Machine learning for healthcare: A bibliometric study of contributions from Africa

Houcemeddine Turki [1,2,3,*], Houda Sekkal [4,†], Anastassios Pouris [3,5,†], Francis-Alfred Michaelangelo Ifeanyichukwu[3,6,†], Catherine Namayega[3,7,†], Hanae Lrhoul [8,†], Mohamed Ali Hadj Taieb [1,3,†], Sadiq Adewale Adedayo [3,9,†], Chris Fourie[3,10,‡], Christopher Brian Currin[3,11,‡], Mercy Nyamewaa Asiedu[3,12,13,‡], Atnafu Lambebo Tonja[3,14,‡], Abraham Toluwase Owodunni[3,15,‡], Abdulhameed Dere[3,16,‡], Chris Chinenye Emezue[3,17,‡], Shamsuddeen Hassan Muhammad[3,18,‡], Muhammad Musa Isa[3,19,‡], Mus'ab Banat[20,‡] and Mohamed Ben Aouicha[1,3,§]

[1] Data Engineering and Semantics Research Unit, Faculty of Sciences of Sfax, University of Sfax, Sfax, Tunisia

[2] Department of Computer Science, University of the People, Pasadena, California, United States of America

[3] SisonkeBiotik Research Community, Johannesburg, South Africa

[4] Mohammadia School of Engineers, Mohamed V University of Rabat, Rabat, Morocco

[5] Institute for Technological Innovation, University of Pretoria, Pretoria, South Africa

[6] Computer Engineering Department, Federal University of Technology Akure, Akure, Nigeria

[7] Division of Biomedical Engineering, Faculty of Health Sciences, University of Cape Town, South Africa

[8] School of Information Sciences, Mohamed V University of Rabat, Rabat, Morocco

[9] FSG Neuroinformatics, University of Vienna, Austria

 doi:10.20944/preprints202302.0010.v2

[10] Faculty of Computer Science and Applied Math, University of the Witwatersrand, Johannesburg, South Africa

[11] Institute of Science and Technology Austria, Klosterneuburg, Austria

[12] Jameel Clinic for AI in Healthcare, Massachusetts Institute of Technology, Cambridge, MA, USA

[13] Google Research, San Francisco, CA

[14] Centro de Investigación en Computación (CIC), Instituto Politécnico Nacional (IPN), Mexico City, Mexico

[15] Faculty of Life Sciences, University of Ilorin, Ilorin, Nigeria

[16] College of Health Sciences, University of Ilorin, Ilorin, Nigeria

[17] Technical University of Munich, Munich, Germany

[18] Department of Computer Science, University of Porto, Porto, Portugal

[19] Department of Computer Science, Nigerian Defence Academy, Kaduna, Nigeria

[20] Faculty of Medicine, Hashemite University, Zarqa, Jordan

[*] Corresponding Author

[†] Equal Contribution

[‡] Equal Contribution

[§] Senior Author

Corresponding Author:

Houcemeddine Turki

Data Engineering and Semantics Research Unit, Faculty of Sciences of Sfax, University of Sfax, Sfax, Tunisia

Tel: +21629499418

Email: turkiabdelwaheb@hotmail.fr

 doi:10.20944/preprints202302.0010.v2

Machine learning for healthcare: A bibliometric study of contributions from Africa

Houcemeddine Turki[a,b,c,*], Houda Sekkal[d,1], Anastassios Pouris[c,e,1], Francis-Alfred Michaelangelo Ifeanyichukwu[c,f,1], Catherine Namayega[c,g,1], Hanae Lrhoul[h,1], Mohamed Ali Hadj Taieb[a,c,1], Sadiq Adewale Adedayo[c,i,1], Chris Fourie[c,j,2], Christopher Brian Currin[c,k,2], Mercy Nyamewaa Asiedu[c,l,m,2], Atnafu Lambebo Tonja[c,n,2], Abraham Toluwase Owodunni[c,o,2], Abdulhameed Dere[c,p,2], Chris Chinenye Emezue[c,q,2], Shamsuddeen Hassan Muhammad[c,r,2], Muhammad Musa Isa[c,s,2], Mus'ab Banat[t,2], Mohamed Ben Aouicha[a,c,3]

[a] *Data Engineering and Semantics Research Unit, Faculty of Sciences of Sfax, University of Sfax, Sfax, Tunisia*
[b] *Department of Computer Science, University of the People, Pasadena, California, United States of America*
[c] *SisonkeBiotik Research Community, Johannesburg, South Africa*
[d] *Mohammadia School of Engineers, Mohamed V University of Rabat, Rabat, Morocco*
[e] *Institute for Technological Innovation, University of Pretoria, Pretoria, South Africa*
[f] *Computer Engineering Department, Federal University of Technology Akure, Akure, Nigeria*
[g] *Division of Biomedical Engineering, Faculty of Health Sciences, University of Cape Town, South Africa*
[h] *School of Information Sciences, Mohamed V University of Rabat, Rabat, Morocco*
[i] *FSG Neuroinformatics, University of Vienna, Austria*
[j] *Faculty of Computer Science and Applied Math, University of the Witwatersrand, Johannesburg, South Africa*
[k] *Institute of Science and Technology Austria, Klosterneuburg, Austria*
[l] *Jameel Clinic for AI in Healthcare, Massachusetts Institute of Technology, Cambridge, MA, USA*
[m] *Google Research, San Francisco, California, United States of America*
[n] *Centro de Investigación en Computación (CIC), Instituto Politécnico Nacional (IPN), Mexico City, Mexico*
[o] *Faculty of Life Sciences, University of Ilorin, Ilorin, Nigeria*
[p] *College of Health Sciences, University of Ilorin, Ilorin, Nigeria*

[*] Corresponding author.
Email addresses: turkiabdelwaheb@hotmail.fr (Houcemeddine Turki), houda.sekkal@gmail.com (Houda Sekkal), anastassios.pouris@up.ac.za (Anastassios Pouris), michaelangeloifeanyichukwu@gmail.com (Francis-Alfred Michaelangelo Ifeanyichukwu), catherinenamayega@gmail.com (Catherine Namayega), hlrhoul@gmail.com (Hanae Lrhoul), mohamedali.hajtaieb@fss.usf.tn (Mohamed Ali Hadj Taieb), sadiq.adedayo@univie.ac.at (Sadiq Adewale Adedayo), chris.fourie@pm.me (Chris Fourie), chris.currin@gmail.com (Christopher Brian Currin), mnasiedu@csail.mit.edu (Mercy Nyamewaa Asiedu), atnafuatx@gmail.com (Atnafu Lambebo Tonja), owodunniabraham@gmail.com (Abraham Toluwase Owodunni), daabiola3@gmail.com (Abdulhameed Dere), chris.emezue@gmail.com (Chris Chinenye Emezue), shamsuddeen2004@gmail.com (Shamsuddeen Hassan Muhammad), mm.isa@nda.edu.ng (Muhammad Musa Isa), mossab748@gmail.com (Mus'ab Banat), mohamed.benaouicha@fss.usf.tn (Mohamed Ben Aouicha)
[1] Equal Contribution.
[2] Equal Contribution.
[3] Senior Author.

[q] *Technical University of Munich, Munich, Germany*
[r] *Department of Computer Science, University of Porto, Porto, Portugal*
[s] *Department of Computer Science, Nigerian Defence Academy, Kaduna, Nigeria*
[t] *Faculty of Medicine, Hashemite University, Zarqa, Jordan*

Abstract

In the last decade, machine learning has gained significant momentum in healthcare, especially for advanced diagnostics and improved patient care. This technology has immense potential for Africa, where resource scarcity is a significant challenge. However, it is unclear how much African institutes are contributing to machine learning research, which is crucial for addressing the unique challenges faced by the continent. In this bibliometric study, we examined African contributions to research publications related to machine learning for healthcare indexed in Scopus between 1993 and 2022. We identified 3,772 research outputs, with most of them published after 2020 due to the African COVID-19 scholarly research response. Deep learning (*Convolutional Neural Networks*), classical machine learning (*Support Vector Machine*, *Decision Tree*, *k-Nearest Neighbors*, and *Random Forest*), and transfer learning are the most commonly used learning algorithms in these publications. The works mainly deal with lethal, highly contagious, and chronic diseases. The popular application areas include computer-aided diagnosis, classification, prediction, and biomedical image processing. North African countries lead with 64.5% of publications for the reported period, but Sub-Saharan Africa is also rapidly increasing its output. Our analysis reveals that research output is correlated with international support in the form of funding and collaborations. Understanding African research contributions to machine learning for healthcare is crucial for surveying the academic landscape, forming stronger research communities, and providing advanced and contextually aware biomedical access to Africa.

Keywords: Machine learning, Scientometrics, Africa, Research community, Open science, Health informatics

 doi:10.20944/preprints202302.0010.v2

1. Introduction

Machine learning is a sub-field of artificial intelligence (AI) that focuses on the development of algorithms that learn statistical models of data (Jordan & Mitchell, 2015). These resulting machine learning models allow for reasoning over existing data and for making predictions from new data. There are many machine learning models that have been developed for decades with varying levels of accuracy and applicability to science, technology, commerce, and more. Of particular relevance for this study is the research and application of machine learning for healthcare: a promising avenue for advanced diagnostics, rapid drug design, and improved patient care. These aspects are of particular importance in Africa where many countries are resource-scare or underdeveloped and where reaching remote communities with advanced expertise is a challenge within existing frameworks.

Machine learning has advanced significantly in the last decade, both in its theory and real-world applications. Within the past 3 years, the COVID-19 pandemic has seen increased focus placed on health and healthcare systems, with flaws and opportunities exposed. Africans represent a vast and growing intellectual capital developing ideas and solutions that are locally impactful and relevant globally (Currin et al., 2019). These three aspects form a confluence of opportunities for applying the latest technological tools in a vital field within an intellectually rich but historically exploited context.

In recent years, the application of machine learning across many medical disciplines has been prolific (Shailaja et al., 2018; Qayyum et al., 2021; Deo, 2015; Jiang et al., 2017; Lidströmer & Ashrafian, 2022). However, the majority of government-approved AI tools for healthcare are limited to radiology and cardiology (Reuter, 2022). If we are to see the benefits of machine learning for healthcare, it is important to better understand the complex path from research to approved application (Sculley et al., 2015; Reuter, 2022). Furthermore, machine learning research communities often neglect under-served geographical and socioeconomic areas and instead tend to focus on higher-income countries

 doi:10.20944/preprints202302.0010.v2

Chen et al. (2021b). To highlight this disparity, Africa generally has been deeply impacted by a devastating colonial history Heldring & Robinson (2012); Ocheni & Nwankwo (2012); Ziltener & Künzler (2013). An example of this impact on research production is that there are fewer Ph.D. graduates coming from lower-income countries (Jowi et al., 2018). A more comprehensive understanding of the specific patterns for how machine learning and healthcare research are connected and published can help toward improving the direction, accessibility, and eventual realisation of this research in an African context.

With an increasing need for understanding, connecting, and growing the present landscape of machine learning for healthcare in Africa, SisonkeBiotik has sprung up as a new grassroots community focusing on the holistic integration of both machine learning and healthcare in an African context. In this bibilometric study, we start the process of understanding the current academic state of researchers in Africa developing and applying machine learning for healthcare.

We believe this work forms a crucial first step towards identifying promising avenues of research and collaboration that will form the basis for future work for the decades ahead.

We begin by providing a literature review of previous bibliometric studies on machine learning, healthcare, and health informatics (Section 2). Then, we explain our methods for the analysis of the bibliographic metadata of African research publications on machine learning for healthcare (Section 3). After that, we show data insights into how African biomedical machine learning research is conducted and published (Section 4). Finally, we draw conclusions for our analysis and provide future directions for this research work (Section 5).

2. Related Works

As of 2014, there was not an extensive number of scientometric studies in general from the African continent. Collaborative studies, in particular, were still in the early stages of development. Furthermore, the scientometric analysis identified that research in Africa emphasizes medical and natural resources disciplines to the detriment of disciplines supporting knowledge-based economies and societies. However collaborative patterns in Africa are substantially higher than in the rest of the world (Pouris & Ho, 2014). Pouris & Ho (2014) comment that it is important to note that the share of articles co-authored with at least one author outside the African continent is many times bigger than the share of articles co-authored with only authors on the African continent.

Pouris & Ho (2014) asks rhetorically: "What drives researchers, say in Botswana and Zimbabwe, to produce more than 74% of their collaborative publications outside of Africa? South African universities are a few hours away by car. Europe and the USA are a number of hours away by plane. Similarly, why does Egypt collaborate almost exclusively with non-African countries?". They argue that African collaboration is not driven by local researchers searching for collaborators in general, but by the availability of resources and interests outside the continent. We stand for the importance of global collaborations and partnerships. However, as demonstrated by the decolonizing global health movement, it is important that research partnerships focusing on deployments on the African continent are driven by African researchers, and that "inequitable power dynamics and neocolonialist assumptions" do not unevenly influence machine learning for Africa initiatives (Eichbaum et al., 2021).

Pouris & Ho (2014) end by raising a number of policy concerns:

- Should Africa's science and development not be better served by the creation of regional research and innovation systems, that as an example, is aiming to create an African Research Union?
- How do the high dependencies on non-African collaboration affect the continent's research evolution and priorities?
- Is African research individualism and inspiration stifled by excessive collaboration?

More recently, by examining the period 2001-2018, Sooryamoorthy (2022)

 doi:10.20944/preprints202302.0010.v2

confirmed the continued above-mentioned overemphasis of African research in particular disciplines such as healthcare and natural resources. Additionally, they show that research efforts are concentrated in only a small set of particular African countries. They estimated that approximately three-quarters of the publications originated from six countries, namely, South Africa, Egypt, Tunisia, Algeria, Nigeria and Morocco. Of these countries, only two of which are in Sub-Saharan Africa (SSA). In other words, 11% of African countries produced 74% of the total number of publications in science from Africa. This leaves the remaining 89% of African countries publishing only a meagre 26% of the total for the continent.

Sooryamoorthy (2022) argued that the substantial contribution Africa has made to research areas such as tropical medicine, parasitology, infectious diseases and immunology is seemingly quite impactful for world science. Furthermore, they suggested that these are crucial research areas for Africa in general and for some African countries in particular. Given the presence of diseases and the sporadic outbreaks of epidemics that are common in several African countries, investment in these research areas is important in order for them to address such problems.

The literature related to health in Africa is succinctly summarised by the World Bank and Elsevier in 2013 (Blom et al., 2015). They state: "The impressive improvement in SSA's research capacity in the Health Sciences demonstrates that persistent support and funding from development partners and governments pays off. There is clearly a large scientific talent base in Africa, but this needs to be trained and nurtured." They identify that a very large share of SSA research is the result of international collaboration. The World Bank notes that the high reliance on international collaboration for research signals that there is a lack of internal research capacity, funding and the critical mass to produce international quality research in isolation.

CAAST-NET investigated the impact of Framework Programmes on African health science (Pouris, 2017). It is identified that health articles constitute 36% of all articles produced in Africa during 2015. The percentage across this region

 doi:10.20944/preprints202302.0010.v2

is higher than the countries of India (28%), China (27%) and Russia (13%). This indicates the importance and sensitivity of the topic for Africa. Investigation around the source of funding for articles co-authored among African and EU authors shows that the Wellcome Trust was the most often mentioned funder. Pouris (2017) identified that the CAAST-NET and CAAST-NET PLUS supported by the Framework Programmes (a set of funding programmes created by the European Union) appear to have been successful in encouraging more and better quality bi-regional science, technology and innovation cooperation for enhanced outcomes related to health. Success is of particular importance as Africa has the lowest life expectancy in the world. Life expectancy is an overarching health indicator for the post-2015 development agenda. Hence, it is suggested that similar approaches can be utilised to develop African research capacity and cooperation in additional fields of common interest and priority beyond just healthcare.

Artificial intelligence (AI) and machine learning research related to healthcare in Africa have been reported by a number of articles. Owoyemi et al. (2020) described the main medical AI (MAI) activities in health from Africa. These include MAI deployed in Kenya that improved health worker–patient interaction quality with evidence of an increased number of symptoms elicited (Hunter et al., 1989) and an investigation to improve the detection of common and potentially blinding eye disorders.

