracy, timeliness and reliability of information irrespective of the type of data. Besides, data science enriches data. Due to versatility of data science, it can be applied widely, used in different settings, applied to different types of data, as well as different health outcomes. Data science enhances and speeds up medical research processes and the accuracy, as well as reproducibility, of the findings.

Improvement in diagnosis and patient healthcare are at the centre of many benefits of data science to medicine, which are expressed in different forms such as precision medicines, reduction of prescription errors and readmissions, reduction of healthcare costs and drug development, among others[52,57]. In research and development, data science uses big data analytics to come up with actionable insights to improving patients' outcomes. Through ML and AI techniques, data science is able to make use of the enormous volumes of big data either in a scalable or/and integrated format to discover hidden pattern and converted into useful, and actionable information that continues to deliver great insights to doctors, healthcare providers and patients in an effective way[58].

Tools for data science

Data science including ML and AI are common mechanisms of solving computational problems. They are currently applied in various fields such as entertainment, finance, education and medical research. Several computational tools and models have been developed and applied generally in biology using these mechanisms either to simulate a system, automate a system or to expedite biomedical research[59–61]. The process gave birth to disciplines called computational biology and bioinformatics, that is, the application of computational knowledge in solving biological problems. These fields of study essentially involve quantitative analysis of information relating to biology with the aid of computers. Data science is the use of scientific techniques and algorithms to extract patterns and knowledge from both structured and unstructured data[62]. Hence, data science is the researching, building, development, implementation and interpretation of models and algorithms. The astronomical generation of data with today's always-improving technologies has exposed the need and efficacy of computer applications in handling the resultant 'big data'.

Effective healthcare delivery requires knowledge from data science to inform medical diagnostics, treatment and research. Healthcare informatics solutions have been mainly developed in high-income countries[63,64]. Due to resource limitations, SSA has had limited opportunity to develop and benefit from the application of data science to health care challenges[65]. The focus here is to assess the computational efforts on healthcare delivery in developing African nations through data science.

Medical-oriented data science research are in high demand giving rise to repositories containing coding tools devoted entirely to facilitate genomics-focused medical research. Examples are Bioconductor, Biopython, Bioperl, BLAST and a special operating system called Biolinux. In fact, a catalogue of bioinformatics tools according to specific functions can be found here. This shows a strong relationship between the medical and data science communities but unfortunately, the authors believe that this has not been fully harnessed in SSA. Machine Learning and deep learning algorithms have played essential roles in developing models and tools in various areas of medical research where some are targeted to specific diseases. A few of these models and tools, are discussed in this study.

Several review studies have assessed opportunities and challenges of data science in medicine and medical research with the instrumentalities of ML and deep learning[66,67]. AI's future role in medical practice has also been exposed through many studies supposedly to optimize the care trajectory of chronic disease patients[12]. This is with the view of speeding up processes and achieving greater accuracy in making diagnosis, opening the path to providing better healthcare, suggest precision therapies for complex illnesses, reduce medical errors, and improve subject enrollment into clinical trials[68–70]. Challenges specific to data science and medicine have also been identified such as data pre-processing, actual clinical problem-based model training and refinement, ethical issues, and data privacy and security[71]. Another important and recent application of data science methods is Genome Wide Association Studies (GWAS) with increasing number of models and tools using genetic variant data[72].

Enhancing Data Science in Sub-Saharan Africa

Sub-Saharan Africa nonetheless has not taken full advantage of data science in healthcare delivery and biomedical research, but some attempts have been made in harnessing this opportunity. It is highly essential to take a deep look into the opportunities, challenges, and the future of data science in healthcare delivery and biomedical research in SSA due to some peculiarities such as highly genetic diversified population, environmental conditions, income, unstable government policies, inadequate technical know-how and power problem. For instance, critical reviews about the state and direction of data science and big data analytics in healthcare industry are described previously[1,63,73,74].

Akinnagbe *et al*. opined in their review that big data analytics applications in healthcare are still at its infancy in Africa despite the fact that platforms and technologies used to amass big data are already in use in Africa, and reporting success stories of its application can help fast track the adoption of data science on the continent[63]. Meanwhile, Owoyemi *et al*. critically considered challenges surrounding the implementation of AI in healthcare in Africa such as data availability and quality, legal and policy issues, costs of application and adoption of AI, and inadequate infrastructure[74]. The authors noted that AI offers tremendous promises in transforming and improving healthcare in low-resource areas like Africa by building intelligence into existing systems and institutions rather than attempting to start from scratch or hoping to replace existing systems. Achieving this will be possible if some critical measures are taken into consideration such as accelerated improvement in Africa's infrastructure (reliable power supply and affordable Internet services). In addition to this, Africa's healthcare sector needs to up her game by inclusion of electronic medical records as against use of traditional methods of handwritten documentation of medical issues related to the patient. There is also the need for a critical build up of data scientists who are specialised in such fields as computational biology, bioinformatics, genomics, and applied mathematics. The governmnet on its part will need to institute policies that will enhance cooperation among governments and states, trade, and commerce and will be instrumental in achieving these goals. Shaffer *et al*.[1] were able to justify one of these measures by assessing the impact of capacity development in bioinformatics and data science training programs in SSA, and the progress recorded during and after the training necessitated incremental and collaborative strategies for feasible and sustainable development of bioinformatics and data science[1]. It is therefore obvious that application of data science is still at infancy stage generally in SSA except few computational efforts to implement data science in biomedicals according to the categories summarized in the Table 1 below.

Apart from the few efforts reported above in articles, a Pan African Bioinformatics network (H3ABioNet) for the Human Heredity and Health in Africa (H3Africa) has made tremendous efforts in the area of genomics by developing computational tools and workflows using data science approaches. The tools and workflows were developed to be applied to data generated from African cohorts and taking the genomic diversity and unique environmental properties into consideration. The tools are bioinformatics application software for performing specific genomic analysis, while the workflows are standard operating procedures (SOPs) for common bioinformatics data analysis tasks. Some SOPs contain a software workflow as well as corresponding containers with all the necessary software dependencies to run the workflow.

Table 1. **A Summary of data science approaches applied to biomedical sciences in Sub Saharan Africa.**

Application Area	Author	Technique/Algorithm	Data	Sample Size	Functionality/Feature	Implementation	Country
Public Health and Surveillance		Random Forests, Logistic Regression, and K-Nearest Neighbors	Demographic and Health Survey data	10,641 children	A predictive model for estimating under-five mortality risk factors, which could help to proffer better policy directions regarding under-five childhood survival.	Model	Ethiopia
	75	Single level binary logistic and multilevel logistic regression	Demographic and Health Survey data	5,884 children	A predictive model to identify determinants of under-five mortality, which could help inform health policy and intervention strategies aimed at improving child survival.	Model	Ghana
	76	Multinomial Logistic Regression, Decision Tree, and Naive Bayes Classification	Recruitment and retention data from Africa Health Placements	1,838 healthcare workers	An effective R tool to predict the recruited health workers' length of practice.	Tool	South Africa
	77	J48, Naïve Bayes, Support Vector Machine (SVM), and Artificial Neural Network (ANN)	Demographic and Health Survey data	11,023 women	A predictive model to forecast the likelihood of a pregnant woman seeking skilled delivery assistance to ensure skilled assistance at childbirth	Model	Ethiopia
	78	Decision tree (using J48 algorithm) and rule induction (using PART algorithm)	Demographic and Health Survey data	11,654 records	A web-based child mortality prediction model.	Model and Tool	Ethiopia
	79	Random Forests,	Clinical Data	949 children 1,579 adult	A model that predicts hospital mortality in children with acute infections	Model	Rural Rwanda.
Infectious Disease	55	XGBoost, K-Nearest Neighbors (KNN), Decision Tree, Random Forest, ExtraTree, AdaBoost, and Linear Discriminant Analysis (LDA)	Clinical and Environmental Data	2,951 Cholera patients	Use of machine learning techniques to model cholera epidemics with linkage to seasonal weather changes	Model	Tanzania
	80	Lasso and Empirical esian Kriging	Socio-Demographic and Health indicator survey data	Reported confirmed cases and deaths from coronavirus between February 28, 2020 through March 26, 2020 in 39 SSA Nations	A predictive model to reveal correlates and patterns of COVID-19 Disease outbreak in SSA.	Model	39 SSA Nations
	81	XGBoost algorithm	Population-based HIV Impact Assessment (PHIA) data	41,939 males and 45,105 females	A predictive model to identify people living with HIV and individuals with a higher likelihood of contracting the disease.	Model	Tanzania, Zambia, Malawi and Eswatini
	82	Penalized Logistic Regression (Elastic Net), Generalized Additive Model (GAM), Support Vector Machine (SVM) Gradient Boosting Tree (XGBoost)	Demographic Health Survey data	55,151 males and 69,626 females	Model to predict the HIV status of individuals based on socio-behavioural characteristics.	Model	10 countries in East and Southern Africa