MAI was also piloted in Egypt in 1986 (Kastner et al., 1984). Similarly, in The Gambia, a probabilistic decision-making system assisted rural health workers to identify life-threatening conditions in outpatient clinics. The MAI performed tolerably well in detecting 88% of cases. Computerized Aid To Treat (CATT) was also used in drug prescriptions in South Africa by nurses based on a cost-and-effectiveness algorithm (Byass, 1987; Forster, 1992). The authors argue that the healthcare application of AI in Africa has only seen a few pilots and test cases in recent times. They refer to investigations by Moyo et al. (2018) in South Africa, Onu et al. (2019) in Nigeria, Bellemo et al. (2019) in Zambia, and Busari & Adebayo (2018), Brunskill & Lesh (2010) in Africa generally.

 doi:10.20944/preprints202302.0010.v2

A number of authors attempt to develop suggestions for how AI can leapfrog in Africa for the field of health. Akpanudo (2022) argues that the automated nature of artificial intelligence systems makes these systems uniquely suited for the challenge of delivering healthcare to remote and under-resourced settings. For example, if an algorithm can supplement the capabilities of community health workers (CHWs) using human-in-the-loop approaches, it could facilitate task shifting from limited expert health specialists to CHWs (Baxter et al., 2021). Even though AI systems may hold great promise for the future of healthcare in Africa, there are limited active use cases on the continent. This is not surprising given the challenges in practical AI deployment in healthcare globally (Kelly et al., 2019). Research-wise, only a limited number of countries on the continent have embraced and pursued artificial intelligence in healthcare. These countries are notably Kenya, South Africa, Nigeria, Ghana, and Ethiopia. Even in these countries, many healthcare providers are unaware that these systems exist and are currently being used in their respective countries. There is also a lack of active research into algorithmic biases and fairness, identified in AI applications (Chen et al., 2021a) to healthcare may manifest in Africa. Additionally, foundational digital health infrastructure to safely collect, store and make accessible health data, a requirement to enable the research and use of AI, is largely not present across Africa. Akpanudo (2022) argues that "without the availability of needed infrastructure for secure data collection with data privacy, education, and good governance, the dream of artificial intelligence leapfrogging healthcare in Africa into the future may not be realized."

3. Materials and Methods

To assess the contribution of Africa to research efforts about machine learning and healthcare (Process in Figure 1), we queried the controlled bibliographic database *Scopus*. We used Scopus instead of an automatically-generated bibliographic database, despite the latter potentially having a better coverage of biomedicine-related scholarly publications (Halevi et al., 2017), because auto-

 doi:10.20944/preprints202302.0010.v2

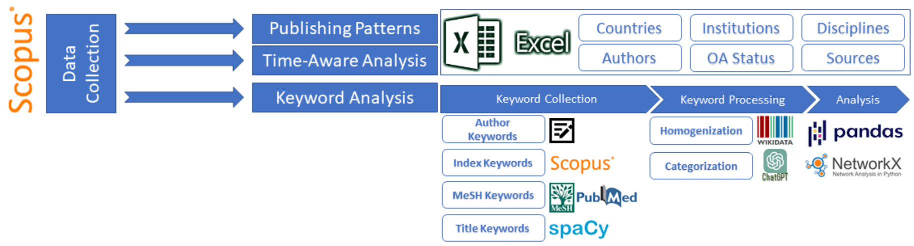

Figure 1: Process for the assessment of African research on Biomedical Machine Learning.

matically generated bibliographic databases do not check indexed publications for scientific relevance and misconduct. Publications from these automatically generated databases could potentially involve several non-peer-reviewed publications as well as some duplicate or non-formatted metadata. Therefore we do not believe them to be suitable as stand-alone sources for evaluating research efforts about a given topic (Halevi et al., 2017). Among the controlled databases, *Scopus* is more inclusive than *Web of Science* and *PubMed* (AlRyalat et al., 2019) and its coverage is even comparable to those of automatically generated bibliographic databases in *computer science* and *health and medical sciences* (Martín-Martín et al., 2020).

Using a specific query [4] we identify biomedical machine learning research publications. We filter this database response to only consider publications that include at least one contributing author from an African institution. Subsequently, we eliminate upcoming research publications that will be issued in 2023 and that are already indexed by *Scopus* to ensure the fair comparison of the authors, institutions, and countries of discarded publications to the ones choosing different venues that do not perform premature indexing. We retrieve the bibliographic metadata For these filtered publications from *Scopus*. This

[4] Query: TITLE-ABS-KEY(""Machine Learning"" OR ""Deep Learning"") AND TITLE-ABS-KEY(Medic* OR Clinic* OR Biomed* OR Health*).

 doi:10.20944/preprints202302.0010.v2

retrieved information includes titles, abstracts, keywords, source titles, author information, external identifiers, funding information, publication type, open access status, references, and the number of citations.

For these publications, we evaluate general patterns of publishing in section 4.1, including comparing North Africa (*Tunisia*, *Algeria*, *Morocco*, *Libya*, and *Egypt*) to Sub-Saharan Africa. In section 4.2, we carry out a time-aware analysis, outlining the evolution of various features through four periods: *Until 2013*, *2014-2016*, *2017-2019*, and *2020-2022*.

Finally, we perform keyword analysis for the research publications to identify the main topics of interest for the African community and how they are solved using machine learning and databases (Section 4.3). We will consider the Author Keywords and Index Keywords as retrieved from *Scopus* and we will also retrieve Medical Subject Headings (MeSH) Keywords) for the publications indexed by PubMed[5], a biomedical bibliographic database maintained by the National Center for Biotechnology Information (NCBI). For this, we will use Biopython as a Python Library for retrieving MeSH Keywords based on PubMed IDs (Turki et al., 2022a). We consider the qualifiers of the MeSH terms as independent keywords in our analysis and we eliminate asterisks (*) from every MeSH keyword. As well, we extract noun phrase chunks from the titles of the scholarly publications using *SpaCy* as a Python Library and *en core web sm* as a language model (Chantrapornchai & Tunsakul, 2021). We clean these noun phrases by eliminating stop words as returned by the *adv* Python Library, eliminating punctuations, and singularizing them by getting their lemmas. After that, we homogenize all the keywords by reconciling them to corresponding items in Wikidata[6], an open and collaborative large-scale knowledge graph by the Wikimedia Foundation. For this, we use the *Wikibase Integrator* Python Library to return the homogenized name (label) for every keyword (Turki et al., 2022b). All the reconciliations are validated by human users and

[5] `https://pubmed.ncbi.nlm.nih.gov/`
[6] `https://www.wikidata.org`

 doi:10.20944/preprints202302.0010.v2

all the Keywords featured less than nine times are eliminated from our analysis. The common keywords that have not been matched to Wikidata items are manually reconciled by human annotators. When the data collection is finished, a category is assigned to every entity using *ChatGPT*, a large language model-driven chatbot released by OpenAI (Cahan & Treutlein, 2023). The generated metaclasses are validated by human experts for consistency and then generalized again to have four final keyword categories (i.e., *STEM*[7], *Biomedicine*, *Social Sciences*, and *Other*). Then, we identify the most common topics per keyword category and we study the distribution of their mentions per period (*Overall*, *Until 2019*, and *2020-2022*), per type (*journal article* and *other*), per region (e.g., *North Africa* and *Only Sub-Saharan Africa*), per opên-access status (*OA* and *non-OA*), and per citation status (*Cited more than 10 times* and *other*) using the *Pandas* Python Library (McKinney, 2011). We also generate the co-occurrence network of the most common fifty keywords using the NetworkX Python Library[8] (Hagberg et al., 2008). To let the interpretation of the keyword co-occurrence networks easier, we also generated heatmap matrices for the identified keyword associations using the *Seaborn* Python Library[9] (Waskom, 2021). We set the heatmap representations with logarithmic-scale color bars because the keyword distribution in a set of publications follows a power law (so-called *Zipf's law*) (Bartol & Stopar, 2015). In the following section 4, we present our results alongside a discussion of those results.

[7] Science, Technology, Engineering, and Mathematics.

[8] Credit for source code: `http://andrewtrick.com/stormlight_network.html` and `https://towardsdatascience.com/customizing-networkx-graphs-f80b4e69bedf`.

[9] Credit for source code: `https://towardsdatascience.com/better-heatmaps-and-correlation-matrix-plots-in-python-41445d0f2bec` and `https://stackoverflow.com/questions/74206551/seaborn-diverging-palette-fix-mid-and-end-points`.

 doi:10.20944/preprints202302.0010.v2

4. Results and Discussion

4.1. Publishing Patterns

As of October 4, 2022, using our methods as described in section 3, we identified 3,772 research publications for African academic papers related to machine learning (ML) and healthcare. Throughout this section we refer to these publications as African machine earning for health publications, making use of acronym *AML4H* to describe this set of publications. We observe that AML4H publications increased exponentially from 2011 as shown in Figure 2 (Blue). Figure 2 is depicted in a log scale, highlighting this exponential growth. Globally, research interest in ML techniques began to rise after 2013 as a result of their demonstrable efficacy on popular standard benchmark tasks (Li et al., 2020; dos Santos et al., 2019). This rise in interest is likely to have contributed to our observed exponential rise in ML and healthcare research in Africa. We suspect that regional factors that include the development of many consortia, funding initiatives, and organisations to enhance African scholarly contributions to Biomedical Informatics have had an impactful contribution to this exponential rise of AML4H publications. Examples of these contributions can be seen from *H3Africa* (Mulder et al., 2018), *BETTEReHEALTH* (Larbi et al., 2022) and *HELINA* (Tchuitcheu et al., 2020) initiatives. Using biomedical informatics techniques, Luna et al. (2014), describe six broad challenges that are considered to impact the general development of physical and digital infrastructure for information technology in Africa. As information technology infrastructure is a fundamental requirement to enable ML research, these challenges can be considered to impact ML for health research as well. Luna et al. (2014) note that for successful implementation of health informatics, knowledge of the challenges to be faced is an important factor.

The majority of AML4H publications (77.7%) occurred after the beginning of the COVID-19 pandemic on the continent in 2020. These exhibit a shift toward topics related to the pandemic, such as ML for virology and ML for epidemiology. The increase in African research productivity for healthcare since 2020 is in

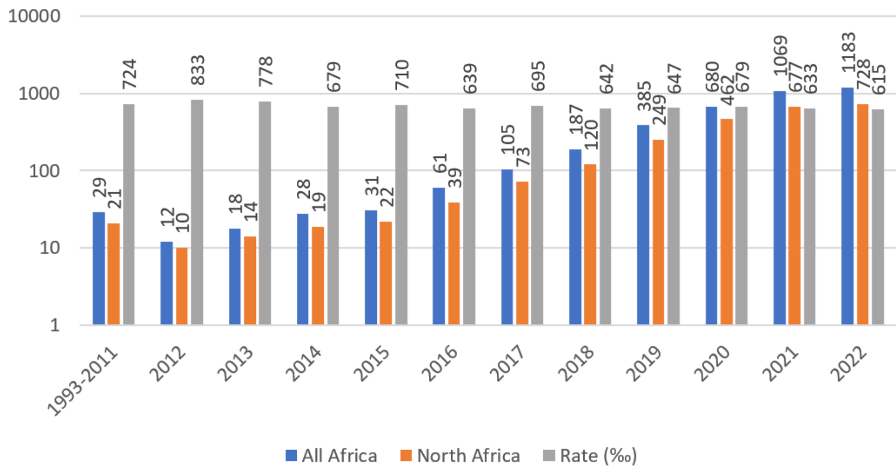

Figure 2: Number of Scopus-indexed African publications about ML for healthcare (log scale) per year

line with the global efforts to solve urgent matters related to the outbreak, particularly those related to timely data curation and management (Zhao et al., 2022; Bayoudhi et al., 2022). This also reaffirms previous observations that disease outbreaks can lead to the growth of research in underdeveloped fields (Cheng et al., 2012) for affected nations (Turki et al., 2023). However, despite this increased productivity in the field of ML and healthcare, the impact of the COVID-19 pandemic on global research production more generally is mixed and will likely be an area of active research for some time (Harper et al., 2020; Gao et al., 2021).

We observe that contributions from North Africa (*Tunisia*, *Algeria*, *Morocco*, *Libya*, and *Egypt*) make up 64.5% of the total AML4H publications between 1993 and 2022 (2,434 publications). This is substantially larger than any other African region. However, recently the publication ratio for North Africa declined from over 70% between 1993 and 2015 to below 70% in 2016, dropping to 61.5% by 2022. An increase in AML4H publications from Sub-Saharan Africa has shifted this ratio rather than a decline in AML4H publications from North

 doi:10.20944/preprints202302.0010.v2

Africa.

In Sub-Saharan Africa recent establishment of government-led initiatives to introduce Telemedicine and Digital Health, mainly in public hospitals (Tapera & Singh, 2021) and ongoing international development funding for research (Shaffer et al., 2019) are possible factors contributing to this increase in AML4H publications. There is still a persistent lack of digital health infrastructure in sub-Saharan Africa. Efforts to introduce various forms of electronic health record systems have started to alleviate this. However, the adoption and scaling of these systems may likely still be slow relative to higher resource contexts (Dodoo et al., 2021). The limited availability of digital health infrastructure consequently limits the development of machine learning research (Dodoo et al., 2021).

The funding information from Dodoo et al. (2021) shows a correlation between AML4H publications (particularly Sub-Saharan Africa) and foreign funding, as shown in Figure 3. The National Institutes of Health (United States of America) has funded 145 AML4H publications. This funding stems mainly from the *Harnessing Data Science for Health Discovery and Innovation in Africa* (DS-I Africa) Program. This program aims to enhance data science in Africa for healthcare, public health, and biomedical research (Wonkam, 2021). Wellcome Trust (United Kingdom) has supported 48 AML4H publications. This is made possible due to the *Developing Excellence in Leadership, Training, and Science* (DELTAS) Africa Initiative co-organized with the African Academy of Sciences (Carpenter et al., 2022). The Medical Research Council (MRC) of the United Kingdom (UK) is also featured as a major funding institution for African research funding 36 AML4H publications. In particular, the *UK MRC funded scholarship* program is likely to have been an important contributing factor to these publications. The MRC maintains long-term support for graduate students of Sub-Saharan Africa (Carpenter et al., 2022). The European Commission is a funder for 34 AML4H publications. This is related to the prioritization of Africa-focused research thanks to the *Horizon Europe* Framework Programme providing 350 million euros to fund research projects including Europe-Africa

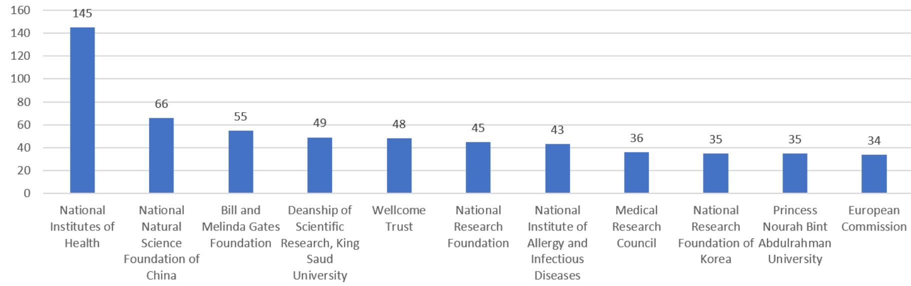

Figure 3: Main funding entities for AML4H publications. National Institute of Health (United States of America), National Research Foundation (South Africa)

collaborations (Estreguil & Buschke, 2022). Furthermore, the Bill and Melinda Gates Foundation (United States) provided funding for 55 AML4H publications. As a non-governmental charity organization, this foundation is interested in encouraging research projects that translate health-related knowledge into life-saving interventions, particularly in developing countries where access to clinical information and consistent health infrastructure is very limited (Fadlelmola et al., 2021). Moreover, the National Institute of Allergy and Infectious Diseases from the United States of America financially contributed to the development of 43 AML4H publications. This is done within the framework of the contribution to studies about infectious and respiratory diseases in Africa (Fauci et al., 2005). Beyond these funding bodies that have a broad and global reach, several national research institutions with a generally narrower scope and more localised focus have been identified among the main funders of AML4H publications as shown in Figure 3. *King Saud University* and *Princess Nourah Bint Abdulrahman University* from Saudi Arabia have respectively funded 49 and 35 AML4H publications. A likely contributing factor to this observation is the mass funding provided by Saudi Government to local research institutions to independently organize scholarly projects and publish high-quality papers toward achieving better standings in world university rankings (Turki et al., 2019). The National Natural Science Foundation of China and the National Research Foundation of Korea (South Korea) have respectively supported 66 and 35 research AML4H publications. By contrast to those of the Saudi institutions, these are centralized and government-led, supervising domestic national research funding (Zhou et al., 2020). These two institutions are not providing programs exclusively for foreign scientists unlike the National Institutes of Health (United States) (Zhou et al., 2020). Rather, these two institutions fund African research papers when Chinese or Korean scientists are significantly involved (Zhou et al., 2020). Their presence as primary funders is likely motivated toward growing the presence of BRICS Countries in the research landscape in Africa, particularly in Health Informatics (Tapera & Singh, 2021).

In a similar context, we find that the National Research Foundation (NRF) of South Africa is the only centralized, government-led institution based in Africa that is significantly funding AML4H with 45 publications. However AML4H publications supported by this local funding appear to draw less attention than those supported by international funders when considering patterns of citations. The number of citations garnered by AML4H publications supported by the South African NRF are less numerous than citations for AML4H publications supported by international funders (Zhou et al., 2020). When we consider South African research in Health Informatics, it seems that international funding and collaboration have a higher impact on citation count as compared to local governmental support (Arvanitis et al., 2022; Tapera & Singh, 2021). This observation seems to apply to all of the continent and not just South Africa. This disparity in the effectiveness of local and international funding, at least when considering patterns of citation, may suggest that a general reform for government-led research funding across the continent should be investigated (Tapera & Singh, 2021; Arvanitis et al., 2022).