Application Area	Author	Technique/Algorithm	Data	Sample Size	Functionality/Feature	Implementation	Country
NCDs	83	An adapted VGGNet architecture and a residual neural network architecture	Image and Clinical data	Training samples from Singapore (76,370 retinal fundus images, 13,099 participants). Validation Samples from Zambia (4504 retinal fundus images, 1574 participants)	An artificial intelligence model using deep learning for detecting referable diabetic retinopathy, vision-threatening diabetic retinopathy, and diabetic macular oedema in population-based diabetic retinopathy screening.	Model	Zambia
Genomics	84	Logistic Regression	Single Nucleotide Polymorphisms (SNPs) and Covariate data	> 2.45 M markers and 883 Children (< 3) with P. falciparum malaria	An association model to investigate the relationship between genotypes and/or haplotypes of SNPs and susceptibility to both Severe malarial anemia and inefficient erythropoiesis in anemic children	Model	Kenya

Within H3Africa, tools and workflows have been developed (H3ABioNet) and can be assessed here and here respectively. Since more models have been developed than tools, the authors opine that there is need to convert the models to user friendly tools and workflows for biomedical researchers. The tools and workflows should be developed to handle genomics diversity and environmental peculiarity of SSA. A website resource, which will house all the tools and the workflows should be developed for easy access to biomedical researchers in this region. A strong collaboration among disciplines such as statistics, computer science and healthcare/medicine should be established to propose feasible solutions to issues around data science and healthcare. Training, workshops and conferences should be organized so as to develop experts towards application of data science in healthcare.

Data science training needs in Africa

Despite the huge benefits and advantages of data science, Africa's capacity to fully utilize data science is limited. While the need for training and utilization of data science in Africa is glaring, lack of human capital and resources to maximize the advantages of data science poses a serious hindrance to its application in SSA[19]. Data science in SSA requires substantial investment in training, infrastructure and a growing portfolio of research grant funding to close the knowledge gap[85].

It is the authors' opinion that the critical elements to be included in the data science training should include 1) data generation 2) data storage and availability and 3) data utilization. Data generation requires links to cohorts and/or sequencing and laboratory cores to complement and enhance offerings while data storage is related to ensuring a staffed central data center. Data utilization will require the capability to offer customized analysis for complex data to answer diverse research questions across different disciplines requiring close collaborations with researchers to ensure optimal bioinformatics approaches for individual labs and research questions ("one size does not fit all") as well as engage biostatisticians in optimizing analytic approaches.

Review of institutions offering data science courses in Africa

Table 2 provides a list of institutions that currently (as of May 2023) have data science curricula in Africa – this information was obtained from a web search and exploration of websites from each of these institutions. These results further corroborate our claim that there are urgent needs in developing capacity of Africans for health data science. Besides the courses listed in the table, there are a number of non-degree awarding institutions offering a range of training on data science – some institutions offer Data Science trainings on Data Science but are yet to incorporate the same into their curriculum. Furthermore, the search revealed that the available training institutions were skewed to the Southern region of Africa (especially South Africa), and this shows a huge dearth of training opportunities in the other regions of Africa.

Gaps in data science training

Academic institutions in SSA have been playing catch-up with western countries in the field of data science. Over the last decade, practically all universities in Europe and North America have heeded the call to the opportunities of domiciled data science by establishing new institutes, departments and degree programs in the field[1,19]. From our review, we surmise that, in most SSA countries, there is a dearth of scientists (i) with data science skills and knowledge (from basic programming to appropriate statistical and ML techniques); (ii) have funding and resources to harness diverse sources and forms of health data for translational science; and (iii) understand ethics, data accessibility, integrity, quality and data security and (iv) with experience working in a multi-disciplinary team. To catch up with the trend and advances in health data science in other parts of the world, there are urgent needs to develop and implement multi-dimensional, context-specific and outcome-tailored training programs in health data science in SSA and bridge the gaps. However, our searches may be limited by the difficulty encountered in obtaining accurate and up to date information on institutional websites. Some of the websites were not up to date while others have insufficient information on the courses they run.

Based on literature search, the authors opine that applying data science methodologies in healthcare delivery and medical research should involve a collaboration or complete knowledge of the following:

i. Need for biostatisticians: Mathematical and statistical concepts are needed to develop methods for study designs and data analysis to advance scientific knowledge in biology, biomedical science and public health, with the ultimate goal of improving human health.

ii. Need for computer programming knowledge: Knowledge of programming languages (including R, Python, SQL, SAS, etc.) are very essential in the field of data science and importantly Unix-based operating system such as Linux is needed to handle biomedical data.

iii. Need for prior biological knowledge: The performance and effectiveness of ML and deep learning models are enhanced and made robust when biological prior knowledge of problem domain is incorporated into the methodology of solving the problem at hand[86].

Data science applications for genomic studies in SSA

Bacterial genomics research

Microbiome data though highly variable, often irregular and complex, has shown for emerging research to have a clear

Table 2. List of data science training institutions in Sub Saharan Africa.

School	Programme	Country	Course link
University of the Witwatersrand,	MSc Data *Science*	South Africa	https://www.wits.ac.za/msc-data-science/?gclid=CjwKCAjw44mlBhAQEiwAqP3eVvdnD38BfnLX6kda-zatpIRjdUqZKLS955WRxThD4Dua4zUULEv83BoC0F8QAvD_BwE
University of the Witwatersrand, Johannesburg	MSc Artificial Intelligence	South Africa	https://www.wits.ac.za/course-finder/postgraduate/science/msc-artificial-intelligence/
University of the Witwatersrand, Johannesburg	MSc e-science	South Africa	https://www.wits.ac.za/course-finder/postgraduate/science/e-science/?gclid=CjwKCAjw44mlBhAQEiwAqP3eVqTP26nnK6q_knUWGX68J8H41rU7pLCccabx4KOtUEhNGsyOG1ZPFhoCS4IQAvD_BwE
University of the Witwatersrand, Johannesburg	MA e-science	South Africa	https://www.wits.ac.za/course-finder/postgraduate/humanities/ma-e-science/
University of the Witwatersrand, Johannesburg	BSc Big Data Analytics	South Africa	https://www.wits.ac.za/course-finder/postgraduate/science/big-data-analytics/?gclid=CjwKCAjw44mlBhAQEiwAqP3eVqgae6a6zcTuhsBzkoypLPb5OtebYHKjuBg-bqHhny4ssVxNPY2pZxoCr6YQAvD_BwE
University of Pretoria	Data Science	South Africa	https://www.up.ac.za/data-science
University of Pretoria	MIT in Big Data Science	South Africa	https://www.up.ac.za/school-of-information-technology/article/2324622/mit-in-big-data-science-stream-c
University of Cape Town	Masters Programmes in Data Science	South Africa	https://www.gsb.uct.ac.za/data-science-leadership/?utc_source=google&utc_medium=cpc&autc_campaign=20276618877&utc_term=data%20science%20courses&gclid=CjwKCAjw44mlBhAQEiwAqP3eVpaRpL_ILYFkvZexaw7D2WWYgt7OllUt90aiHgfv6rAt_4nQ_YR4JBoCNLoQAvD_BwE
Rhodes University	MSc Bioinformatics	South Africa	https://rubi.ru.ac.za/
Namibia University of Science and Technology	Master of Science in Applied Statistics	Namibia	https://www.nust.na/programmes/master-science-applied-statistics
- The African Centre of Excellence in Data Science (ACE-DS) National University of Rwanda (NUR)	The master's degree with specialization in Data Mining	Rwanda	https://aceds.ur.ac.rw/specialization-data-mining
National Institute of Statistics		Rwanda	https://statistics.gov.rw/content/data-revolution/nisr-training-centre
University of Dar es Salaam	MSc. Data Science	Tanzania	https://cse.udsm.ac.tz/programs/program-details/msc-in-data-science-%28msc-dsc%29#:~:text=This%20is%20a%20two%2Dyear,evolving%20data%2Ddriven%20business%20world.
Makerere University	AI & Data Science	Uganda	https://cs.mak.ac.ug/research/groups_and_centers
Eden University	MSc Data Analytics	Zambia	https://www.edenuniversity.edu.zm/courses/become-a-php-master-and-make-money-fast/
University of Zimbabwe	BSc Data Science and Informatics	Zimbabwe	http://www.emhare.uz.ac.zw/
University of Zimbabwe	MSc Data Science and Informatics	Zimbabwe	http://www.emhare.uz.ac.zw/