Patterns of funding shape the international collaboration networks that African institutions develop to conduct research for ML and healthcare. We summarise the contributions of non-African countries to AML4H publications in Figure 4. Effectively, we see that the United States of America and Saudi

Arabia dominates international collaborations in this context respectively with 550 and 532 publications. Although the flexibility of these two countries can explain in part their relative domination of African Biomedical ML research, this fact can be due to other factors. The research policy of Saudi Arabia emphasizes international collaborations by contrast to other major funding countries (Hu et al., 2020). Saudi Arabia has also established for decades a tradition of research collaboration with North Africa through the mediation of Egypt thanks to geographic proximity and their joint affiliation to the Arab region (Landini et al., 2015). As for the United States of America, it is the most prolific country in the world for research on biomedical informatics (Tran et al., 2019) as well as on ML (Li et al., 2020). It is also behind the establishment of multiple international biomedical research consortia that encourage the move to digital health (Vogel et al., 2019). We also notice that many European countries significantly contribute to African research on the matter: the United Kingdom (359 publications), France (257 publications), Germany (158 publications), Spain (126 publications), Netherlands (82 publications), and Italy (73 publications). Financial support from the European Commission and local charity organizations like *Wellcome Trust* can explain this finding, mainly for the United Kingdom. Nevertheless, probably, this fact is also due to the existence of these countries among the most productive ones in ML (Li et al., 2020) and health informatics (Tran et al., 2019) research: United Kingdom (3$^{\text{rd}}$ in deep learning), France (6$^{\text{th}}$ in health informatics), Germany (5$^{\text{th}}$ in deep learning, 4$^{\text{th}}$ in health informatics), Spain (8$^{\text{th}}$ in health informatics), Netherlands (13$^{\text{th}}$ in health informatics), and Italy (10$^{\text{th}}$ in deep learning, 3$^{\text{rd}}$ in health informatics). Similarly, we can find that the presence of China (198 publications), Canada (150 publications), Australia (110 publications), and South Korea (97 publications) among the main collaborating countries with Africa in Biomedical ML research can be explained by the status of these countries as highly productive ones in ML (Li et al., 2020) and health informatics (Tran et al., 2019) research and by the existence of nationwide funding institutions in these countries (Zhou et al., 2020). However, the identification of India (390), Pakistan (102), and the United Arab Emirates

(88) among the main collaborators of Africa in this research area is quite surprising as these countries have not been featured as sponsors for African research papers. For the United Arab Emirates, the situation is quite similar to the one of Saudi Arabia as geographic proximity to North Africa enables the country to easily establish research collaborations with North Africa (Turki et al., 2019). The United Arab Emirates has large flexibility in establishing research collaborations, particularly higher than the one of Saudi Arabia (Sarwar & Hassan, 2015). However, its limited efficiency to contribute to African Biomedical ML research outputs is mainly linked to the considerably smaller scholarly productivity of the United Arab Emirates and the trend of the country to establish collaborations with Asian neighbors rather than with North African countries by contrast to Saudi Arabia that maintains a robust research collaboration with Egypt and consequently with North Africa (Sarwar & Hassan, 2015). As for Pakistan, it is among the best-published Islamic countries in computer science and it has long-term research collaboration traditions with Saudi Arabia (Sarwar & Hassan, 2015). Its contribution to African research outputs is probably an effect of the involvement of Saudi Arabia in biomedical ML research in Africa. Concerning India, it is among the best ten most published countries in ML (Li et al., 2020) and health informatics (Tran et al., 2019). The limited history of research collaborations between India and Africa (Tran et al., 2019; Li et al., 2020) except for several joint projects between South-Eastern Africa and India (Toivanen & Ponomariov, 2011) proves that this tendency is new and is probably a consequence of COVID-19 where Indian scientists were invited to join large-scale research projects online for their proficiency in this research field. This is confirmed for global COVID-19 research where India is identified as the third research collaborator of the Arab countries with a strong scholarly bond with Egypt (Zyoud, 2021).

From what we have already discussed, it seems that collaborating countries tend to be selective towards North Africa or Sub-Saharan Africa. There are limited countries that develop scholarly collaboration programs for the African continent. This is confirmed through the computation of the rate of the pa-

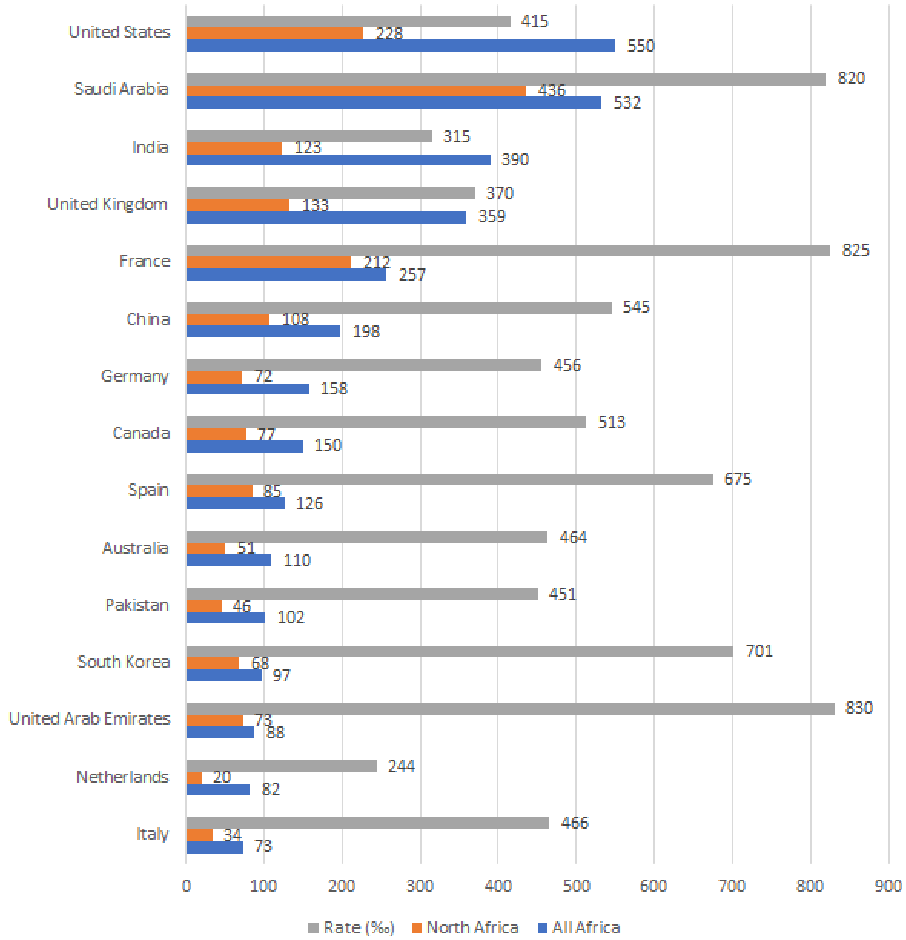

Figure 4: Overview of non-African countries contributing to AML4H publications.

pers coauthored with North African institutions among the paper coauthored by a non-African country with African ones (Grey in Figure 4). Saudi Arabia, France, Spain, South Korea, and the United Arab Emirates are biased towards establishing collaboration with North Africa (¿55% coauthored with North Africa). Yet, the United States of America, India, the United Kingdom, and the Netherlands are favoring collaboration with Sub-Saharan Africa (¡45% coauthored with North Africa). While the bias of Saudi Arabia and the United Arab Emirates can be explained by the close relations with other Arab nations including North African ones (Turki et al., 2019; Sarwar & Hassan, 2015), the exclusive collaboration of France towards North Africa is rather due to the long-term effect of the colonization of Tunisia, Algeria, and Morocco by this country (Landini et al., 2015). The similarity of the higher education and research systems between the best-published three African countries and the use of French as the main language of scholarly research in these nations facilitate the establishment of joint research programs between France and North Africa (Landini et al., 2015). The lack of collaboration between France and Sub-Saharan African countries that have been formerly colonized by it like Senegal, Benin, Cameroon, and Ivory Coast is explained by the current lack of development of research in health informatics (Tran et al., 2019) and ML (Li et al., 2020) in these countries. As for Spain and South Korea, the bias is rather linked to the establishment of government-led bilateral research cooperation programs between these two countries and North Africa, particularly Tunisia and Morocco[10] (Currie-Alder et al., 2017). Spain is also a country that is located very close to North Africa and has consequently the ability to easily establish research collaborations with this region through Morocco (Toivanen & Ponomariov, 2011; Landini et al., 2015). Concerning the United States of America, the United Kingdom, and probably the Netherlands, their higher interest in Sub-Saharan Africa is mainly motivated by the funding programs that are exclusively done by these three countries that disregard North Africa due to the assumption that North African countries are

[10] Examples: `https://cutt.ly/fB0l9Ie` and `https://cutt.ly/TB0l6hb`.

richer than South African ones although all Africa is currently underdeveloped (Elliott et al., 2015). However, the bias of India towards Sub-Saharan Africa is not explained by funding because research collaborations with North African institutions can be easier through Saudi support (Sarwar & Hassan, 2015). It is rather explained by a historical scholarly association between South-Eastern Africa and India (Toivanen & Ponomariov, 2011) and by the invitation of Indian individuals by institutions in Sub-Saharan Africa to join projects for their efficiency in computer science research (Tran et al., 2019; Li et al., 2020).

Beyond regional and political motivations behind the bias in collaboration with non-African countries and the choice of funding sources between North Africa and Sub-Saharan Africa, it can be explained by the lack of coordination of research between the two regions. Only twelve Sub-Saharan African nations contributed to North African research about ML for healthcare: *South Africa* (12 publications), *Nigeria* (11 publications Sudan (6 publications), *Gabon* (4 publications), *Ethiopia* (3 publications), *Ghana* (3 publications), *Kenya* (3 publications), *Senegal* (2 publications), *Cameroon* (1 publication), *Congo* (1 publication), *Malawi* (1 publication), and *Tanzania* (1 publication). Such behavior is unfortunately historical and common to all research fields in Africa (Toivanen & Ponomariov, 2011). More collaboration between North Africa and Sub-Saharan Africa is required so that all countries can benefit from all the scholarly resources .

When seeing the research productivity of African countries (54 nations), we found that only eight countries published more than 100 publications, and twelve countries published more than 40 publications as shown in Figure 5. These countries are led by *Egypt* (1255 publications), *South Africa* (505 publications), *Morocco* (446 publications), *Tunisia* (407 publications), *Algeria* (327 publications), and *Nigeria* (270 publications). This proves relative domination of African ML research for healthcare by North Africa over Sub-Saharan Africa. This is mainly due to the leading position of *Egypt* and *South Africa* in scholarly research (Toivanen & Ponomariov, 2011), particularly the one related to biomedical informatics (Tran et al., 2019) and ML (Li et al., 2020). The relative higher standings of *Tunisia*, *Algeria*, *Morocco*, and *Nigeria* is confirmed by previous findings on research productivity in Africa (Toivanen & Ponomariov, 2011). As well, only four French-speaking countries are represented among the fifteen most published African nations: *Morocco*, *Tunisia*, *Algeria*, and *Rwanda*. This can partly be explained by the fact that quite all the research production about ML and healthcare in Africa is written in English (3,788 out of 3,789). This constitutes a language barrier for French-speaking countries where higher education is mostly delivered in French (Toivanen & Ponomariov, 2011). This is also explained by the lack of collaboration between French-speaking North Africa and the rest of French-speaking Africa which has an underdeveloped research infrastructure (Toivanen & Ponomariov, 2011). It is also important to know that only four countries from the fifteen most productive ones have a population that is inferior to 30 million citizens: Zambia (19.4 millions), Rwanda (12.9 millions), Tunisia (11.7 millions), and Libya (6.9 millions) (United Nations Economic Commission for Africa, 2021). This confirms the effect of the population size on the research productivity of a country (Rahman & Fukui, 2001). When adjusting the research productivity by the population size for the considered nations, we found that only four countries achieve a rate of research publications per 1 million citizens superior to 8: *Tunisia* (34.64), *Egypt* (12.06), *Morocco* (12.04), and *South Africa* (8.30). This means that the registered advantage of these four continents over the remaining parts of the continent is not due to population size. This is rather due to other factors such as the higher quality of research capacities in these countries, the existence of a robust research infrastructure, and the development of efficient research policies and funding programs (Uwizeye et al., 2022). This can be also related to the higher density of medical workers in these four countries (¿20 per 10,000 citizens) (Boniol et al., 2022). The relatively limited productivity of several countries with high rate of medical specialists like *Libya* and *Mauritius* is the lack of a computer science research community in these nations (Tran et al., 2019; Li et al., 2020).

When seeing the effect of the Gross Domestic Product (GDP) in USD on the

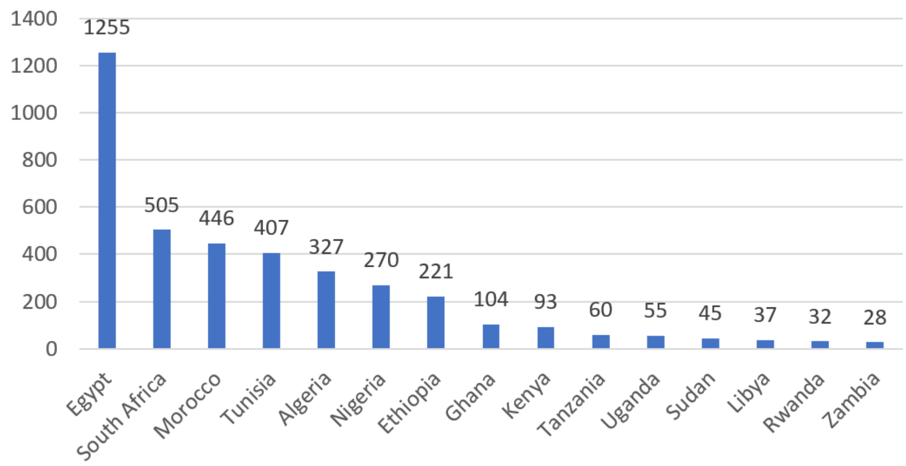

Figure 5: Top publishing African countries of AML4H research.

research productivity of the fifteen African nations, we found that only *Rwanda* has both a nominal GDP (12.098 billion USD) and a GDP per capita (912 USD) that are not ranked among the best twenty in the continent (United Nations Economic Commission for Africa, 2021). This country has succeeded to emerge thanks to the government-led research policy that tries to grow its local Artificial Intelligence community from the perspective of research and development (Vernon, 2019). This involves the hosting of international branch campuses in the country such as Carnegie Mellon University Africa, the creation of a local artificial intelligence ecosystem involving startups and corporation branches, and the development of capacity-building programs in ML (Vernon, 2019). However, we found that eight of the fifteen considered countries have a GDP per capita that is not ranked as one of the best twenty in Africa: *Kenya* (2255 USD), *Nigeria* (2326 USD), *Zambia* (1348 USD), *Tanzania* (1245 USD), *Uganda* (1105 USD), *Ethiopia* (1097 USD), *Sudan* (916 USD), and *Rwanda* (912 USD). This proves that the funding programs provided for African countries, the capacity building events like *Deep Learning Indaba*, and the free online and offline courses and mentorships in ML and biomedical informatics have succeeded to

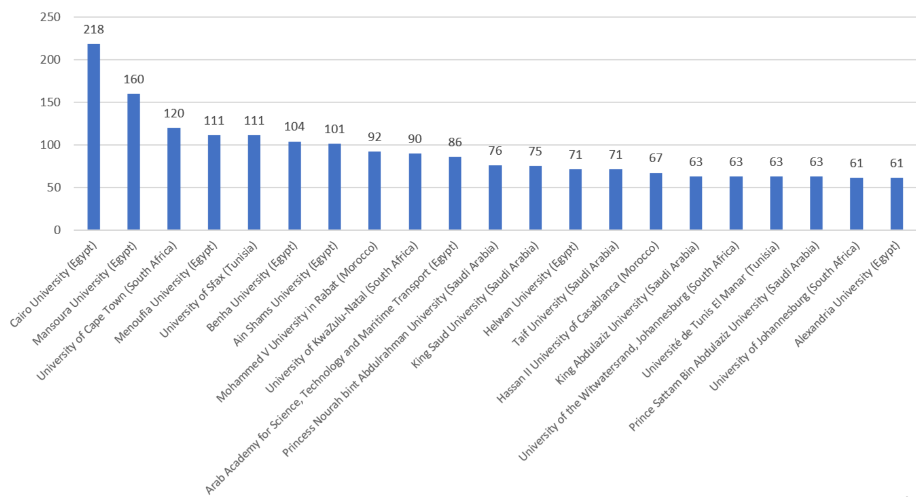

Figure 6: Top publishing African institutions of AML4H research

bridge the gap between African countries caused by financial burdens (Vernon, 2019). When seeing the number of publications per one billion USD of GDP, we found that only four countries achieved a rate of 2 or more: *Tunisia* (8.79), *Morocco* (3.12), *Egypt* (2.67), and *Rwanda* (2.64). This proves that the higher productivity of these countries is not motivated by their better financial situation. As explained before for Rwanda, this advantage is mainly due to better research policies and capacities in these countries, particularly related to the field of artificial intelligence and digital health (Ibeneme et al., 2021). Although several African countries have a better contribution to research about ML and healthcare, their productivity and impact are below the international average (Tran et al., 2019; Li et al., 2020) and they also require foreign funding and capacity-building programs from developed countries to evolve.

When identifying the twenty-one most productive institutions in this field, we found that only four African countries are featured as shown in Figure 6: *Egypt* (8 universities), *South Africa* (4 universities), *Tunisia* (2 universities), and *Morocco* (2 universities). This goes in line with our findings regarding the superiority of these nations in African ML research for healthcare. These

 doi:10.20944/preprints202302.0010.v2

universities are led by *Cairo University* (Egypt, 218 publications), *Mansoura University* (Egypt, 160 publications), *University of Cape Town* (South Africa, 120 publications), *Menoufia University* (Egypt, 111 publications), *University of Sfax* (Tunisia, 111 publications), *Benha University* (Egypt, 104 publications), *Ain Shams University* (Egypt, 101 publications), and *Mohammed V University in Rabat* (Morocco, 92 publications). This confirms previous findings proving the status of these universities as the most productive ones in Africa (Toivanen & Ponomariov, 2011). This is achieved thanks to a strong government-led collaboration network between Tunisia, Algeria, and Morocco and another one between Egypt and South Africa (Toivanen & Ponomariov, 2011). This is also due to robust collaborations between these African universities and non-African ones, particularly research collaborations of Tunisian, Algerian, and Moroccan universities with French and German ones, the collaborations between Egyptian and Saudi universities, and the research partnership between South African universities and the ones in the United Kingdom, Sweden, Netherlands, and the United States of America (Toivanen & Ponomariov, 2011). The analysis also revealed the significant contribution of five Saudi universities to ML research for healthcare in Africa. This means that the effect of the Saudi-Egyptian scholarly collaborations on the research productivity of Egyptian universities is higher than the one for the alliance between Tunisian, Algerian, Moroccan, and South African universities and their collaborators in Europe and North America (Toivanen & Ponomariov, 2011). Further efforts should be done to enhance the quality of research collaborations between African universities and world-class ones like the University of Oxford through initiatives like the *Africa Oxford Initiative*[11]. In another context, it is important to notice that there is no AI corporation or startup has been featured among the most published African institutions about ML and healthcare. This can be explained by the lack of development of the AI industry in Africa which is still in its very beginning (Gwagwa et al., 2020). This situation is different from the one of the AI

[11] https://www.afox.ox.ac.uk/

 doi:10.20944/preprints202302.0010.v2

industry in the developed world, mainly the United States of America, where AI corporations such as Google and Microsoft significantly contribute to the development of this research field where industries contribute up to 5% of the biomedical informatics research production (Jia et al., 2018).

The analysis of the 22 prolific authors of African biomedical ML scholarly outputs revealed that 12 of the scientists were Egyptian ones, 2 were from Morocco, and 2 were from South Africa as shown in Table 1. These scientists are led by *Aboul Ella Hassanien* (52 publications - Cairo University, Egypt), *Romany F. Mansour* (27 publications - New Valley University, Egypt), *Shaker El Sappagh* (26 publications - Benha University, Egypt), and *Fahmi Khalifa* (23 publications - Mansoura University, Egypt). This proves that the better standings of Egypt in biomedical ML research are mainly due to the existence of a large community of highly productive scientists in the field. These scientists contribute to the development of research customs and collaborations and the shaping of effective research directions inside their institutions leading to a sharp increase in their productivity (Kwiek, 2015). Two scientists in our list have even been featured among the most productive scientists in the world by coauthoring over 40 papers in one year several times between 2000 and 2016: *Aboul Ella Hassanien* (Cairo University, Egypt) and *Dan J. Stein* (University of Cape Town, South Africa) (Ioannidis et al., 2018). As well, these individuals will be the engine of highly-cited publications by guiding the rest of their colleagues in the same institution and by identifying trendy topics based on their large experience (Abramo et al., 2014). The lack of these scientists in other countries, particularly Morocco, South Africa, and Tunisia, proves that the development of research outputs in these countries about ML for healthcare is based on a collaborative effort rather than on individual ones. This situation does not seem to be similar to the one of African research about biomedical informatics where several highly-productive scientists are leading the field in South Africa and probably in Tunisia and Morocco through multiple international collaborations and large-scale research projects (Tapera & Singh, 2021). This is mainly explained by the fact that the field of ML for healthcare in these countries is not as mature as in Egypt and that several highly-productive scientists can appear in the next few years in Tunisia, Morocco, and South Africa where the field becomes more developed. Table 1 also revealed that several scientists working outside Africa are featured among the ones mostly publishing African research outputs on ML for healthcare: *Ayman El-Baz* (34 publications - University of Louisville, United States of America), *Mohammed Ghazal* (26 publications - Abu Dhabi University, United Arab Emirates), *Islem Rekik* (20 publications - Istanbul Technical University, Turkey), *Mohamed Elhoseny* (18 publications - University of Sharjah, United Arab Emirates), *Sanjay Misra* (18 publications - Østfold University College, Norway), and *Ahmed Soliman* (17 publications - University of Louisville, United States of America). Most of these researchers are Egyptian ones serving as liaisons between their home country and their host institution building and maintaining and growing research collaborations between their host country and their country of origin. This confirms a general trend of scientists working abroad to collaborate with their country of origin (Scellato et al., 2015). Such a collaboration is enhanced through the establishment of a joint supervision of Ph.D. students between the host country and the home nation of the productive scientist (Confraria et al., 2019). This type of collaboration has been enhanced during the COVID-19 pandemic thanks to the organization of online conferences allowing the easier development of collaborations to fight the disease outbreak and because of the fact that several scientists have been blocked in their home country due to travel restrictions (Korbel & Stegle, 2020). The two non-Egyptian scientists *Islem Rekik* (Tunisia) and *Sanjay Misra* (India) are involved thanks to their past research experience in Africa respectively with *African Network for Artificial Intelligence in Radiology and Imaging* (Morocco) and *Covenant University* (Nigeria).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 May 2023 doi:10.20944/preprints202302.0010.v2

Rank	Scientist	Affiliation	Publications
1	Aboul Ella Hassanien	Cairo University (Egypt)	52
2	Ayman El-Baz	University of Louisville (United States of America)	34
3	Romany F. Mansour	New Valley University (Egypt)	27
4	Shaker El-Sappagh	Benha University (Egypt)	26
4	Mohammed Ghazal	Abu Dhabi University (United Arab Emirates)	26
6	Fahmi Khalifa	Mansoura University (Egypt)	23
7	Omneya Attallah	Arab Academy for Science, Technology, and Maritime Transport (Egypt)	22
8	Ali Idri	Mohamed V University of Rabat (Morocco)	21
8	Serestina Viriri	University of KwaZulu-Natal (South Africa)	21
10	Islem Rekik	Istanbul Technical University (Turkey)	20
11	Mohammed Elmogy	Mansoura University (Egypt)	19
12	Mohamed Elhoseny	University of Sharjah (United Arab Emirates)	18
12	Sanjay Misra	Østfold University College (Norway)	18
14	Abdelmgeid A. Ali	Minia University (Egypt)	17
14	Essam H. Houssein	Minia University (Egypt)	17
14	Ahmed Soliman	University of Louisville (United States of America)	17
14	Dan J. Stein	University of Cape Town (South Africa)	17
18	Ahmad Taher Azar	Benha University (Egypt)	16
18	Bouchaib Cherradi	Hassan II University of Casablanca (Morocco)	16
18	Ashraf Darwish	Helwan University (Egypt)	16
18	Abdel-Badeeh M. Salem	Ain Shams University (Egypt)	16
22	Fathi E. Abd El-Samie	Menoufia University (Egypt)	15

Table 1: Most published authors of African research publications about ML for healthcare.

The analysis of the types of African research outputs revealed that these publications are mostly *articles* (2,223 publications), *conference papers* (1,158 publications), and *reviews* (221 publications) as shown in Figure 7. Although reviews are among the best three publication types, their number is small when compared to the ones of *articles* and *conference papers*. Developing more review papers is important to provide an overview of the state-of-the-art of ML for healthcare to the African research audience (Zhang et al., 2022). Limited interest is shown in short communications like notes, editorials, short surveys and letters to the editor as these kinds of publications have a less significant weight than *articles*, *reviews*, and *conference papers* in evidence-based research (Zhang et al., 2022). However, short communications and letters to the Editor can be very useful to brainstorm about a given topic and to develop the discussion, accuracy, and dissemination of valuable research findings (Turki et al., 2018; Joaquin & Tan, 2021). African scientists should use this route of publication to interact with other scientists across the continent and enhance the quality of their research outputs. Besides, African scientists do not significantly

 doi:10.20944/preprints202302.0010.v2

publish data papers to describe their datasets for ML for healthcare. Data papers are very important to provide detailed information about Africa-related datasets and ensure their availability for other scientists working on biomedical applications in the African context (Candela et al., 2015). The lack of documented and local datasets limits the development of customized solutions for digital health in the continent. More emphasis should be provided to the development of data papers in Africa. As for the number of *book chapters* (130 publications), it is higher than the one of short communications but less than the one of journal articles and conference papers. Books can involve literature reviews and articles and that is why book chapters are not more numerous than short communications (Yang et al., 2021). However, computer scientists tend to cite and interact with book chapters less than social scientists (Yang et al., 2021). This explains the tendency of African scientists to publish journal articles and conference papers rather than book chapters. One positive aspect of the African research production about ML for healthcare is that only two retractions occurred during 29 years. This proves in part the integrity of African research (Cokol et al., 2008). Nevertheless, this should be carefully considered as most of the research publications have been issued since 2020 and have not been examined by proficient scientists for a sufficient time to identify research flaws and scientific misconduct (Cokol et al., 2008).

When seeing the contribution of North Africa to every publication type, we found that most of the kinds of research publications are dominated by North African scientists (Grey in Figure 7). Particularly, *journal articles* and *conference papers* are mostly co-authored by North African institutions respectively at a rate of 61% and 75%. The significantly higher rate of conference papers by North African countries is mainly explained by the large cost of registration and travel to attend scholarly conferences that cannot be afforded by research scientists in Africa and the lack of organization of top-level conferences in the continent (Ondari-Okemwa, 2007). The situation could be worse if the COVID-19 pandemic did not occur allowing online participation in scholarly conferences (Wijesooriya et al., 2020). *Diversity, Equity, and Inclusion*

 doi:10.20944/preprints202302.0010.v2

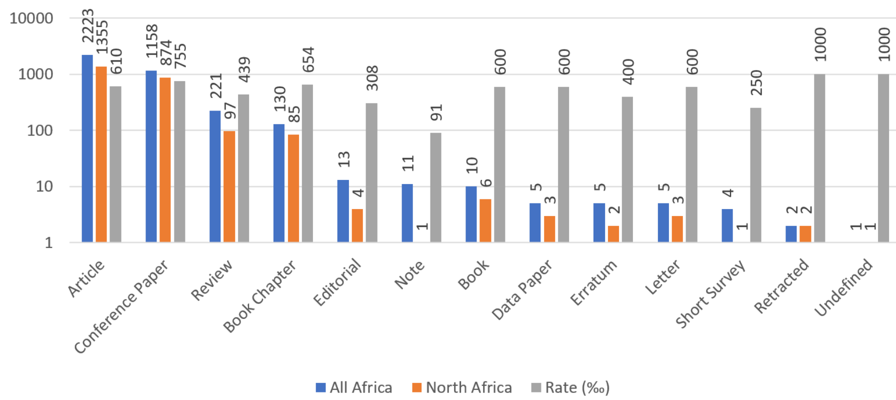

Figure 7: African research outputs about ML for healthcare (Log-Scale) by publication type.

(DEI) Programs are currently established to solve the lack of participation of the Global South, particularly Africa, in highly-referred conferences through fee waivers and mentorships (Roscoe, 2022). Several top-tier ML conferences will even be organized in Africa such as *ICLR 2023*[12] (Rwanda) and *MICCAI 2024*[13] (Morocco). It is interesting to see that *reviews*, *editorials*, *notes*, and *short surveys* are mostly published by Sub-Saharan African institutions. The higher rate of editorials by Sub-Saharan Africans is mainly due to the establishment of Africa-related special issues edited by Sub-Saharan African scientists in scholarly journals as a part of the DEI Program (Roscoe, 2022). Other reasons can be the involvement of African scientists in scientific society-driven special issues (Pugsley et al., 2020) or the organization of special issues in Sub-Saharan African journals about ML for healthcare (Kaswa et al., 2022). The higher interest of Sub-Saharan Africa to publish *reviews* and *short surveys* is explained by the fact that the Sub-Saharan African scientists are trying to explore the field of

[12]Eleventh International Conference on Learning Representations: `https://iclr.cc/Conferences/2023`

[13]27$^{\text{th}}$ International Conference on Medical Image Computing and Computer Assisted Intervention: `http://www.miccai.org/news/2020/12/31/and-the-location-of-miccai-2024-is`

 doi:10.20944/preprints202302.0010.v2

biomedical ML prior to contributing to it by identifying and understanding the state-of-the-art, recent advances, and limitations for this field (Palmatier et al., 2017). The significant publication of *Notes* by Sub-Saharan Africa is quite surprising as this region is not very interested in publishing short communications. This finding can be explained by the fact that notes can be short journal articles or unstructured reviews as in *BMC Research Notes* and that the development of notes can be easier for Sub-Saharan African scientists, particularly aspiring ones, than articles, conference papers and reviews (Krüger & Marshall, 2017).

When finding the publishers of the African outputs on ML for healthcare, we found that 2,783 out of 3,772 publications (76%) have been issued by ten world-class publishing houses as shown in Figure 8. These main publishers are led by *Springer* (Germany, 722 publications), *IEEE* (United States of America, 693 publications), *Elsevier* (Netherlands, 486 publications), and *MDPI* (Switzerland, 273 publications). This confirms the oligopoly of scholarly publishers where less than ten corporations particularly *Reed-Elsevier*, *Wiley-Blackwell*, and *Springer*, dominate the market (Larivière et al., 2015). An interesting finding was that *Taylor & Francis*, one of the top five research publishers, was not featured among the best publishing corporations issuing African papers on ML and healthcare. This is explicated by the lack of creation of scholarly journals related to Biomedical Informatics by *Taylor & Francis* [14]. Among the ten best publishers, only *Hindawi* is featured as an African publisher with 233 publications as it was founded in Cairo, Egypt before being moved to London, United Kingdom. Africa should increase its number of confirmed scholarly publishers to enhance the inclusion of its research outputs, particularly the ones related to biomedical ML, in English and local languages in large-scale bibliographic databases (Collazo-Reyes, 2013). The rise of *MDPI* and *BioMed Central* as con-

[14] Only 2 out of 104 Scopus-indexed health informatics journals are published by *Taylor & Francis*: Health systems and reform and Health Systems. Even these two journals are not fully related to the field and deal with it as a part of Public, Environmental and Occupational Health (SCImago, 2022).

 doi:10.20944/preprints202302.0010.v2

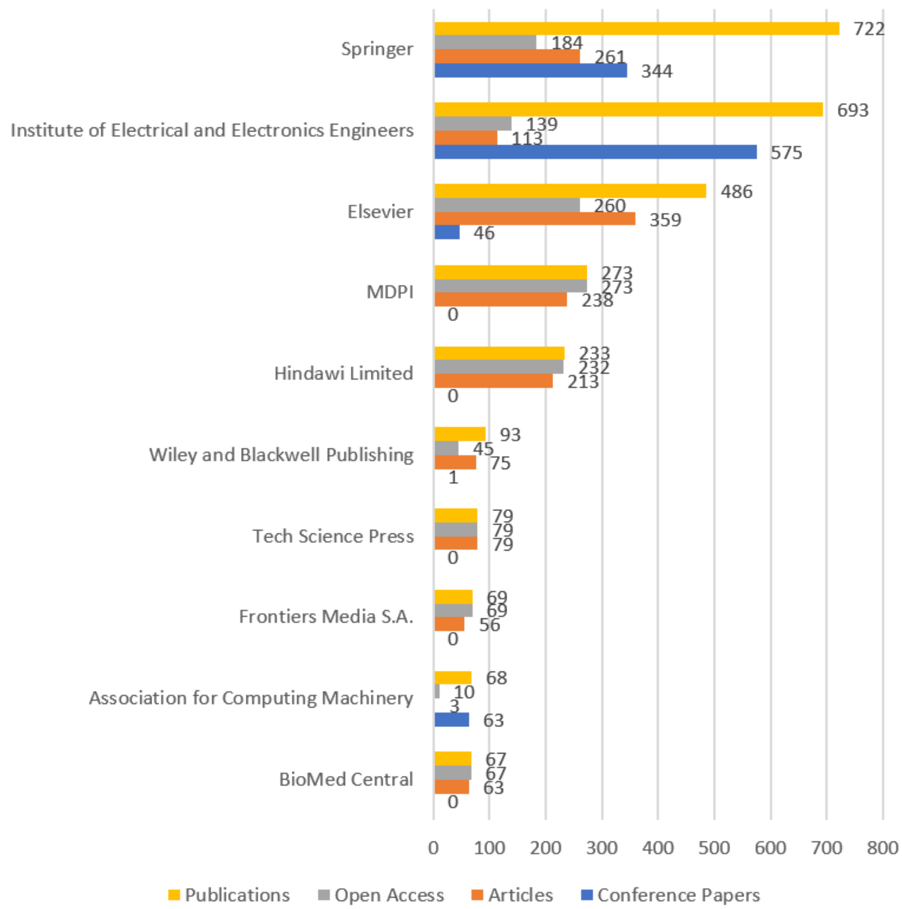

Figure 8: Main publishers of African research outputs about ML for healthcare.

firmed publishers of African biomedical ML research outputs is mainly caused by their development of biomedical mega-journals having a publishing model that reduces publication delays through easier and more flexible peer review (Schwarz Rodrigues et al., 2020).