School	Programme	Country	Course link
University of Zimbabwe	BSc Artificial Intelligence and Machine Learning	Zimbabwe	http://www.emhare.uz.ac.zw/
University of Zimbabwe	MSc Genomics and Biotechnology	Zimbabwe	http://www.emhare.uz.ac.zw/
University of Zimbabwe	BSc Data Science and Systems	Zimbabwe	http://www.emhare.uz.ac.zw/
University of Zimbabwe	MSc Bioinformatics and Genomics	Zimbabwe	http://www.emhare.uz.ac.zw/
Chinhoyi University of Zimbabwe	MSc Data Analytics	Zimbabwe	
University of Cape Coast	MSc. Data Management and Analysis	Ghana	https://cdamaa.ucc.edu.gh/training/msc_data_analysis.html
African Institute of Mathematical Sciences	Big Data Analytics	Ghana	https://nexteinstein.org/industry-initiative-2/bd4d-scp1-aims-cameroon/
Jomo Kenyatta Univeristy of Agriculture and Technology	MSc. Mathematics (Data Science)	Kenya	https://www.jkuat.ac.ke/jkuat-launches-new-course-in-data-science-and-analytics/
Addis Ababa University	MSc. Information Science and Systems (Health Information/ Information Science/Information System)	Ethiopia	http://www.aau.edu.et/cns/school-of-information-science/
Jamhuriya University of Science and Technology	MSc Data Science	Somalia	https://just.edu.so/course/master-of-science-in-data-science/
University of Stellenbosch	MSc Computer Science (Data Science)	South Africa	https://www.cs.sun.ac.za/teaching/masters/#:~:text=Our%20Masters%20degree%20is%20a,edge %20research%2C%20it%20often%20is.
University of KwaZulu-Natal	BSc Data Science (Combination of Computer Science & Statistics)	South Africa	https://smscs.ukzn.ac.za/data-science/
North West University	Data Mining/Data Science	South Africa	https://natural-sciences.nwu.ac.za/data-mining
North West University	BSc, MSc, PhD Pure and Applied Analytics	South Africa	https://natural-sciences.nwu.ac.za/paa
African University of Science & Technology	MSc Geoinformatics & GIS	Nigeria	https://aust.edu.ng/programs/school-of-science-and-science-education-1/geoinformatics-and-gis
Lead City University	Msc Computer Science (Computational Intelligence)	Nigeria	https://lcu.edu.ng/index.php/2-uncategorised/165-csc-courseware
University of Uyo	MPhil/PhD Computational Intelligence	Nigeria	https://www.uniuyo.edu.ng/index.html
University of Ibadan	PhD Computer Science (Data Mining)	Nigeria	http://sci.ui.edu.ng/compsciwelcome
American University of Nigeria	MSc Data Science and Analytics	Nigeria	https://www.aun.edu.ng/index.php/sitc/msc-data-science

impact on growth, immune health, metabolomic functioning, disease susceptibility, and drug response to name a few[87,88]. As most bacterial species in our microbiomes have been, until recently, largely unculturable, high throughput DNA sequencing has led to a deluge of data and the development of repositories and reference databases. These resources have served and continue to serve as a foundation for new technologies and data analysis tools, but are mostly utilized from the western/developed world where they are mostly generated[89]. There is a clear need for data science applications in microbiome research. For example, the need to accurately assemble complete genomes from small fragments of DNA sequences extracted from biologic samples represents a data science challenge. In general, microbiome researchers have to keep developing, applying or/and reapplying new tools and databases to answer new questions, making the field fast-paced and study findings not often comparable. Another growing challenge is that, bacterial pangenome needs to be overlaid with the mycobiome and virome, other omics (transcriptomics, metabolomics) as well as robust and comprehensive biologic and phenotypic datasets. Analyzing, interpreting and applying these multi-omic datasets requires significant data science resources, not just in SSA but globally.

Resources are needed to ethically[90] and robustly support microbiome for discovery, prediction, prevention and therapeutics in LMIC in SSA. Replicating large networks such as the NIH-funded Human Microbiome Project[91,92], and continued collaboration with robust consortiums and groups such as H3 Africa[1,8,14,15,93], will bring us closer to harnessing the unique diversity of populations living in SSA. For a long-lasting impact for diseases prevalent to SSA[94–96], discipline-specific and multidisciplinary training in computational biology, genomics and microbiology need to be a priority.

Viral genomics research

The advent of next-generation sequencing has ushered in a golden era of viral genomics studies. Viral genomics studies have rapidly transformed infectious disease public health surveillance making it possible to carry out near real-time and precise detection/investigation of disease outbreaks, which then provides insights into the emergence of diseases and transmission patterns across different populations[97–100]. Genomics and bioinformatics have greatly increased our understanding of infectious diseases caused by viruses such as HIV, Lassa virus, Ebola virus, and SARS-CoV-2[101,102]. This ranges from disease outbreak investigations and pathogenesis, host-pathogen interactions and variation, diagnostics marker identification and vaccine targets[17,101]. Also, high throughput viral genomics data can be combined with host genomics and patients' health records to give advice on treatment options giving rise to the field of precision medicine[14]. It can also provide important data on drug and vaccine interaction[103]. Molecular characterization of pathogens based on next generation sequencing data is now possible, and data science analytic software has also been developed and used to recognize changes in virus composition, study their diversity, track viral transmission and monitor their pathways. This can be used to influence or drive public health policies and implementation for viral disease control and prevention.

A major application of data generated from next generation sequencing (NGS) of pathogens is in development of molecular diagnostics such as Polymerase Chain Reaction (PCR) and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) based diagnostics. From data generated, primers and guide RNA can be designed to target specific pathogens such as Lassa and SARS-CoV-2[104,105], and multiplex detection of several pathogens in one reaction[106]. Protein structures and functions for understanding pathogen/therapeutic interaction can be inferred from genomic data. In-line with these, various gene expression studies are designed based on data generated.

In Africa, data science has been used to support viral genomic research; bioinformatic analysis of Ebola virus genomes from the 2014 outbreak in Nigeria was very instrumental in containing the outbreak and limiting it to 20 reported cases and 8 fatalities[99]. Phylogenetic analysis revealed that there was a single source for Nigerian infections and that the Nigerian Ebola virus lineage nested within a lineage previously seen in Liberia. It also showed that transmission from Nigeria to elsewhere did not occur and this helped to inform public health decisions and response[99]; molecular dating analysis revealed that Lassa virus originated in Nigerian over a thousand years ago and within the last several hundred years spread into neighboring West African countries[107]. During the current COVID-19 pandemic, data science and bioinformatics tools have helped in understanding spread of the virus (SARS-CoV-2) across different regions and countries, delineating the different lineages/variants of the virus and developing vaccines in record time to help stop the virus in its tracks[108,109].

Pharmacogenomics research

Africa with her rich diversity in genetics and ethnolinguistics stands to benefit hugely from data science. One major area is in knowledge and implementation of pharmacogenomics, which is a branch of pharmacology. The term "pharmacogenetics" gets often used interchangeably with pharmacogenomics, which is the science focused on evaluation of genetic influence on observed variations on drug absorption, distribution, metabolism, and excretion (ADME), thereby significantly affecting its pharmacokinetic (PK) and pharmacodynamic (PD) properties[93,110].

Pharmacogenomics (PGx) has been gaining tremendous grounds in Africa in the past decade and a half[93,110]. It's no longer news that Africa holds the genetic mysteries behind the origin of humankind and when compared to other distinct populations, Africa bears huge heterogeneity. Despite its promise fortransforming the field of pharmacogenomics, data from African populations is disproportionately severely limited. Laboratory analysis of biosamples lead to generation of very huge amount of data that without the skills-set of bioinformatics data managers, nothing substantial can be made out of such results. Also, generation of pharmacogenomics data is one thing, translation of such acquired data to clinical relevance is another. Both are two-pronged, having their prospects and challenges with respect to SSA.

Africa's healthcare is premised on both orthodox and traditional complimentary medicines, giving room for 'polypharmacy' and its attendant challenges such as: drug-drug interactions, drug-herb interactions, toxicity/adverse drug reactions and therapeutic failures. With data science via PGx, these challenges can be adequately managed for the benefit of the patients as well as upgraded image for the healthcare systems on the continent[111–114]. SSA and, the entire continent of Africa, stands to reap benefits from the huge market of conducting clinical trials. There is rising global concerns about pharmaceuticals been tested only in selected populations, optimised for their maximal benefit but then sold and/or distributedin African populations that were not included in clinical trials, meaning the drugs were not optimised for their use[115]. For example, dose adjustments based on research evidence have been suggested for efavirenz due to genetic variations in CYP2B6 gene, which results in HIV patients in SSA requiring a reduced dose[116,117]. With implementation of data science through the field of PGx in Africa, pharmaceutical companies are likely to carry out clinical trials in African populations[93]. This has the potential technology transfer for both basic and clinical scientists on the continent, as well as improve economic inflow for many African countries like it has done for Asian countries in recent times[93,117–119].

Gaps in PGx research in SSA

Sub-Saharan Africa in comparison to other regions has produced fewer publications in the field of PGx due largely to inadequate infrastructure and resources required for such high throughput studies[1]. As a result, most of the human genetic variations present in Africa is yet unexplored. There are African countries with little or no pharmacogenomics data. E.g., Djibouti, Eritrea[93,115]. This however does not apply to many West African/Central African and South African countries that have been included in genomic databases such as the 1000 genome project and the HapMap project. However, some level of advancements have occurred in recent times. For instance, studies undertaken to examine genetic markers for diabetes and those investigating the genetic selection in Africa as a result of disease exposure or environmental adaptation[120]. However, there is still need for more genomic studies along ethnolinguistic groups, since high heterogeneity have been seen on the continent[120].

If truly environment impacts on pharmacogenomics trend over the years, then there still exists a huge gap in publicly available pharmacogenomics information on Africans. What do we mean? Several existing pharmacogenomics data on African populations were those obtained from Africans in the diaspora: populations whose diet, lifestyle and environment significantly differs from what obtains on the African continent itself. Studies investigating pharmacogenetics in Africans and Africans in diaspora have revealed increased diversity in Africans resident on the continent but declining diversity as Africans migrated from the continent[93,115]. The studies have shown that adaptation to diet, disease, and climate have generated into inherited traits that led to production of distinct populations with characteristic physical and physiological differences, including differences in responseto pathogens and xenobiotic challenges[121,122].

Pharmacogenomics is necessary to curb increasing trend in anti-retroviral induced adverse drug reactions coupled with surge in therapeutic failure of medicines used in management of chronic non-communicable diseases such as seen with breast cancer and hypertension treatments[123–125]. African healthcare systems are complex, involving contemporary and herbal medicines. Thus, pharmacogenomics could enable a better understanding of the basis of both western and traditional medicine leading to better integration[93,111,113]

As emphasised above, Africa can no longer be said to be naïve to the knowledge of PGx or acquaintance of expertise in the field. However, huge disparities still lie between Sub-Saharan Africa's uptake of the technology and developed societies. Translational PGx is still very much in the embryonic stage. The PharmGKB site documents over 40 gene-drug pairs that have influenced the literature inserts of some medications. Instances include CYP2D6-Tamoxifen, CYP2C19-Clopidogrel, and VKORC1/CYP2C9–warfarin, but how many practising physicians on the continent implement genotyping of patients to determine the dose appropriate for them prior to commencement of therapy? High cost of genetic testing, cultural biases, paucity of pharmacogenetic diagnostic tools pertinent to African genes are some of the limiting factors contributing to this[121].

The near absence of electronic medical records (EMR) in most healthcare systems in many LMIC in SSA is another clog-in-the-wheel of translation of PGx in SSA and by extension, Africa in whole. Nigeria, the third largest African population cannot boast of EMR in most of her government owned health institutions. A few private healthcare centres own and operate EMR but such facilities are only at the reach of high-profile citizens[126,127]. Whereas there is sufficiently high disposition and readiness by healthcare givers to switch over to EMRs, lack of infrastructure and funding, among other factors, militate against its uptake in the country[128,129]. Without this structural backbone, uptake of PGx in the clinical settings in SSA will still linger for a while. Table 3 highlights the significant milestones with respect to PGx[93,130].

Role of data science in drug discovery and development

The advent of big data science has revolutionised the process of drug discovery, development and production. In the past, the average time taken from drug discovery to product launch was around 15-20 years. However, with the introduction of in silico analysis, this timeline has been significantly reduced to approximately 7-10 years.[59,131]. Vast information acquired from target molecules that did not make it to

Table 3. A list of some of the common genomics initiatives in Africa[93,130].

INITIATIVE	FOCUS	ADDRESS/CONTACT
African Pharmacogenomics Consortium (APC)	The genetics of drug effectiveness (meetings, training workshops, conferences, collaborations)	Current initiative (website to be developed) (bsiddondo@strathmore.edu)
The African Society for Human Genetics (AfSHG)	Annual conferences/meetings	https://www.afshg.org/
African Human Genome Initiative	Lectures, conferences, discussions	www.africagenome.co.za
H3Africa	Genomics and environmental determinants of disease	https://h3africa.org
MalariaGEN	Malaria genomic epidemiology Network, focussing on effects of genetic variation on the biology and epidemiology of malaria	www.malariagen.net
H3ABioNet	pan-African bioinformatics network	https://www.h3abionet.org/
The Southern African Human Genome Project	Understanding of DNA variation among southern Africans and how this impact on the health of the people of our country.	https://sahgp.sanbi.ac.za
African Genome Variation Project	Aims to collect essential information about the structure of African genomes to provide a basic framework for genetic disease studies in Africa	https://www.sanger.ac.uk/science/collaboration/african-genome-variation-project

product launch have been collated to form chemical libraries and are now available for re-modification, and/or repurposing for previously non-indicated disease treatment/management and thus have helped to shorten drug discovery process[132]. AI and machine learning tools are now being used in drug design (via in silico methods), chemical synthesis, drug repurposing, drug screening, and polypharmacology[133]. Prediction of physicochemical properties, bioactivity, toxicity, target protein structure and drug interaction are now possible by AI tools[101,134]. Other applications in pharma include pharmaceutical product development, pharmaceutical manufacturing, clinical trial design, quality control and quality assurance, pharmaceutical product management (supply chain) and market delivery amongst others[133,135–137].

If biomedical scientists in Africa invest more into data science in terms of increased manpower, it stands to benefit immensely from the application of data science in drug discovery process since the demands for infrastructure isn't as large in comparison to setting up laboratories for wet analysis. Major tools for *in silico* analysis are reliable internet connections, servers, high capacity laptops and power. It is more of a lower hanging fruit for Africa to get involved in drug discovery process. For instance, during the COVID-19 pandemic, in collaboration with Northern partners, *in silico* analysis was utilised in predicting existing drug molecules that could be re-purposed as anti-covid agent in management of COVID-19 patients[101]. To arrive at the likely anti-covid candidates, over one million compounds comprised of approved drugs, investigational drugs, natural products, and organic compounds, were screened, a feat that never could have been accomplished under the old traditional processes of drug discovery. Africa has a rich untapped biodiversity - untapped in the sense that little of the studied plant system have translated into medicines. With capacity building in data science, these can be explored, and with the aforementioned benefit of shorter turn-around time, be utilized for solving challenging health issues such as is been faced with the COVID-19 pandemic[134].

Conclusions and recommendations

In this narrative review, the shortcomings of the health system in combating the infectious and non-communicable diseases that abound in sub-Saharan African countries and the potential role of data science in addressing the gaps were elucidated. Aside from lack of political will, infrastructure and manpower resources, a major factor found to be key in addressing the current inadequate state of health care system and services is poor harnessing of available data for use in decision-making and policy for planning interventions to improve the quality of healthcare delivery and public health. Data science has a key role in public health surveillance especially for gathering large health information and data on health trend as well as enhancing predictive analysis to inform decision-making on early warning of unprecedented disease outbreak and public health action. Through advances in bioinformatics and information technology, big data research is now very relevant in biomedical research, especially with bacterial and viral genomics, and pharmacogenomics, as areas of great potential for big data revolution.