When seeing the main publication types by publishers, we found three different situations depending on the policy of each corporation (Orange and blue in Figure 8). The two main publishers *Springer* and *IEEE* publish both conference papers and journal articles. This variety of publication types explains the domi-

 doi:10.20944/preprints202302.0010.v2

nation of these two publishers in the field of ML for healthcare. *BioMed Central*, *MDPI*, *Frontiers Media S.A.*, *Hindawi Limited*, *Wiley* and *Blackwell Publishing* exclusively publishes journal articles while *Elsevier* mostly issue journal articles with a few conference papers. This can be explained by the policy of *BioMed Central*, *MDPI*, *Frontiers Media S.A.*, and *Hindawi Limited* that mainly interests in the curation of open-access mega-journals (Schwarz Rodrigues et al., 2020). As for *Wiley*, *Blackwell Publishing*, and *Elsevier*, they are mostly known for maintaining subscription-based and open-access journals although they have several venues for conference papers such as *Procedia Computer Science*. *ACM* is the unique publisher that only issues African conference papers about ML for healthcare. This is quite surprising as there are many computer science journals published by the *ACM* and that can include African research about ML for healthcare such as *Communications of the ACM* and *ACM Transactions on Database Systems* (SCImago, 2022). The choice of the publication type by the African scientists for every publisher is mainly related to what every corporation provides as topics and publishing models for research journals and conferences. As shown in Table 2, *Springer* provides specific venues for the proceedings of health informatics conferences (*Lecture Notes in Bioinformatics*) as well as the general venue for computer science conferences (*Lecture Notes in Computer Science*). These series include 113 out of the 344 conference papers issued by *Springer*. The other part of the conference papers is published in other specific venues about sub-fields of computer science such as *Advances in Intelligent Systems and Computing*, *Lecture Notes in Networks and Systems*, and *Communications In Computer And Information Science*. Similarly, *ACM* hosts a venue for the proceedings of computer science conferences (*ACM International Conference Proceeding Series*) including 56 out of the 63 conference papers issued by this scholarly publisher. Even for *Elsevier*, most of the conference papers (35 out of 46 publications) are issued as a part of their venue for computer science proceedings (*Procedia Computer Science*). By contrast, *IEEE* does not provide a series for conference proceedings and publishes the outputs of every conference as an independent book. The importance of proceeding series in indexing

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 May 2023 doi:10.20944/preprints202302.0010.v2

conference papers has been confirmed for the computer science field (Fiala & Tutoky, 2017), probably because these venues allow easier indexing of scholarly conferences by bibliographic databases. The advantage of the main conference paper publishers over *Elsevier* is mainly due to the involvement of *ACM* and *IEEE* in the regular organization of scholarly conferences and the broad scope of *Springer* conference series that accepts to include conferences with a narrow regional representation and research topic (Wainer et al., 2009). African community should study how every publisher considers accepting conference proceedings and consequently work to enhance the indexing of continent-level conferences in biomedical informatics, particularly in Sub-Saharan Africa.

As for the most published journals, we found that most of them are open-access mega-journals publishing research papers after flexible peer review and short editorial delay, such as *IEEE Access* (91 publications, IEEE), *Applied Sciences* (40 publications, MDPI), *BioMed Research International* (28 publications, BioMed Central), and *Scientific Reports* (28 publications, Nature Publishing Group). Several open-access journals with narrower scope but having the same editorial policy of quick manuscript processing and sometimes a higher acceptance rate and efficient editorial services like proofreading and typesetting are also identified among the main target journals for African research about biomedical ML, particularly *Computational Intelligence and Neuroscience* (68 publications), Hindawi), *International Journal of Advanced Computer Science and Applications* (46 publications, Science and Information Organization), *Journal of Healthcare Engineering* (40 publications, Hindawi), *Sensors* (34 publications, MDPI), *Electronics* (27 publications, MDPI), and *Informatics in Medicine Unlocked* (26 publications, Elsevier). The leading position of *Elsevier* and *Springer* in publishing journal articles about African biomedical ML is mainly related to the creation of scholarly journals specific to a particular topic of biomedical informatics like *Computers in Biology and Medicine* (31 publications, Elsevier), *Biomedical Signal Processing and Control* (30 publications, Elsevier), and *Neural Computing and Applications* (30 publications, Springer). When seeing the representation of Sub-Saharan African institutions in top journals, it was clear that only five journals were not dominated by North Africa: *Computational Intelligence and Neuroscience* (Hindawi, 32%), *BioMed Research International* (BioMed Central[15], 14%), *Scientific Reports* (Nature Publishing Group, 35%), *Journal of Healthcare Engineering* (Hindawi, 32%), and *Informatics in Medicine Unlocked* (Elsevier, 46%). This advantage for *Hindawi* can be explained by their lack of manuscript formatting requirements as well as their free-of-charge language proofreading report at point of submission[16]. As well, *Hindawi* also provides a fee waiver for publication in its scholarly journals on a rotating basis (Morrison et al., 2015). Such editorial services are friendly to early-career Sub-Saharan African scientists who lack research support (Mullan et al., 2011). Concerning *Scientific Reports* and *Informatics in Medicine Unlocked*, the only reason for their choice by Sub-Saharan Africa as target scholarly journals is due to their fast editorial delay and that is why they are less considered than the journals maintained by *Hindawi*. The list of the main journals publishing African outputs about biomedical ML also reveals that 6 out of the 13 top journals (46%) are not related to computer science and that 2 journals (15%) are multidisciplinary mega-journals publishing health-related research among other outputs about topics ranging from exact sciences to social sciences. This lack of consideration of health-related research venues is confirmed when seeing the number of considered publications indexed by PubMed, a bibliographic database for biomedical scholarly publications. We found that only 948 out of 3,772 publications (25.1%) are indexed by PubMed. This proves a lack of involvement of health specialists, particularly physicians, pharmacists, and dentists, in biomedical informatics and digital health research in Africa and that the field is largely dominated by computer scientists. This is linked to the lack of availability of biomedical informatics courses in medical schools (Shortliffe, 2010). Further efforts should be done to include medical spe-

[15] Also co-published by *Hindawi*.

[16] Further information are available at `https://www.hindawi.com/publish-research/authors/`.

cialists in biomedical ML research. This situation has been better during the first wave of the COVID-19 pandemic in 2020 where 147 out of 679 publications (21.6%) were indexed by Scopus. This is mainly due to the timely awareness of the clinical community that health information systems can be important to monitor the evolution of the disease due to the emergency of the situation (Bakken, 2020).

The analysis of the number of open-access publications by the publisher (Grey in Figure 8) finds that the publishers of open-access journals providing rapid and flexible peer review and low-cost editorial services (*MDPI* and *Hindawi*) lead the open-access publishing industry in issuing African research about biomedical ML. Several publishers such as *Tech Science Press*, *Frontiers Media S.A.*, *BioMed Central* also fully publish their outputs as open-access publications. However, they are not aimed at the African community as *MDPI* and *Hindawi* due to their higher publication fees. Surprisingly, *Springer*, *IEEE* and *Elsevier* that are not specialized in open-access publishing publish a significant rate of African research as open-access outputs. This is explained for *Springer* and *Elsevier* by the availability of full open-access fee waivers for low-income countries, including Sub-Saharan African ones, through funding programs like *Research4Life* and *cOAlition S* (Simard et al., 2022). The open-access publications of *IEEE* are mainly related to its open-access mega-journal, *IEEE Access*, providing flexible peer review and editorial policies (Raman et al., 2022). When seeing the rate and types of open-access licenses assigned to publications, we found that 1,954 out of 3,772 publications (51.8%) are open-access. This is a direct result of the tendency of Sub-Saharan African institutions to publish their research outputs as open-access scholarly publications (Simard et al., 2022) and of the free sharing of COVID-19-related publications by publishers during the first wave of the pandemic in 2020 to increase the efficiency of the scholarly response to the disease outbreak (Lee & Haupt, 2020). These outputs are mostly gold (1,336 out of 1,954 publications) or green open access (1,297 out of 1,954 publications) as shown in Figure 9. Most of these publications are under gold and green open-access licenses at once. This tendency to favor gold

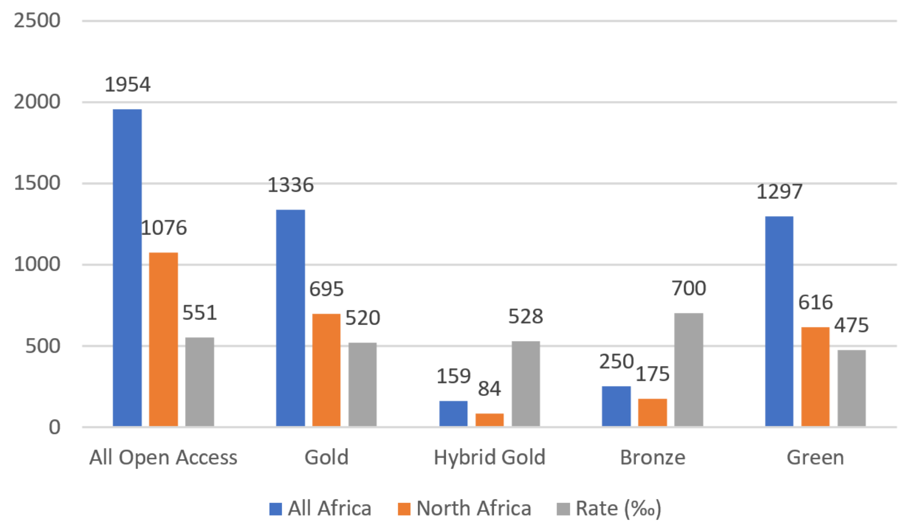

Figure 9: Open-Access African research publications about ML for healthcare by OA type.

and green open access is higher for Sub-Saharan Africa (Grey in Figure 9) and is mainly explained by the freedom given to authors and institutions to upload their research papers before and after final publications to repositories and freely shared them with the scientific community by contrast to bronze open access that prohibits the reuse of the publications by the authors (Singh et al., 2020). Figure 9 also shows a limited tendency of the authors of papers in pay-walled journals to pay article processing fees to let their research output open-access as only 159 open-access publications are issued in hybrid journals. This is due to the high open-access fees that authors have to pay for being granted open access (≊3,000 USD) (Björk, 2012) and the possibility of doing self-archiving for free (Laakso, 2013). Further efforts should be done to enhance open-access publishing in Africa by providing sustainable funding resources and spreading open science practices.

The screening of the subject areas of the African research outputs about ML for healthcare as inferred from their source titles has revealed that most of the

 doi:10.20944/preprints202302.0010.v2

work is published in computer science-related, engineering-related, medicine-related, and mathematics-related research venues are found in Figure 10. Another evident finding is the lack of publication of biomedical ML outputs in venues dealing with unrelated fields such as *Business, Management, and Accounting*, *Earth and Planetary Sciences*, *Economics, Econometrics, and Finance*, and *Arts and Humanities*. However, what should be considered is the limited publishing about biomedical ML applications in *Pharmacology, Toxicology, and Pharmaceutics*, *Nursing*, *Psychology*, *Veterinary Medicine*, and *Dentistry*. Although artificial intelligence has been used for many years to support clinical medicine practices (Bernstam et al., 2010), the application of computer science to other health professions is quite a new field, particularly in Veterinary Medicine (Smithakin et al., 2007), Nursing (Booth et al., 2021), Pharmacology (Zhao et al., 2005), and Dentistry (London et al., 2022). Psychology is developing computational methods for remote diagnosis and treatment based on the principles of human-computer interaction (Dix, 2017) and social network analysis (Kim et al., 2021). An interesting fact is that these emerging fields are mostly dominated by Sub-Saharan Africa. This seems to be an excellent alternative for Sub-Saharan Africa to grow its biomedical ML research in a field where competition is limited and the need for it is growing worldwide. Besides, Sub-Saharan Africa has proven its efficiency in conducting biomedical ML research related to *Agricultural and Biological Sciences* and *Immunology and Microbiology*. These outputs respectively reflect research about *food chemistry and safety* as well as *infectious diseases, immunology, microbiology*. The advantage of research on ML applications for infectious disease monitoring is confirmed by previous publications on the matter (Phoobane et al., 2022) and is mainly related to the development of research consortia and programs in Sub-Saharan Africa for the control of epidemic infectious such as the *West Africa International Centers of Excellence in Malaria Research* (Shaffer et al., 2019). ML applications in Agriculture is a limited research area in Africa mainly driven by the individual efforts of several scientists in South Africa (Benos et al., 2021), explaining the relative domination of Sub-Saharan Africa on health safety research linked to agricul-

 doi:10.20944/preprints202302.0010.v2

ture. Furthermore, there are several research topics of ML for healthcare where North Africa and Sub-Saharan Africa contribute comparably, mainly *Public, Environmental and Occupational Health* (Social Sciences, Health Professions, Environmental Science), *Bioinformatics* (Biochemistry, Genetics, and Molecular Biology), and *Neurology and Neurosurgery* (Neuroscience). The similar distribution of ML research related to public, environmental, and occupational health between North Africa and Sub-Saharan Africa does not seem to correlate with the contributions of African countries in public health research where Sub-Saharan Africa currently dominates the research field (Chuang et al., 2011). The equal distribution in public health-related ML research can be explained by the existence of research topics in this scholarly area that is only too relevant for one of the two regions such as *traffic accidents* and *pollution-related diseases* in North Africa (dos Santos et al., 2019). Bioinformatics research is developed and equally distributed across Africa thanks to the establishment and ongoing efforts of *H3ABioNet* as a continent-level consortium for computational biology (Mulder et al., 2015). Neuroscience-related ML research is fairly split between North Africa and Sub-Saharan Africa thanks to the existence of highly productive research communities on both sides of the African continent, particularly in Tunisia, Morocco, Egypt, Nigeria, and South Africa (Maina et al., 2021). By contrast, several fields are dominated by North Africa. These fields explain the advantage of North Africa in terms of scholarly productivity related to biomedical ML research:

- *Decision Sciences* stands for ML research about clinical decision support and recommendation system engineering.
- *Chemical Engineering*, *Chemistry*, and *Materials Science* mainly reveal research outputs linked to biochemistry, nanomedicine, device engineering, and drug discovery.
- *Physics and Astronomy* and *Energy* identify research related to Biophysics, Nuclear Medicine, Oncology, and Radiology.

Several fields covered by these three clusters are still incubating and do not explain the huge gaps between North Africa and Sub-Saharan Africa such as Nanomedicine (Bragazzi, 2019), clinical decision support (Adunlin et al., 2014; Chien et al., 2022) and Drug Discovery (Winks et al., 2022). The publication bias towards North Africa is mainly explained by the higher research productivity of this region in fields like Oncology (Mutebi et al., 2022), Radiology (Dang et al., 2015), and Biochemistry (Sooryamoorthy, 2018).

The assessment of the ten most cited scholarly publications about ML for healthcare (Table 3) has confirmed the disciplinary distribution of the scholarly publications as three of the most cited papers deal with breast cancer diagnosis and classification using medical imaging and three other ones deal with COVID-19 diagnosis and profiling by applying machine-learning models on biological analyses and medical images. This evaluation has revealed as well that eight of these outputs have been issued by *Elsevier* as shown in Table 3: Three papers are published in *Expert Systems with Applications* and two in *Computer Methods and Programs in Biomedicine*. This proves that *Elsevier* publishes more impactful research outputs than the other corporations due to its highly selective editorial policy characterized by a limited acceptance rate (Haensly et al., 2008; Herbert, 2020). Other publishers should revise their peer review policies to achieve better research quality and consequently a better citation impact. A surprising finding is that only five out of the ten most cited publications are open access. This proves the limitation of open access as a factor for a single publication to achieve highly-cited paper status. In other words, there is a level of citation impact where the open-access advantage no longer works and where only publication quality matters. When seeing the venues of the ten highly-cited publications, we found out that only one of them was a conference paper. This finding also applies to the list of highly-cited papers about COVID-19 pandemic (De Felice & Polimeni, 2020) among other lists of highly-cited papers (Garousi & Fernandes, 2016). This is mainly explained by the relative inability of conferences to generate groundbreaking publications due to page limits and short time for reviewing manuscripts. One solution to that is the *ACL Rolling*

 doi:10.20944/preprints202302.0010.v2

Review launched to enable longer peer review period for conferences (Dabre, 2022). The list of the authors of the best-cited research publications revealed that *Mohamed Loey* (Benha University, Egypt) is featured as the first author of two highly-cited research publications. This author is not included in the list of most productive scientists at Table 1, proving that early-career scientists can write highly-cited publications if they meet high standards of research. When seeing the titles of the ten most cited papers, we find that the two first ranked scholarly publications are reviews, confirming the citation advantage of review papers over other publication types (Tahamtan et al., 2016). We also found one data paper published at *Symmetry* among our list, proving that biomedical dataset creation is an important work that can be worth citing. Finally, we found that six works deal with biomedical image classification, proving the emphasis of this research topic by the African community.