This review revealed that a large amount of data can be generated with the advances in technology and implementation of large studies in SSA. In addition, information relating to medicine for effective healthcare delivery is now coming in torrents requiring knowledge from data science for effective medical diagnosis, treatment and research. Furthermore, studies on genomics and molecular epidemics of diseases, which utilise big data is upcoming in SSA. However, many of the healthcare informatics solutions through data science are mostly developed in high-income countries and there is dearth in the skill for data science in SSA.

In the area of drug discovery and development, SSA can become a hub for drug repurposing and discovery of new molecules which hitherto was non-existent. Data science can assist in converting the rich biodiversity of SSA for drug discovery in the face of drug resitance and emerging diseases.

To bridge the gaps and catch up with the trend and advances in health data science in other parts of the world, there are urgent needs to develop and implement multi-dimensional, context-specific and outcome-tailored training programs in health data science in SSA. These will include use of artificial intelligence and machine learning to build and apply models alongside computational tools and models to simulate systems or expedite biomedical research like bioinformatics and computational biology. In addition, opportunities available and the advances that the region has made so far should be built upon. Ongoing support from and partnership with several funding agencies and institutions from western countries such as Europe, Canada and United States are critical in supporting and building this expertise and should be encouraged.

The dearth in institutions offering training in data science in Africa should be addressed and buttressed by financial as well as technical support from the developed countries with the expertise using sustainable approach. Degree awarding programs would help to promote the value of data science education. Other learning platforms for data science in universities and colleges would encourage many departments to tap into the possibilities of data sciences, and in the long run students in a wide range of fields would learn the skills necessary to be proficient in data science. Hubs created for data science outside the university system would also increase frontiers in data science. These hubs could serve as a mobile access for students who cannot afford university education.

Africa, with the genetic diversity, potential to generate big data, and the clamour to harness data for improved as well as innovative medicine is a major stakeholder in development of data science, but there is need for funding and adequate training that will build capacity and subsequently benefit by using data science for economic growth and improved health of the populace.

Data availability

No data are associated with this article.

References

1. Shaffer JG, Mather FJ, Wele M, *et al.*: **Expanding Research Capacity in Sub-Saharan Africa Through Informatics, Bioinformatics, and Data Science Training Programs in Mali.** *Front Genet.* 2019; **10**: 331.
PubMed Abstract | Publisher Full Text | Free Full Text
2. Mandreoli F, Ferrari D, Guidetti V, *et al.*: **Real-world data mining meets clinical practice: Research challenges and perspective.** *Front Big Data.* 2022; **5**: 1021621.
PubMed Abstract | Publisher Full Text | Free Full Text
3. Kiryu Y: **[Medical Big Data Analysis Using Machine Learning Algorithms in the Field of Clinical Pharmacy].** *Yakugaku Zasshi.* 2022; **142**(4): 319–326.
PubMed Abstract | Publisher Full Text
4. Sigwadi L: **Data Science and the Fourth Industrial Revolution (4IR).** 2020.
Reference Source
5. Earnshaw R: **Data Science.** In: *Data Science and Visual Computing.* Cham: Springer International Publishing; 2019; 1–10.
6. ALU Team: **Data Science Africa's Returrn Ticket to the Future.** 2020.
Reference Source
7. Zhang Q: **Data science approaches to infectious disease surveillance.** *Philos Trans A Math Phys Eng Sci.* 2022; **380**(2214): 20210115.
PubMed Abstract | Publisher Full Text | Free Full Text
8. Aron S, Gurwitz K, Panji S, *et al.*: **H3ABioNet: Developing Sustainable Bioinformatics Capacity in Africa.** 2017; **23**.
Publisher Full Text
9. United Nations Department of Economic and Social Affairs PD: **World Population Prospects 2022: Summary of Results.** 2022.
Reference Source
10. Osseni IA: **COVID-19 pandemic in sub-Saharan Africa: preparedness, response, and hidden potentials.** *Trop Med Health.* 2020; **48**: 48.
PubMed Abstract | Publisher Full Text | Free Full Text
11. Low W, Tan S, Schwartz S: **The effect of severe caries on the quality of life in young children.** *Pediatr Dent.* 1999; **21**(6): 325–326.
PubMed Abstract
12. Medeiros MM, Hoppen N, Macada ACG: **Data science for business: benefits, challenges and opportunities.** *Bottom Line.* 2020; **33**(2): 149–163.
Publisher Full Text
13. World Health Organization: **Big data and artificial intelligence for achieving universal health coverage: an international consultation on ethics.** Geneva, 2018.
14. Adoga MP, Fatumo SA, Agwale SM: **H3Africa: a tipping point for a revolution in bioinformatics, genomics and health research in Africa.** *Source Code Biol Med.* 2014; **9**: 10.
PubMed Abstract | Publisher Full Text | Free Full Text
15. Mulder N, Abimiku A, Adebamowo SN, *et al.*: **H3Africa: current perspectives.** *Pharmgenomics Pers Med.* 2018; **11**: 59–66.
PubMed Abstract | Publisher Full Text | Free Full Text
16. Tastan Bishop O, Adebiyi EF, Alzohairy AM, *et al.*: **Bioinformatics education--perspectives and challenges out of Africa.** *Brief Bioinform.* 2015; **16**(2): 355–364.
PubMed Abstract | Publisher Full Text | Free Full Text
17. Mulder NJ, Adebiyi E, Alami R, *et al.*: **H3ABioNet, a sustainable pan-African bioinformatics network for human heredity and health in Africa.** *Genome Res.* 2016; **26**(2): 271–277.
PubMed Abstract | Publisher Full Text | Free Full Text
18. Doumbo OK, Krogstad DJ: **Doctoral training of African scientists.** *Am J Trop Med Hyg.* 1998; **58**(2): 127–132.
PubMed Abstract | Publisher Full Text
19. Beyene J, Harrar SW, Altaye M, *et al.*: **A Roadmap for Building Data Science Capacity for Health Discovery and Innovation in Africa.** *Front Public Health.* 2021; **9**: 710961.
PubMed Abstract | Publisher Full Text | Free Full Text

20. Tishkoff SA, Reed FA, Friedlaender FR, *et al.*: **The genetic structure and history of Africans and African Americans.** *Science.* 2009; **324**(5930): 1035–1044.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

21. Dye C: **After 2015: infectious diseases in a new era of health and development.** *Philos Trans R Soc Lond B Biol Sci.* 2014; **369**(1645): 20130426.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

22. World Health Organization: **World health statistics 2019: monitoring health for the SDGs, sustainable development goals.** Geneva: World Health Organization; 2019.
Reference Source

23. Shears P: **Emerging and reemerging infections in africa: the need for improved laboratory services and disease surveillance.** *Microbes Infect.* 2000; **2**(5): 489–495.
PubMed Abstract | **Publisher Full Text**

24. Boutayeb A: **The Impact of Infectious Diseases on the Development of Africa.** *Handbook of Disease Burdens and Quality of Life Measures.* 2010; 1171–1188.
Publisher Full Text

25. Shetty NP SP: **Epidemiology of Disease in the Tropics.** *Manson's Tropical Diseases.* 2009.

26. **Centers for Disease Control.** *CDC in Nigeria.* 2022.

27. Muhammad F, Abdulkareem JH, Chowdhury AA: **Major Public Health Problems in Nigeria: A review.** *South East Asia J Public Health.* 2017; **7**(1): 6–11.
Publisher Full Text

28. Seguin M, Niño Zarazúa M: **Non-clinical interventions for acute respiratory infections and diarrhoeal diseases among young children in developing countries.** *Trop Med Int Health.* 2015; **20**(2): 146–169.
PubMed Abstract | **Publisher Full Text**

29. World Health Organization: **Accelerating progress on HIV, tuberculosis, malaria, hepatitis and neglected tropical diseases: A new agenda for 2016-2030.** 2015.
Reference Source

30. Sofola OO, Folayan MO, Oginni AB: **Changes in the prevalence of dental caries in primary school children in Lagos State, Nigeria.** *Niger J Clin Pract.* 2014; **17**(2): 127–133.
PubMed Abstract | **Publisher Full Text**

31. Center for Medicaid and Medicare Services: **National Health Expenditures 2017 Highlights.** [accessed 2019].
Reference Source

32. World Health Organization: **The Global health Observatory - Indicators.** WHO; 2023.
Reference Source

33. MacNeil A, Glaziou P, Sismanidis C, *et al.*: **Global Epidemiology of Tuberculosis and Progress Toward Achieving Global Targets - 2017.** *MMWR Morb Mortal Wkly Rep.* 2019; **68**(11): 263–266.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

34. Centers for Disease Control and Prevention: **Malaria's Impact Worldwide**. 2021.
Reference Source

35. World Health Organization: **Business case for WHO immunization activities on the African continent, 2018–2030.** Brazzaville. Licence: CC BY-NC-SA 3.0 IGO. 2018.
Reference Source

36. World Health Organisation: **Global Action Plan for the Prevention and Control of Noncommunicable Diseases 2013–2020**. 2013.
Reference Source

37. Dalal S, Beunza JJ, Volmink J, *et al.*: **Non-communicable diseases in sub-Saharan Africa: what we know now.** *Int J Epidemiol.* 2011; **40**(4): 885–901.
PubMed Abstract | **Publisher Full Text**

38. WHO African Region: **Report on the status of major health risk factors for noncommunicable diseases**. 2015.
Reference Source

39. Zeng Z, Shaffer JR, Wang X, *et al.*: **Genome-wide association studies of pit-and-fissure- and smooth-surface caries in permanent dentition.** *J Dent Res.* 2013; **92**(5): 432–437.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

40. Mudie K, Jin MM, Tan, *et al.*: **Non-communicable diseases in sub-Saharan Africa: a scoping review of large cohort studies.** *J Glob Health.* 2019; **9**(2): 020409.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

41. Witter S, Zou G, Diaconu K, *et al.*: **Opportunities and challenges for delivering non-communicable disease management and services in fragile and post-conflict settings: perceptions of policy-makers and health providers in Sierra Leone.** *Confl Health.* 2020; **14**: 3.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

42. Juma PA, Mohamed SF, Matanje Mwagomba BL, *et al.*: **Non-communicable disease prevention policy process in five African countries.** *BMC Public Health.* 2018; **18**(Suppl 1): 961.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

43. Bhardwaj N, Wodajo B, Spano A, *et al.*: **The Impact of Big Data on Chronic Disease Management.** *Health Care Manag (Frederick).* 2018; **37**(1): 90–98.
PubMed Abstract | **Publisher Full Text**

44. Rumsfeld JS, Joynt KE, Maddox TM: **Big data analytics to improve cardiovascular care: promise and challenges.** *Nat Rev Cardiol.* 2016; **13**(6): 350–359.
PubMed Abstract | **Publisher Full Text**

45. Federal Ministry of Health N: **National technical guidelines for integrated diseases surveillance and response.** Abuja, 2009.

46. Fall IS, Rajatonirina S, Yahaya AA, *et al.*: **Integrated Disease Surveillance and Response (IDSR) strategy: current status, challenges and perspectives for the future in Africa.** *BMJ Glob Health.* 2019; **4**(4): e001427.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

47. World Health Organisation: **Integrated Disease Surveillance and Response Technical Guidelines.** Brazzaville: WHO Regional Office for Africa. Licence: CC BY-NC-SA 3.0 IGO. 2019.
Reference Source

48. World Health Organization: **An Integrated approach to communicable disease surveillance**. 2020.

49. Veenema TG, Toke J: **Early detection and surveillance for biopreparedness and emerging infectious diseases.** *Online J Issues Nurs.* 2006; **11**(1): 3.
PubMed Abstract

50. Kasolo F, Roungou J., Perry H: **Technical guidelines for Integrated Disease Surveillance and response in the African region.** 2nd edition. 2010.
Reference Source

51. Isere EE, Fatiregun AA, Ajayi IO: **An overview of disease surveillance and notification system in Nigeria and the roles of clinicians in disease outbreak prevention and control.** *Niger Med J.* 2015; **56**(3): 161–168.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

52. Dash S, Shakyawar SK, Sharma M, *et al.*: **Big data in healthcare: management, analysis and future prospects.** *J Big Data.* 2019; **6**(1): 54.
Publisher Full Text

53. Linder JE, Bastarache L, Hughey JJ, *et al.*: **The Role of Electronic Health Records in Advancing Genomic Medicine.** *Annu Rev Genomics Hum Genet.* 2021; **22**: 219–238.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

54. Huang SC, Pareek A, Seyyedi S, *et al.*: **Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines.** *NPJ Digit Med.* 2020; **3**: 136.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

55. Bretz WA, Corby PM, Schork NJ, *et al.*: **Longitudinal analysis of heritability for dental caries traits.** *J Dent Res.* 2005; **84**(11): 1047–1051.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

56. Subrahmanya SVG, Shetty DK, Patil V, *et al.*: **The role of data science in healthcare advancements: applications, benefits, and future prospects.** *Ir J Med Sci.* 2022; **191**(4): 1473–1483.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

57. Batko K, Slezak A: **The use of Big Data Analytics in healthcare.** *J Big Data.* 2022; **9**(1): 3.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

58. Carter AB, Abruzzo LV, Hirschhorn JW, *et al.*: **Electronic Health Records and Genomics: Perspectives from the Association for Molecular Pathology Electronic Health Record (EHR) Interoperability for Clinical Genomics Data Working Group.** *J Mol Diagn.* 2022; **24**(1): 1–17.
PubMed Abstract | **Publisher Full Text**

59. Shaker B, Ahmad S, Lee J, *et al.*: ***In silico* methods and tools for drug discovery.** *Comput Biol Med.* 2021; **137**: 104851.
PubMed Abstract | **Publisher Full Text**

60. Purcell S, Neale B, Todd-Brown K, *et al.*: **PLINK: a tool set for whole-genome association and population-based linkage analyses.** *Am J Hum Genet.* 2007; **81**(3): 559–575.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

61. Zhan X, Hu Y, Li B, *et al.*: **RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data.** *Bioinformatics.* 2016; **32**(9): 1423–1426.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

62. Deo RC: **Machine Learning in Medicine.** *Circulation.* 2015; **132**(20): 1920–1930.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

63. Akinnagbe A, Dharini Amitha Peiris K, Akinloye O: **Prospects of Big Data Analytics in Africa Healthcare System.** *Glob J Health Sci.* 2018; **10**(6): 1.
Publisher Full Text

64. Chifamba N: **Readiness for E-Health in Sub-Saharan Africa.** 2019.
Reference Source

65. Adeloye D, Adigun T, Misra S, *et al.*: **Assessing the Coverage of E-Health Services in Sub-Saharan Africa. A Systematic Review and Analysis.** *Methods Inf Med.* 2017; **56**(3): 189–199.
PubMed Abstract | **Publisher Full Text**

66. Camacho DM, Collins KM, Powers RK, *et al.*: **Next-Generation Machine Learning for Biological Networks.** *Cell.* 2018; **173**(7): 1581–1592.
PubMed Abstract | **Publisher Full Text**

67. Núñez Reiz A, Armengol de la Hoz MA, Sánchez García M: **Big Data Analysis and Machine Learning in Intensive Care Units.** *Med Intensiva (Engl Ed).* 2019; **43**(7): 416–426.
PubMed Abstract | **Publisher Full Text**

68. Bohr A, Memarzadeh K: **The rise of artificial intelligence in healthcare applications.** *Artificial Intelligence in Healthcare.* 2020; 25–60.
Publisher Full Text | **Free Full Text**

69. Miller DD, Brown EW: **Artificial Intelligence in Medical Practice: The Question to the Answer?** *Am J Med.* 2018; **131**(2): 129–133.
PubMed Abstract | **Publisher Full Text**

70. Mintz Y, Brodie R: **Introduction to artificial intelligence in medicine.** *Minim Invasive Ther Allied Technol.* 2019; **28**(2): 73–81.
PubMed Abstract | **Publisher Full Text**

71. Ngiam KY, Khor IW: **Big data and machine learning algorithms for health-care delivery.** *Lancet Oncol.* 2019; **20**(5): e262–e273.
PubMed Abstract | **Publisher Full Text**