4.2. Time-Aware Analysis

Based on the assumptions raised in Section 4.1, we need to study the bibliographic data of the considered publications for four periods: *Until 2013*, *2014-2016*, *2017-2019*, and *2020-2022*. This will confirm the association of several developments in African biomedical machine learning research with the occurrence of events such as the ongoing COVID-19 pandemic. Additionally, this will identify the stable characteristics and behaviors of African research productivity in machine learning for healthcare.

Figure 11 shows the distribution of published articles in four - Gold, Bronze, Green, and Hybrid Gold open access types, over the observed categories of years until the time of collation of this study. Our analysis shows a similar number of gold and green accesses across the time period 2020–2022. Although publishers are leaning towards publishing their articles in the gold access, as seen over the time period 2020–2022, similarities are also observed in Bronze and hybrid gold accesses during the time periods 2014–2016 and 2017–2019. We observed from our collection that hybrid gold access has had the lowest percentage over the years and did not even get recognition by publishers before

 doi:10.20944/preprints202302.0010.v2

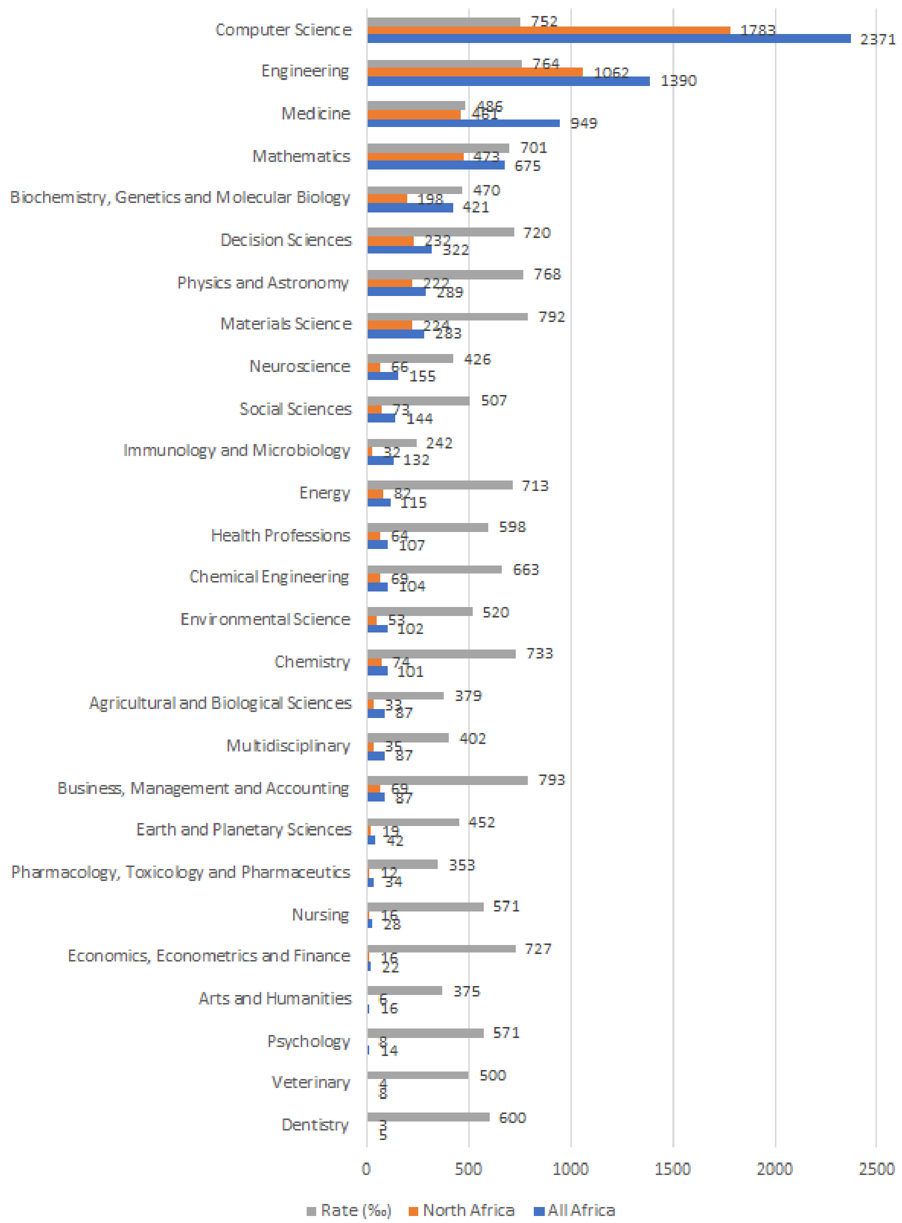

Figure 10: Subject areas for African research outputs about ML for healthcare.

Source Title	Publisher	All Africa	North Africa	Rate (‰)
Lecture Notes In Computer Science Including Subseries Lecture Notes In Artificial Intelligence And Lecture Notes In Bioinformatics	Springer	113	81	717
IEEE Access	IEEE	91	68	747
Advances In Intelligent Systems And Computing	Springer	85	76	894
Lecture Notes In Networks And Systems	Springer	73	59	808
Computational Intelligence And Neuroscience	Hindawi	68	22	324
Computers Materials And Continua	Tech Science Press	67	65	970
ACM International Conference Proceeding Series	ACM	56	42	750
International Journal Of Advanced Computer Science And Applications	Science and Information Organization	46	36	783
Applied Sciences Switzerland	MDPI	40	36	900
Journal Of Healthcare Engineering	Hindawi	40	13	325
Communications In Computer And Information Science	Springer	37	30	811
Procedia Computer Science	Elsevier	35	32	914
Sensors	MDPI	34	30	882
Computers In Biology And Medicine	Elsevier	31	20	645
Biomedical Signal Processing And Control	Elsevier	30	26	867
Neural Computing And Applications	Springer	30	28	933
Biomed Research International	BioMed Central	28	4	143
Scientific Reports	Nature Publishing Group	28	10	357
Electronics Switzerland	MDPI	27	18	667
Informatics In Medicine Unlocked	Elsevier	26	12	462

Table 2: Most published target venues for African research outputs about ML for healthcare.

First Author (Year)	Title	Source	Citations	Publisher	Open Access
Hao Y. (2021)	Integrated analysis of multimodal single-cell data	Cell	706	Elsevier	Yes
El-Dahshan E.A.-S. (2014)	Computer-aided diagnosis of human brain tumor through MRI: A survey and a new algorithm	Expert Systems with Applications	443	Elsevier	No
Nweke H.F. (2018)	Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges	Expert Systems with Applications	418	Elsevier	No
Abdel-Zaher A.M. (2016)	Breast cancer classification using deep belief networks	Expert Systems with Applications	318	Elsevier	No
Asri H. (2016)	Using ML Algorithms for Breast Cancer Risk Prediction and Diagnosis	Procedia Computer Science	307	Elsevier	Yes
Loey M. (2021)	A hybrid deep transfer learning model with ML methods for face mask detection in the era of the COVID-19 pandemic	Measurement: Journal of the International Measurement Confederation	291	Elsevier	Yes
Shrock E. (2020)	Viral epitope profiling of COVID-19 patients reveals cross-reactivity and correlates of severity	Science	264	American Association for the Advancement of Science	Yes
Loey M. (2020)	Within the lack of chest COVID-19 X-ray dataset: A novel detection model based on GAN and deep transfer learning	Symmetry	246	MDPI AG	Yes
Chougrad H. (2018)	Deep Convolutional Neural Networks for breast cancer screening	Computer Methods and Programs in Biomedicine	238	Elsevier	No
Inbarani H.H. (2014)	Supervised hybrid feature selection based on PSO and rough sets for medical diagnosis	Computer Methods and Programs in Biomedicine	229	Elsevier	No

Table 3: Most cited African scholarly publications about ML for healthcare.

 doi:10.20944/preprints202302.0010.v2

2014. However, we notice that a higher percentage of authors chose to publish with gold access during this same period. This might indicate that authors are becoming conscious of green open access, putting out versions of manuscripts during preparation, and most likely choosing to only have published articles out to readers as in gold access. However, as indicated in Figure 9 most of these publications are under gold and green open access licenses due to the policies governing (Simard et al., 2022; Lee & Haupt, 2020; Singh et al., 2020; Björk, 2012; Laakso, 2013) the different open access statuses indicated in Figure 4 on publishing patterns.

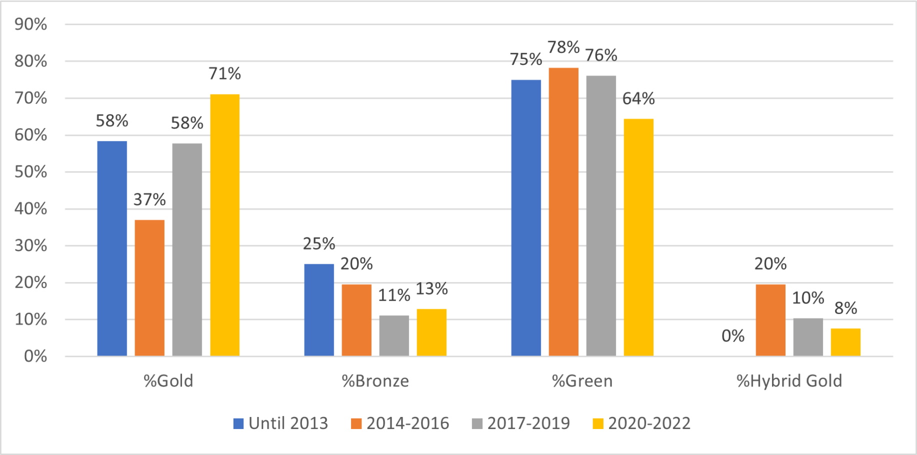

Figure 11: Percentage of different open-access status over the total number of open-access publications per time period

Figure 12 shows the top 11 researchers in Africa who have contributed to the research space "applications of machine learning for healthcare in Africa" over the same time frame. Nine of the top eleven authors had no contributions before 2014, either because they were new to machine learning or because they lacked the resources required to carry out their research. Also, it could be because machine learning research has been revived by the evolution of deep learning models, starting with AlexNet in 2012 (Krizhevsky et al., 2017) and the space has only recently started attracting funding from sponsors, as shown in Figure

 doi:10.20944/preprints202302.0010.v2

16. The last period, 2020-2022, saw very high contributions, especially from Romany F. Mansour who published all his work no later than 2017 (Figure 12). This indicates that the trend of machine learning research in Africa could have been triggered by the COVID-19 pandemic (De Felice & Polimeni, 2020).

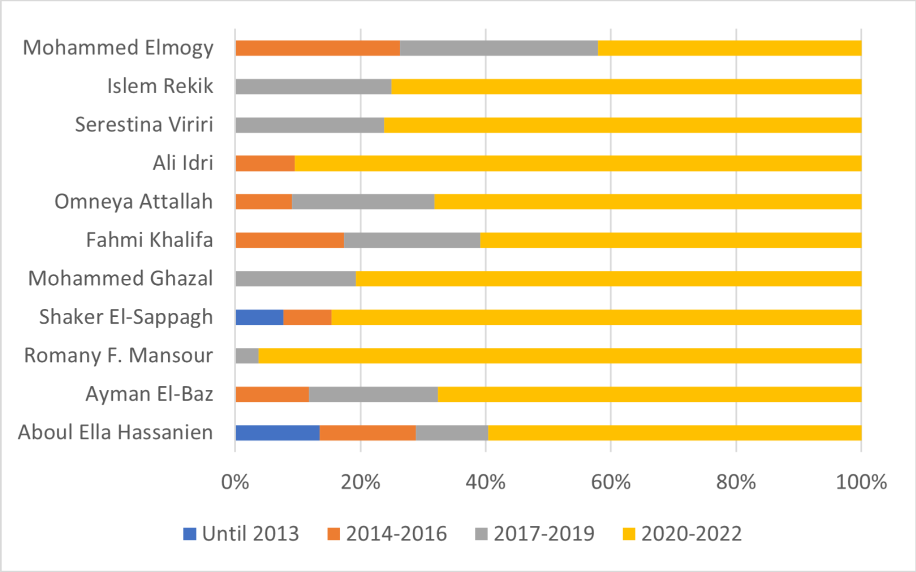

Figure 12: Top Contributing authors from African Institutions

Figure 13 shows the trend of biomedical machine learning in the different subjects in health research areas over time. The most common area of published research among these areas is computer science, which accounts for approximately 30% of the publications. Biochemistry, genetics and molecular biology, engineering, mathematics, and medicine have all found very significant use for machine learning models. Surprisingly, biochemistry, genetics, and molecular biology had significant machine-learning usage before 2013. However, usage has not increased much more than it did in the early years. Business, management, accounting, chemical engineering, energy, and environmental sciences did not have very significant use of machine learning in the early days and still seem to use it very minimally. As seen, African healthcare machine-learning research has primarily been concerned with clinical and medicinal research. Operational health service delivery is rarely studied. Operational healthcare focuses on improving the performance of the healthcare facility and servers. In a system where resources are scarce, it is obvious to think of increasing operational efficiency, which will reduce costs and increase patient satisfaction. Questions like how long patients must wait for care, how long they must stay in the hospital for consultations, and how to improve the quality of healthcare provided by different healthcare providers have never been taken into consideration. There is a demand, so it was decided to identify comparable worldwide indicators and analyze their traits as well as the larger national frameworks and circumstances.

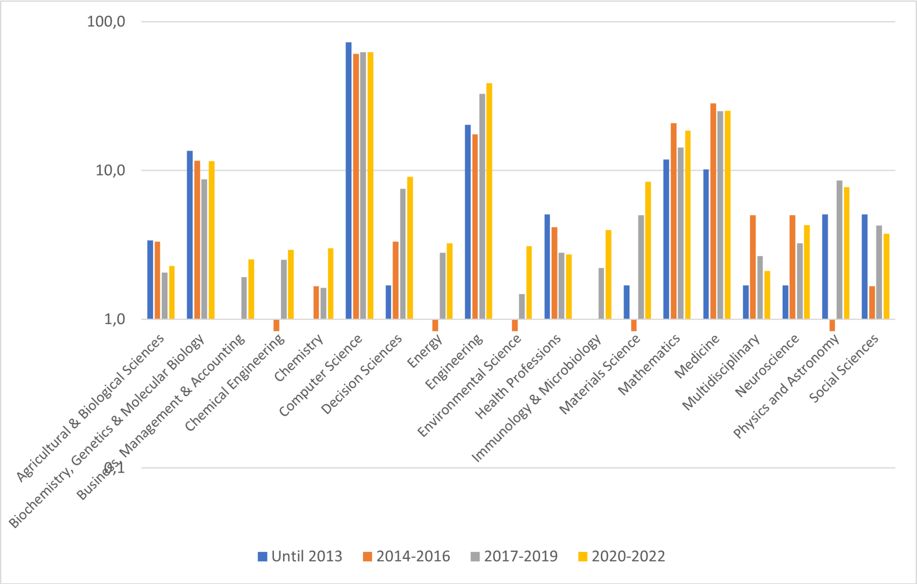

Figure 13: Most common subject areas for Africa biomedical machine learning research (Log-Scale) per period

Figure 14 shows the distribution of publication types over the periods. There has been a general increase in all document types over time. Most notably, all considered document types have had at least one publication in very recent

 doi:10.20944/preprints202302.0010.v2

times. The most common means of research communication are conference papers and articles, which exponentially increase over time. However, it is clear that the increase in journal article publication has been sharper between 2020 and 2022. This is mainly due to travel restrictions during the COVID-19 pandemic that blocked the organization of scholarly conferences (Valenti et al., 2021).