72. Agler CS, Shungin D, Ferreira Zandona AG, *et al.*: **Protocols, Methods, and Tools for Genome-Wide Association Studies (GWAS) of Dental Traits.** *Methods Mol Biol.* 2019; **1922**: 493–509.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

73. Elisa N: **Could Machine Learning be used to address Africa's Challenges?** *Int J Comput Appl.* 2018; **180**(18): 9–12.
Publisher Full Text

74. Owoyemi A, Owoyemi J, Osiyemi A, *et al.*: **Artificial Intelligence for Healthcare in Africa.** *Front Digit Health.* 2020; **2**: 6.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

75. Ekhaguere OA, Oluwafemi RO, Badejoko B, *et al.*: **Automated phone call and text reminders for childhood immunisations (PRIMM): a randomised controlled trial in Nigeria.** *BMJ Glob Health.* 2019; **4**(2): e001232.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

76. Nariyama M, Shimizu K, Uematsu T, *et al.*: **Identification of chromosomes associated with dental caries susceptibility using quantitative trait locus analysis in mice.** *Caries Res.* 2004; **38**(2): 79–84.
PubMed Abstract | **Publisher Full Text**

77. Boraas JC, Messer LB, Till MJ: **A genetic contribution to dental caries, occlusion, and morphology as demonstrated by twins reared apart.** *J Dent Res.* 1988; **67**(9): 1150–1155.
PubMed Abstract | **Publisher Full Text**

78. Bretz WA, Corby PM, Hart TC, *et al.*: **Dental caries and microbial acid production in twins.** *Caries Res.* 2005; **39**(3): 168–172.
PubMed Abstract | **Publisher Full Text**

79. Bretz WA, Corby PM, Melo MR, *et al.*: **Heritability estimates for dental caries and sucrose sweetness preference.** *Arch Oral Biol.* 2006; **51**(12): 1156–1160.
PubMed Abstract | **Publisher Full Text**

80. Conry JP, Messer LB, Boraas JC, *et al.*: **Dental caries and treatment characteristics in human twins reared apart.** *Arch Oral Biol.* 1993; **38**(11): 937–943.
PubMed Abstract | **Publisher Full Text**

81. Shaffer JR, Feingold E, Wang X, *et al.*: **Heritable patterns of tooth decay in the permanent dentition: principal components and factor analyses.** *BMC Oral Health.* 2012; **12**: 7.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

82. Shaffer JR, Wang X, Desensi RS, *et al.*: **Genetic susceptibility to dental caries on pit and fissure and smooth surfaces.** *Caries Res.* 2012; **46**(1): 38–46.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

83. Wang X, Shaffer JR, Weyant RJ, *et al.*: **Genes and their effects on dental caries may differ between primary and permanent dentitions.** *Caries Res.* 2010; **44**(3): 277–284.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

84. Shaffer JR, Feingold E, Wang X, *et al.*: **GWAS of dental caries patterns in the permanent dentition.** *J Dent Res.* 2013; **92**(1): 38–44.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

85. Munthali T, Musonda P, Mee P, *et al.*: **Underutilisation of routinely collected data in the HIV programme in Zambia: a review of quantitatively analysed peer-reviewed articles.** *Health Res Policy Syst.* 2017; **15**(1): 51.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

86. Morrison J, Laurie CC, Marazita ML, *et al.*: **Genome-wide association study of dental caries in the Hispanic Communities Health Study/Study of Latinos (HCHS/SOL).** *Hum Mol Genet.* 2016; **25**(4): 807–816.
PubMed Abstract | **Free Full Text**

87. Shaffer JR, Wang X, Feingold E, *et al.*: **Genome-wide association scan for childhood caries implicates novel genes.** *J Dent Res.* 2011; **90**(12): 1457–1462.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

88. Li J, Quinque D, Horz HP, *et al.*: **Comparative analysis of the human saliva microbiome from different climate zones: Alaska, Germany, and Africa.** *BMC Microbiol.* 2014; **14**: 316.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

89. Pheeha SM, Tamuzi JL, Chale-Matsau B, *et al.*: **A Scoping Review Evaluating the Current State of Gut Microbiota Research in Africa.** *Microorganisms.* 2023; **11**(8): 2118.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

90. Littlejohn PT, Glover JS: **Ethical gut microbiota research in Africa.** *Nat Microbiol.* 2023; **8**(8): 1376–1377.
PubMed Abstract | **Publisher Full Text**

91. Wagner J, Kancherla J, Braccia D, *et al.*: **Interactive exploratory data analysis of Integrative Human Microbiome Project data using Metaviz [version 1; peer review: 3 approved].** *F1000Res.* 2020; **9**: 601.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

92. Integrative HMP (iHMP) Research Network Consortium: **The Integrative Human Microbiome Project: dynamic analysis of microbiome-host omics profiles during periods of human health and disease.** *Cell Host Microbe.* 2014; **16**(3): 276–289.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

93. Dandara C, Masimirembwa C, Haffani YZ, *et al.*: **African Pharmacogenomics Consortium: Consolidating pharmacogenomics knowledge, capacity development and translation in Africa [version 1; peer review: 2 approved].** *AAS Open Res.* 2019; **2**: 19.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

94. Llinas-Caballero K, Caraballo L: **Helminths and Bacterial Microbiota: The Interactions of Two of Humans' "Old Friends".** *Int J Mol Sci.* 2022; **23**(21): 13358.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

95. Maigoro AY, Muhammad M, Bello B, *et al.*: **Exploration of Gut Microbiome Research in Africa: A Scoping Review.** *J Med Food.* 2023; **26**(9): 616–623.
PubMed Abstract | **Publisher Full Text**

96. Allali I, Abotsi RE, Tow LA, *et al.*: **Human microbiota research in Africa: a systematic review reveals gaps and priorities for future research.** *Microbiome.* 2021; **9**(1): 241.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

97. Grad YH, Lipsitch M: **Epidemiologic data and pathogen genome sequences: a powerful synergy for public health.** *Genome Biol.* 2014; **15**(11): 538.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

98. Siddle KJ, Eromon P, Barnes KG, *et al.*: **Genomic Analysis of Lassa Virus during an Increase in Cases in Nigeria in 2018.** *N Engl J Med.* 2018; **379**(18): 1745–1753.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

99. Folarin OA, Ehichioya D, Schaffner SF, *et al.*: **Ebola Virus Epidemiology and Evolution in Nigeria.** *J Infect Dis.* 2016; **214**(suppl 3): S102–S109.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

100. Ajogbasile FV, Oguzie JU, Oluniyi PE, *et al.*: **Real-time Metagenomic Analysis of Undiagnosed Fever Cases Unveils a Yellow Fever Outbreak in Edo State, Nigeria.** *Sci Rep.* 2020; **10**(1): 3180.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

101. Olubiyi OO, Olagunju M, Keutmann M, *et al.*: **High Throughput Virtual Screening to Discover Inhibitors of the Main Protease of the Coronavirus SARS-CoV-2.** *Molecules.* 2020; **25**(14): 3193.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

102. Coker MO, Mongodin EF, El-Kamary SS, *et al.*: **Immune status, and not HIV infection or exposure, drives the development of the oral microbiota.** *Sci Rep.* 2020; **10**(1): 10830.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

103. Bah SY, Morang'a CM, Kengne-Ouafo JA, *et al.*: **Highlights on the Application of Genomics and Bioinformatics in the Fight Against Infectious Diseases: Challenges and Opportunities in Africa.** *Front Genet.* 2018; **9**: 575.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

104. Arizti-Sanz J, Freije CA, Stanton AC, *et al.*: **Streamlined inactivation, amplification, and Cas13-based detection of SARS-CoV-2.** *Nat Commun.* 2020; **11**(1): 5921.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

105. Barnes KG, Lachenauer AE, Nitido A, *et al.*: **Deployable CRISPR-Cas13a diagnostic tools to detect and report Ebola and Lassa virus cases in real-time.** *Nat Commun.* 2020; **11**(1): 4131.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

106. Yang X, Yang S, Qi H, *et al.*: **PlaPPISite: a comprehensive resource for plant protein-protein interaction sites.** *BMC Plant Biol.* 2020; **20**(1): 61.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

107. Andersen KG, Shapiro BJ, Matranga CB, *et al.*: **Clinical Sequencing Uncovers Origins and Evolution of Lassa Virus.** *Cell.* 2015; **162**(4): 738–750.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

108. Giandhari J, Pillay S, Wilkinson E, *et al.*: **Early transmission of SARS-CoV-2 in South Africa: An epidemiological and phylogenetic report.** *medRxiv.* 2020; 2020.05.29.20116376.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

109. Ngoi JM, Quashie PK, Morang'a CM, *et al.*: **Genomic analysis of SARS-CoV-2 reveals local viral evolution in Ghana.** *Exp Biol Med (Maywood).* 2021; **246**(8): 960–970.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

110. Kitzmiller JP, Groen DK, Phelps MA, *et al.*: **Pharmacogenomic testing: relevance in medical practice: why drugs work in some patients but not in others.** *Cleve Clin J Med.* 2011; **78**(4): 243–257.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

111. Gouws C, Steyn D, Du Plessis L, *et al.*: **Combination therapy of Western drugs and herbal medicines: recent advances in understanding interactions involving metabolism and efflux.** *Expert Opin Drug Metab Toxicol.* 2012; **8**(8): 973–984.
PubMed Abstract | **Publisher Full Text**

112. Ghodke Y, Joshi K, Patwardhan B: **Traditional Medicine to Modern**

Pharmacogenomics: Ayurveda *Prakriti* Type and CYP2C19 Gene Polymorphism Associated with the Metabolic Variability. *Evid Based Complement Alternat Med.* 2011; **2011**: 249528.
PubMed Abstract | Publisher Full Text | Free Full Text

113. Liu MZ, Zhang YL, Zeng MZ, *et al.*: **Pharmacogenomics and herb-drug interactions: merge of future and tradition.** *Evid Based Complement Alternat Med.* 2015; **2015**: 321091.
PubMed Abstract | Publisher Full Text | Free Full Text

114. Meng Q, Liu K: **Pharmacokinetic interactions between herbal medicines and prescribed drugs: focus on drug metabolic enzymes and transporters.** *Curr Drug Metab.* 2014; **15**(8): 791–807.
PubMed Abstract | Publisher Full Text

115. Rajman I, Knapp L, Morgan T, *et al.*: **African Genetic Diversity: Implications for Cytochrome P450-mediated Drug Metabolism and Drug Development.** *EBioMedicine.* 2017; **17**: 67–74.
PubMed Abstract | Publisher Full Text | Free Full Text

116. Matimba A, Del-Favero J, Van Broeckhoven C, *et al.*: **Novel variants of major drug-metabolising enzyme genes in diverse African populations and their predicted functional effects.** *Hum Genomics.* 2009; **3**(2): 169–190.
PubMed Abstract | Publisher Full Text | Free Full Text

117. Matimba A, Dhoro M, Dandara C: **Is there a role of pharmacogenomics in Africa.** *Glob Health Epidemiol Genom.* 2016; **1**: e9.
PubMed Abstract | Publisher Full Text | Free Full Text

118. Leng J, Ntekim AI, Ibraheem A, *et al.*: **Infrastructural Challenges Lead to Delay of Curative Radiotherapy in Nigeria.** *JCO Glob Oncol.* 2020; **6**: 269–276.
PubMed Abstract | Publisher Full Text | Free Full Text

119. Ntekim A, Ibraheem A, Adeniyi-Sofoluwe A, *et al.*: **Implementing oncology clinical trials in Nigeria: a model for capacity building.** *BMC Health Serv Res.* 2020; **20**(1): 713.
PubMed Abstract | Publisher Full Text | Free Full Text

120. Ali AA, Aalto M, Jonasson J, *et al.*: **Genome-wide analyses disclose the distinctive HLA architecture and the pharmacogenetic landscape of the Somali population.** *Sci Rep.* 2020; **10**(1): 5652.
PubMed Abstract | Publisher Full Text | Free Full Text

121. Radouani F, Zass L, Hamdi Y, *et al.*: **A review of clinical pharmacogenetics Studies in African populations.** *Per Med.* 2020; **17**(2): 155–170.
PubMed Abstract | Publisher Full Text | Free Full Text

122. Masimirembwa C, Dandara C, Hasler J: **Population Diversity and Pharmacogenomics in Africa.** *Handbook of Pharmacogenomics and Stratified Medicine.* 2014; 971–998.
Publisher Full Text

123. Ntekim AI, Folasire AM, Ali-Gombe M: **Survival pattern of rare histological types of breast cancer in a Nigerian institution.** *Pan Afr Med J.* 2019; **34**: 114.
PubMed Abstract | Publisher Full Text | Free Full Text

124. Kotila OA, Fawole OI, Olopade OI, *et al.*: **N-acetyltransferase 2 enzyme genotype-phenotype discordances in both HIV-negative and HIV-positive Nigerians.** *Pharmacogenet Genomics.* 2019; **29**(5): 106–113.
PubMed Abstract | Publisher Full Text | Free Full Text

125. Mary AR, Olayinka KA, Onoja AM, *et al.*: **Self-reported sulphonamide hypersensitivity reactions in adults living in Ibadan, Nigeria: A cross-sectional, community-based study.** *Niger Med J.* 2015; **56**(6): 404–410.
PubMed Abstract | Publisher Full Text | Free Full Text

126. Kiri VA, Ojule AC: **Electronic medical record systems: A pathway to sustainable public health insurance schemes in sub-Saharan Africa.** *Niger Postgrad Med J.* 2020; **27**(1): 1–7.
PubMed Abstract | Publisher Full Text

127. Waheed A, Okesola K, Afe O, *et al.*: **An Integrated and Secured Web Based Electronic Health Record.** *International Journal of Recent Engineering Science.* 2021; **8**: 19–26.
Publisher Full Text

128. Onigbogi OO, Poluyi AO, Poluyi CO, *et al.*: **Doctors' Attitude and Willingness to Use Electronic Medical Records at the Lagos University Teaching Hospital, Lagos, Nigeria.** *Online J Public Health Inform.* 2018; **10**(2): e211.
PubMed Abstract | Publisher Full Text | Free Full Text

129. Alobo IG, Soyannwo T, Ukponwan G, *et al.*: **Implementing electronic health system in Nigeria: perspective assessment in a specialist hospital.** *Afr Health Sci.* 2020; **20**(2): 948–954.
PubMed Abstract | Publisher Full Text | Free Full Text

130. Rentzsch P, Witten D, Cooper GM, *et al.*: **CADD: predicting the deleteriousness of variants throughout the human genome.** *Nucleic Acids Res.* 2019; **47**(D1): D886–D894.
PubMed Abstract | Publisher Full Text | Free Full Text

131. Kapetanovic IM: **Computer-aided drug discovery and development (CADDD): *in silico*-chemico-biological approach.** *Chem Biol Interact.* 2008; **171**(2):165–176.
PubMed Abstract | Publisher Full Text | Free Full Text

132. Duffy S, Sykes ML, Jones AJ, *et al.*: **Screening the Medicines for Malaria Venture Pathogen Box across Multiple Pathogens Reclassifies Starting Points for Open-Source Drug Discovery.** *Antimicrob Agents Chemother.* 2017; **61**(9).
PubMed Abstract | Publisher Full Text | Free Full Text

133. Paul D, Sanap G, Shenoy S, *et al.*: **Artificial intelligence in drug discovery and development.** *Drug Discov Today.* 2021; **26**(1): 80–93.
PubMed Abstract | Publisher Full Text | Free Full Text

134. Attah AF, Fagbemi AA, Olubiyi O, *et al.*: **Therapeutic Potentials of Antiviral Plants Used in Traditional African Medicine With COVID-19 in Focus: A Nigerian Perspective.** *Front Pharmacol.* 2021; **12**: 596855.
PubMed Abstract | Publisher Full Text | Free Full Text

135. Wilson WI, Peng Y, Augsburger LL: **Generalization of a prototype intelligent hybrid system for hard gelatin capsule formulation development.** *Aaps Pharmscitech.* 2005; **6**(3): E449–57.
PubMed Abstract | Publisher Full Text | Free Full Text

136. Wang X: **Intelligent quality management using knowledge discovery in databases.** *IEEE presented at International Conference on Computational Intelligence and Software Engineering.* 2009 Dec 11, 2009.
Publisher Full Text

137. Harrer S, Shah P, Antony B, *et al.*: **Artificial Intelligence for Clinical Trial Design.** *Trends Pharmacol Sci.* 2019; **40**(8): 577–591.
PubMed Abstract | Publisher Full Text