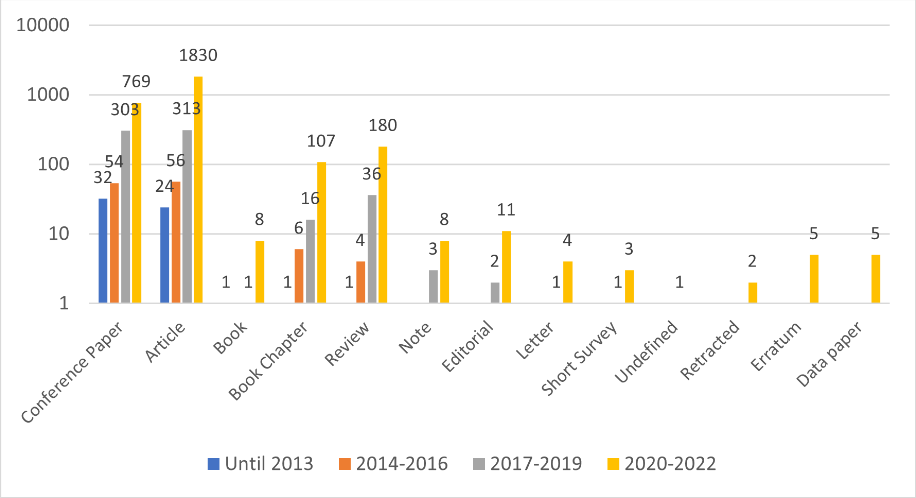

Figure 14: Number of publications (Log-Scale) by type per time period

Figure 15 shows that these top 10 universities contributed more in the period 2020–2022, as opposed to the periods 2014–2016 and 2017–2019. Cairo University seems to be the only university that has been contributing to this research space right from the first period until 2013. Figure 6 shows Cairo University is the most contributing university over these four periods based on the number of publications within these four periods. Figures 15 and 6 also show that except for the University of Cape Town and the University of Kwazulu-Natal, the remaining eight universities are from North Africa, including six in Egypt, one in Tunisia, and one in Morocco. This identifies huge inequities in access to machine learning and publication resources in sub-Saharan Africa and the need to address these inequities and democratize machine learning research.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 May 2023 doi:10.20944/preprints202302.0010.v2

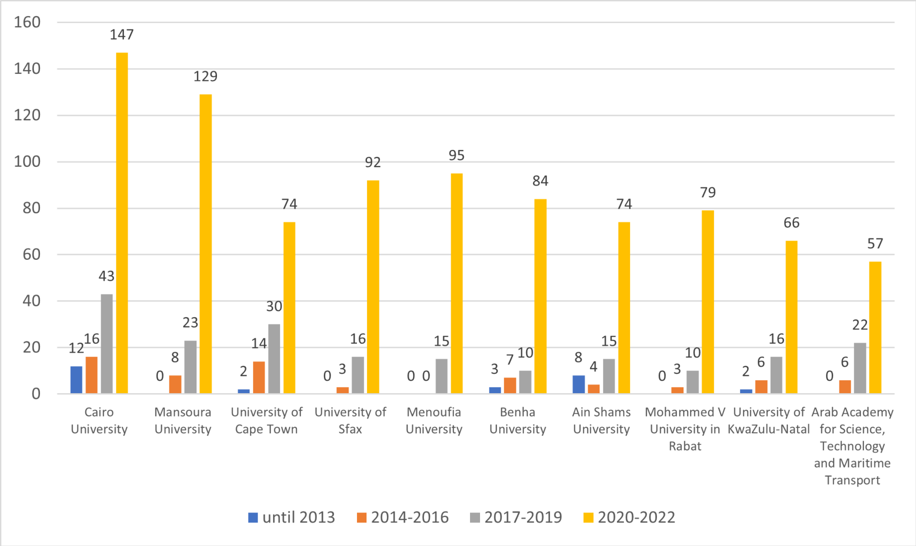

Figure 15: Research contribution from universities per period

Figure 16 summarizes funding for African biomedical machine learning research. The research space saw an exponential rise in funding from its usual funding institutions within 2020-2022. The National Institute of Health (NIH), from 2017 onward, accounted for approximately 75% of the funding across all observed periods. This might be due to the greater adoption of the Gold and Green Open Access publication systems throughout the 2017–2019 and 2020–2022 time periods, as observed in Figure 11. Because the Gold and Green Open Access systems allow for the unrestricted dissemination of research articles in journals and repositories, respectively, the NIH likely saw this trend as a good opportunity to promote biomedical research in Africa so that others can build on previous works to come up with better research works, so they went hard into funding to support such initiatives. We also assume that the NIH funded heavily during these periods because the following subject areas, Biochemistry, Genetics & Molecular Biology, Medicine, and Neuroscience, which doubled as some of the organization's key focus areas, saw high engagement and interest from both major African (Egypt, South Africa, and Morocco) and non-African countries

 doi:10.20944/preprints202302.0010.v2

(United States, India, United Kingdom, France, China, Germany, Canada, and Australia) Figure 18. The European Commission ranks last in these observations, likely due to the study's focus (understanding the current academic state of researchers in Africa developing and applying machine learning for healthcare) rather than implying that the European Commission does not support funding in the biomedical space. Perhaps they would be better represented if the study's scope was expanded to include places other than Africa.

The trend chart in Figure 17 displays the top 20 African nations actively engaged in the research and use of machine learning in healthcare. It highlights our results that Egypt is the nation with the biggest contribution to the research space. From this chart alone, we can see that Egypt has a greater spike than the other nations depicted here at all times. We also see that, in contrast to past periods with a decreasing tendency, the period 2017–2019 attempted to push upwards, but not dramatically. This might be because conference papers and article proceedings began to gain attraction during that time period, resulting in an increase in the acceptance rate of the Gold and Green Open Access systems, drawing financing from both African and non-African academic institutions, Figure 11. Cairo University in Egypt topped the research productivity on biomedical machine learning throughout the four periods with a total of 218 publications throughout all time periods. We also draw the conclusion that Egypt appears to be the forerunner of research works in Africa that focus on the study and application of machine learning in healthcare because Egypt claims five of the top ten universities that have made significant contributions to the research space (Cairo University, Mansoura University, Menoufia University, Benha University, and Ain Shams University) as shown at Figure 15. These institutions contributed a total of 693 of the 1191 publications produced by the top ten affiliations or universities represented, implying that Egypt is responsible for nearly 58% of research efforts and publications from 2013 to the present moment of this study.

African research publication outputs have increased over time, as shown in Figure 18. Six top-most published African countries of which four belong

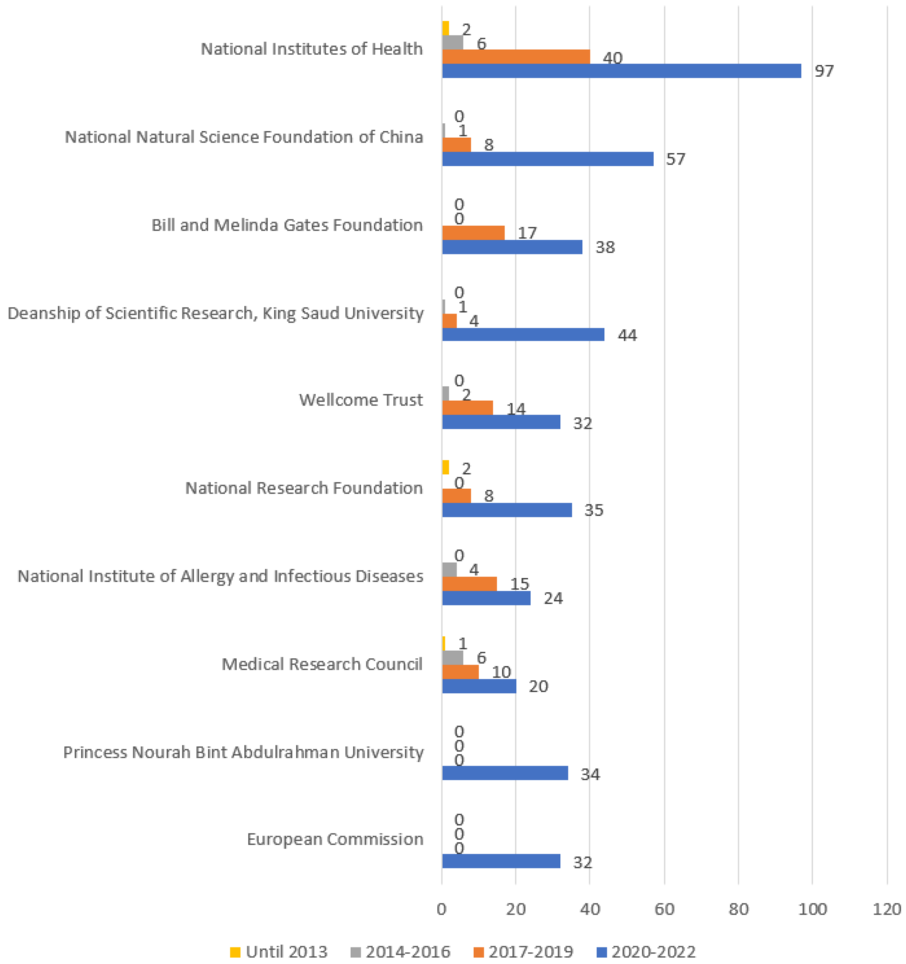

Figure 16: Main funding institutions for African research outputs in machine learning for health care per period

to North Africa (Egypt - 32%, Morocco - 12%, Tunisia - 11%, and Algeria - 8%), one to Southern Africa (South Africa - 13%) and one belongs to West Africa (Nigeria - 7%). This can be explained by several factors, as described in Section 4.1 on Publishing Patterns. Egypt is the highest-contributing nation to biomedical research and remains a heavily funded nation that participates actively in research for this space (Landini et al., 2015). Despite a few countries

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 May 2023 doi:10.20944/preprints202302.0010.v2

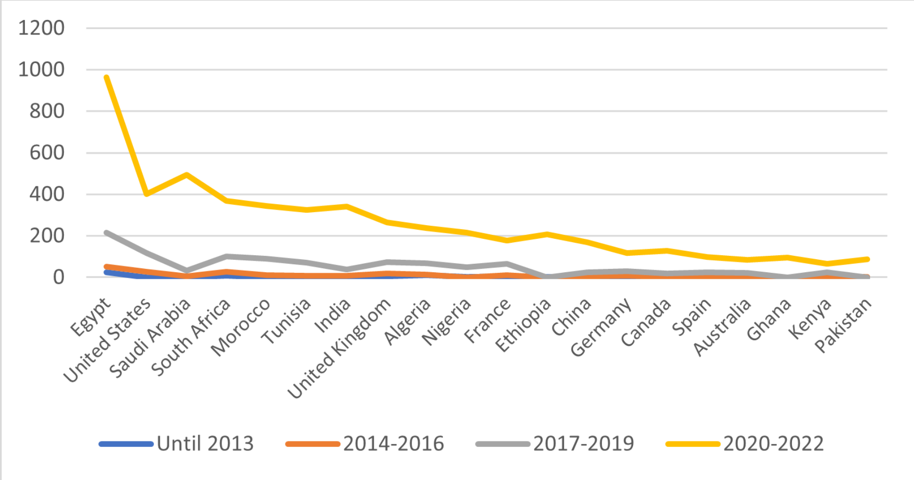

Figure 17: Trendline of research contributions by different countries per period

being the most published over the period, we see a gradual increase in the representation of other African countries from 2014. Figure 18 also shows an exponential increase in the number of publications in the period between 2020-2022, which accounts for 78% of the total number of publications.

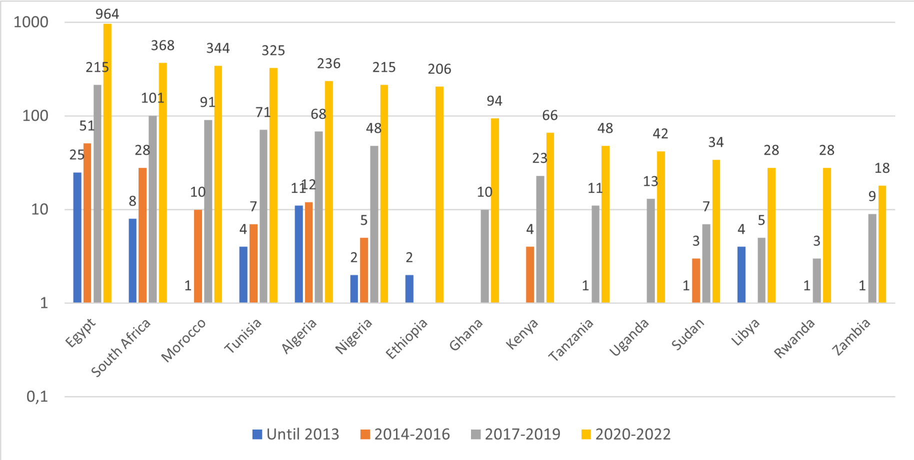

Figure 18: Most common African countries that published (Log-Scale) over the periods

African research has rapidly increased over the years, as shown in Figure 18 which has partially been promoted by international collaborations. This can be seen in Figure 19 which shows a general increase in the level of collaboration between African and non-African institutions over the different periods. The collaboration is shown to have started mainly in the period until 2013, with a gradual increase in the period of 2014–2016 and a steep increase in the period between 2017–2019 and 2020–2022, which account for 1%, 3%, 15%, and 81% of the total, respectively. The highest increase is shown in the period between 2020- 2022, publishing a total of 2643 (81%) publications, which may be due to different factors, which include increased funding shown in Figure 16 and these funding institutions are shown in Figure 4. Collaborations by non-African institutions can be attributed to increased funding (Zhou et al., 2020) (Hu et al., 2020), long-standing relationship with North Africa (Landini et al., 2015) for the case of Saudi Arabia, developed expertise in biomedical informatics (Tran et al., 2019) and machine learning (Li et al., 2020) and flexibility in establishing research collaborations (Sarwar & Hassan, 2015). The five top collaborations with African institutions are by non-African institutions from the United States of America (17%), Saudi Arabia (16%), India (12%), United Kingdom (11%), France (8%), and China (6%). The international collaboration shown by the research outputs can be further supported by Figure 4 which shows the main funding institutions for African research outputs.

4.3. Keyword Analysis

According to Figure 20, it is clear to see that keywords related to the category “STEM” are the most frequent (25000 keyword mentions) compared to those belonging to biomedicine (keyword mentions between 10000 and 15000). This proves that the African research publications on machine-learning applications in biomedicine are mainly emphasizing the computer science and engineering perspective of the considered matters rather than their clinical side. This is confirmed by previous analyses of the general trends of machine learning research that confirm the lack of deployment of machine learning applications into clini-

 doi:10.20944/preprints202302.0010.v2

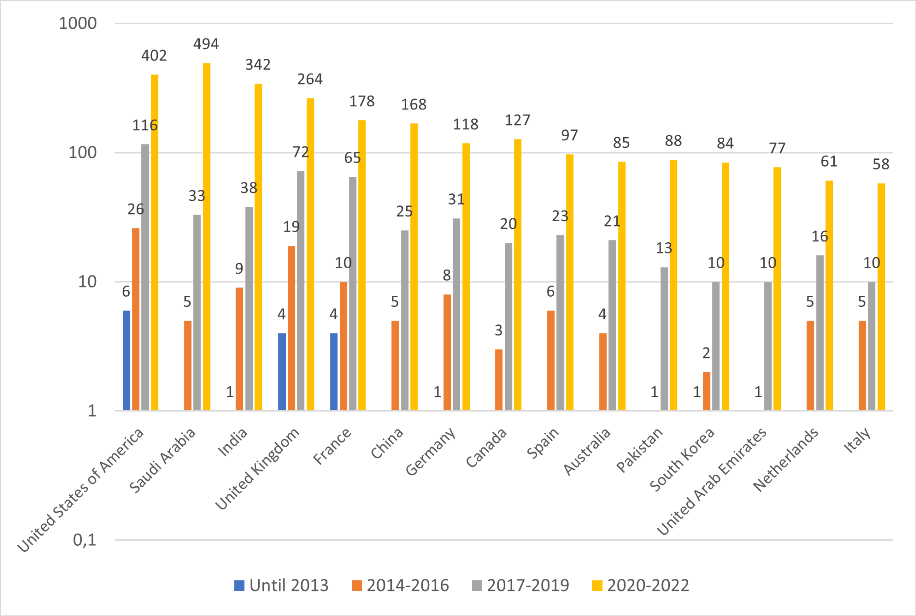

Figure 19: Major non-African countries contributing to African research (Log-Scale) over the periods

cal practice (Guo et al., 2020). Current research mainly focuses on developing machine-learning techniques to achieve better accuracy of biomedical artificial intelligence rather than transforming clinical practice for such a purpose (Li et al., 2021). We also notice according to the same figure that the rate of highly cited publications (cited more than 10 times) in our dataset is slightly higher for the social science and biomedicine-related keyword mentions. This proves that application-oriented research that tries to solve matters from the perspective of the problem itself can be more impactful than works that use generic methods to solve biomedical matters as shown in previous research publications (Tijssen & Winnink, 2016). Furthermore, we find that most keyword mentions are included in North African publications, except social science keywords. This goes in line with the higher tendency of Sub-Saharan African countries to work on the social aspects of medicine such as public health, epidemiology, and community health

(Asubiaro & Shaik, 2021). Another aspect in Figure 20 concerns publishing in Open Access where we notice that the majority of the keyword mentions are included in open-access publications, mainly the social sciences and biomedicine keywords. This is mainly related to the free publication of COVID-19-related papers under a free license (Lee & Haupt, 2020). These publications are mainly linked to the prediction of the epidemiological and social evolution of the disease (Fonkou et al., 2021; Phoobane et al., 2022). This is confirmed by the significantly high rate of keyword mentions in the scholarly publications issued after 2020 as shown in the same figure. The higher rate of keyword mentions in journal articles, mostly in the social sciences domain and biomedicine, is linked to the COVID-19 travel restrictions that blocked the organization of scholarly conferences between 2020 and 2022 (Valenti et al., 2021), i.e., during the period where African biomedical machine learning research has been the most active (Figure 14).

In Figure 21, we can more closely analyze the keyword mentions related to the category Science, Technology, Engineering, and Mathematics. In fact, according to the figure, most keywords belonging to this class concern machine learning and deep learning. This includes both models (e.g., *convolutional neural network*, *support vector machine*, *decision tree*, *random forest*, *transfer learning*, and *k-nearest neighbors algorithm*) and tasks (e.g., *classification*, *feature extraction*, *image segmentation*, *computer-aided diagnosis*, *data mining*, *prediction*, *automation*, and *image analysis*). Particularly, biomedical image processing tasks like *image analysis* and novel ones linked to the development of telemedicine in Africa such as *automation* have a better rate of mentions in highly-cited publications and journal articles. This confirms the distribution of computer science-related keywords in previous bibliometric studies on machine learning for healthcare where the list of keywords is mainly dominated by machine learning terminology (Guo et al., 2020; Li et al., 2021; Jimma, 2023). This also proves that most of the African biomedical machine-learning research outputs are mainly focused on applying classical machine-learning techniques and artificial neural networks for image and signal processing with a minor

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 May 2023 doi:10.20944/preprints202302.0010.v2

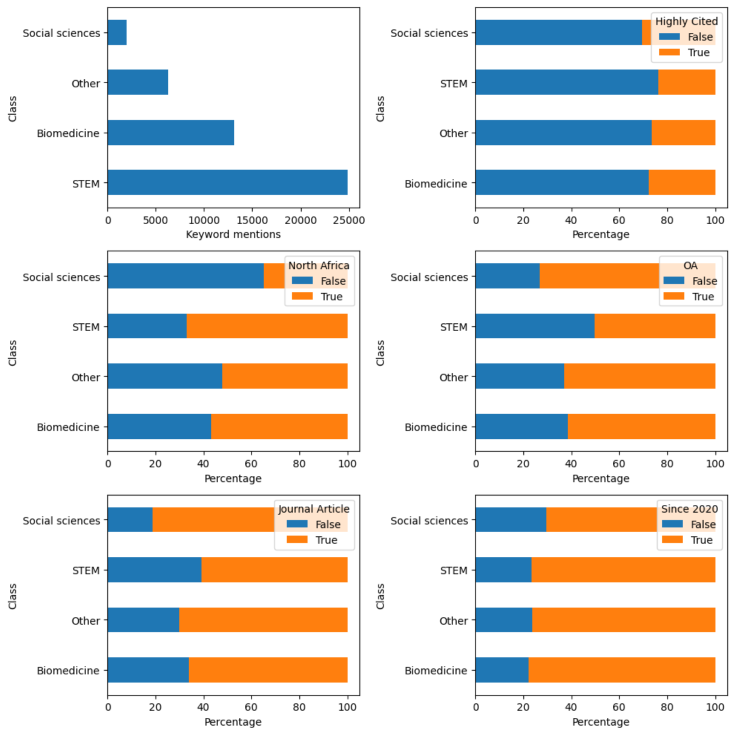

Figure 20: Keyword mentions per class: Overall number and rate of their availability in highly-cited publications, North African research production, open-access outputs, journal articles, and works published since 2020.

 doi:10.20944/preprints202302.0010.v2

interest in data mining and automation. This is unfortunately not common in Sub-Saharan African countries that mainly use classical machine learning algorithms like *random forest* for *disease prediction*. This also shows a lack of interest by the African research community in analyzing semantic resources such as biomedical ontologies, knowledge graphs, bibliographic databases, and electronic health records using large language models, natural language processing, and graph learning (Houssein et al., 2021). As well, African works on developing algorithms to ensure the trustworthiness of machine learning applications in healthcare (e.g., explainability, fairness, robustness, and privacy) are limited (Li et al., 2021; Kaur et al., 2022). Beyond this, common keywords also list multiple sources for biomedical machine learning, particularly *Big Data*, the *Internet of Things*, and *image* databases, as well as evaluation metrics, mainly *accuracy* and *receiver operating characteristic*. This confirms the general trends of using social network interactions to monitor the epidemiological surveillance of populations and of using Internet of Things devices for tracking the vital signs of patients, gathering health data using biomedical imaging and electrography, and driving healthcare robotics (Guo et al., 2020; Jimma, 2023). This research direction has been catalyzed by the COVID-19 pandemic that urged the urgent move towards telemedicine and clinical decision support for better management of the disease outbreak (Wijesooriya et al., 2020; Bayoudhi et al., 2022). This is confirmed by the higher rate of mentions of these keywords alongside keywords related to image processing and deep learning methods for image and signal processing (e.g., *convolutional neural network*) in scholarly publications since 2020 (Figure 21). Before this, only the *data mining* of biomedical resources such as electronic health records was significantly common in Africa. The two evaluation metrics, *accuracy* (mostly used) and *receiver operating characteristic*, have not been evocated in the context of the development of new measures for evaluating machine learning algorithms as in Turki et al. (2021). They are just the two main probabilistic metrics used by African scientists to evaluate their biomedical machine-learning applications. As shown in Figure 21, *accuracy* is mainly used as a proof of concept of the efficiency of a given algorithm, partic-

 doi:10.20944/preprints202302.0010.v2

ularly in scholarly conferences, while.*receiver operating characteristic* is mainly used in journal articles, open-access publications, and highly-cited papers where better evidence of the precision and recall of machine learning algorithms is required. Surprisingly, the rate of the *receiver operating characteristic* mentions in Sub-Saharan African publications is higher than the one of *accuracy* despite that the latter is simpler to compute.

Figure 22 focuses the analysis on the biomedical term category. The figure shows an interest in the applications of machine learning algorithms for *medical diagnosis* (More than 800 publications), automating *medical procedures* (Around 280), deciding *treatments* (Almost 150), reading *diagnostic tests* (Almost 150), identifying *risk factors* (Almost 120), recognizing disease *prognosis* (Almost 80), and detecting disease *biomarkers* (Almost 80). These tasks are mainly applied to lethal (e.g., *breast cancer* and *brain tumor*), widespread (e.g., *COVID-19*), and chronic (e.g., *Alzheimer's disease* and *heart disease*) diseases. The higher tendency to work on computer-aided diagnosis rather than treatment and prognosis is mainly due to the contributions of North African scientists and is a common pattern of machine learning research for healthcare, particularly for heart disease and breast cancer (Shailaja et al., 2018; Tran et al., 2019). However, works mainly dominated by North African scientists on the diagnosis of *heart disease*, *Alzheimer's disease*, *brain tumor* and *breast cancer* are not as cited as the rare and underdeveloped applications of machine learning for *pathology* and *medical procedures*. More efforts to increase the diversity of machine learning algorithms in clinical practice beyond computer-aided diagnosis should be provided for a better citation impact of African research outputs. Health data are collected for such a purpose using *pathology*, *epidemiology*, medical imaging (e.g., *magnetic-resonance imaging*, *computed tomography*, and *chest radiograph*), and electrography (e.g., *electroencephalography* and *electrocardiography*) methods. Research on machine learning from medical images (e.g., *chest radiograph* and *computed tomography*) has mainly evolved since 2020, due to the emergence of the *COVID-19* pandemic caused by *SARS-CoV-2* as a viral agent. This is confirmed by previous bibliometric studies on the use of machine

doi:10.20944/preprints202302.0010.v2

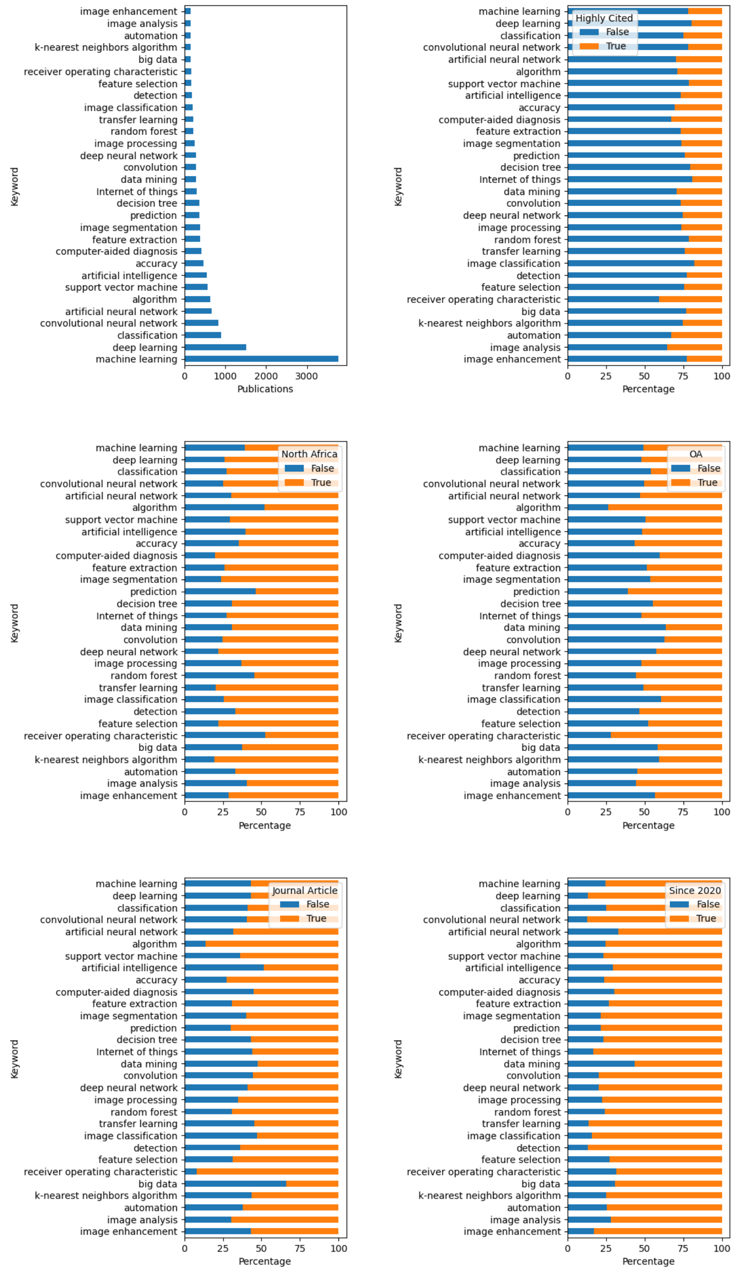

Figure 21: Keyword mentions for STEM terms: Overall number and rate of their availability in highly-cited publications, North African research production, open-access outputs, journal articles, and works published since 2020.

 doi:10.20944/preprints202302.0010.v2

learning for COVID-19 diagnosis (Alyasseri et al., 2021; Heidari et al., 2022) and is mainly linked to contributions by North African scientists (Alyasseri et al., 2021). These COVID-19-related works alongside several other works on COVID-19 epidemiology and diagnostic tests are more often issued as open-access journal articles due to travel restrictions caused by the disease outbreak that blocked the organization of scholarly conferences (Valenti et al., 2021) and the urgent need to access timely scholarly findings about the ongoing disease to manage the outbreak (Lee & Haupt, 2020).

Figures 23 and 24 respectively present the results related to the social science category and the unclassified term one. Several types of entities are mistakenly distributed among two categories due to the limitations of ChatGPT in classification tasks, particularly when a term can be included in more than one category (Hassani & Silva, 2023). For example, Autism Spectrum Disorder has been mistakenly considered a social science term (Figure 23) due to the fact that psychology is a part of medicine and social sciences at once. This term has been identified in almost 30 scholarly publications mostly done by North African local scientists. This topic is highly-cited and can be easily presented at scholarly conferences as it is trendy all over the world (Hyde et al., 2019). More interest should be provided to it by Sub-Saharan African scientists. As a result of the classification problem, we need to analyze both social science and unclassified term categories together. Keywords in these categories mainly include age ranges, particularly *adult* (around 400 keyword mentions), *aged* (around 110), *middle age* (around 100), *child* (around 90), *young adult* (around 80), *teenager* (around 70), *baby* (around 40), *pre-school student* (around 30), *pre-school age* (around 30), and *aged 80 and over* (around 20). *Age of a person* as a variable is also featured in around 70 scholarly publications. The distribution of age ranges proves the lack of customized machine-learning applications for extreme ages, although such works can be more susceptible to having a higher citation rate and being issued in journal articles as open-access publications (Figure 23). Unfortunately, the COVID-19 pandemic has not urged the development of customized machine-learning algorithms for extreme ages, mainly the elderly, although aged

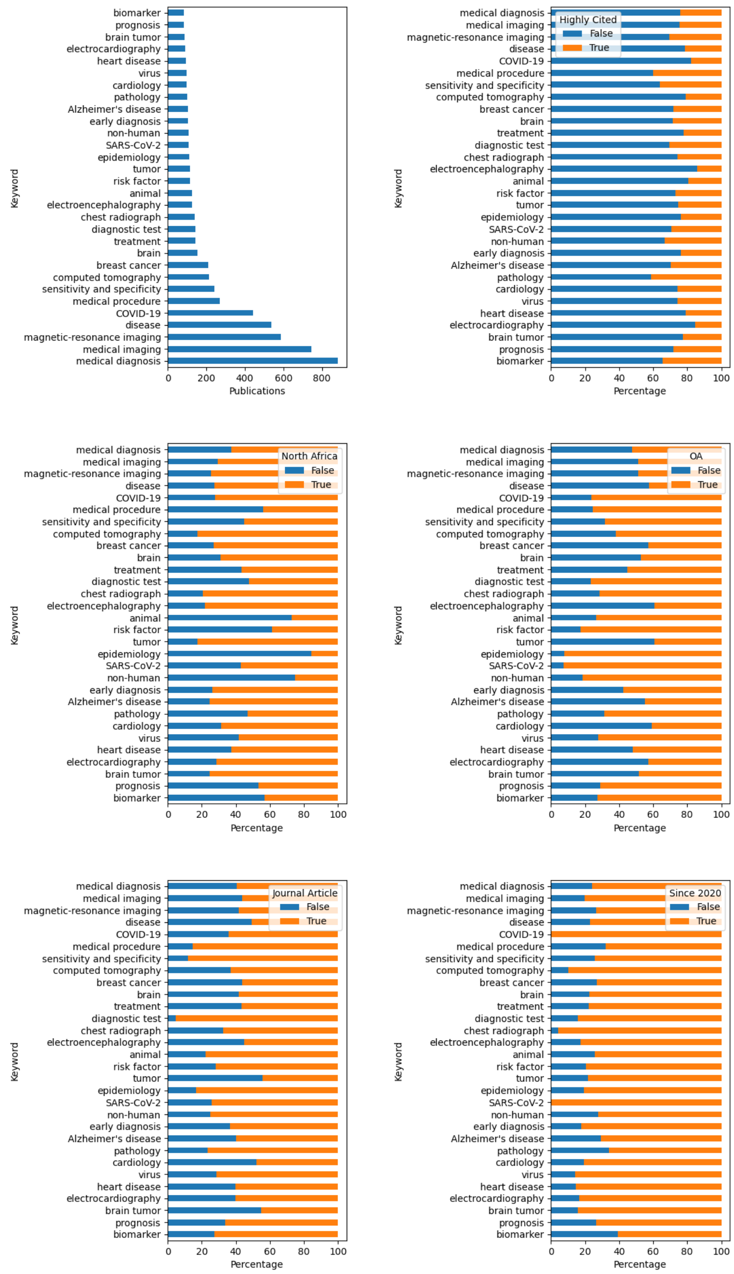

Figure 22: Keyword mentions for biomedical terms: Overall number and rate of their availability in highly-cited publications, North African research production, open-access outputs, journal articles, and works published since 2020.

 doi:10.20944/preprints202302.0010.v2

people are more likely to have severe and deadly symptoms of the infection (Davies et al., 2020). Interestingly, Sub-Saharan Africa is more involved in developing machine-learning algorithms for extreme ages. Further efforts should be provided in this context. Beyond age ranges, several publications also consider the gender of human subjects: *female* (around 390) and *male* (around 320). As a result, research customized for females is more common than the one for males. This is mainly explained by the better development of gynecology and obstetrics (female reproductive medicine) than andrology (male reproductive medicine) in Africa (Zhang et al., 2016). These works are mainly published by Sub-Saharan African scientists as journal articles and did not achieve sufficient citation impact despite the growth of gender-aware biomedical machine-learning applications since 2020 in Africa (Figure 24). That being said, there is room for improvement in this field as there are many pathologies in these disciplines that have not been explored using machine learning and that can be easily analyzed by African scientists using machine learning techniques (Dhombres et al., 2022). Moreover, the social science category includes the names of several regions (Figure 23): *developing country* (around 80), *South Africa* (around 50), *Africa* (around 40), *United States of America* (around 30), *Nigeria* (around 30), *Kenya* (around 30), *Sub-Saharan Africa* (around 30), *Tanzania* (around 30), *Uganda* (around 30), and *Ethiopia* (around 20). We clearly see here that there is a higher trend of developing research works that are continent-wide (*Africa* and *Sub-Saharan Africa*) or customized to developing countries. We also find that there are more machine-learning applications that are customized to Sub-Saharan Africa than to North Africa despite the better research productivity of North African countries. Most of these works have been done since 2020 and this is probably due to efforts related to COVID-19 epidemiology prediction. This explains in part why most of these works are open-access (Lee & Haupt, 2020). This is confirmed by the fact that such region-focused research works are common in public, environmental, and occupational health (Moradi & Asnafi, 2016). These works are mainly done by Sub-Saharan Africa without any participation of North African countries. This is mainly due to the nature of this

 doi:10.20944/preprints202302.0010.v2

kind of research that requires the exclusive participation of local scientists in its design and development (Moradi & Asnafi, 2016). Exceptions to this are the few works featuring the *United States of America*. The United States of America is studied by African scientists, particularly Sub-Saharan African scientists, for two reasons: Comparison of African countries with this developed country to identify improvement directions and better data availability and quality in the United States of America (van Velthoven et al., 2016). Region-aware research works have a fewer rate of highly-cited publications than other research works and are mainly published as journal articles. This is a general trend of such works and this is mainly due to the significantly restricted visibility of such research topics outside their region of interest (Abramo et al., 2016).

As for the fields of interest and research methods included in the keywords (Figures 23 and 24), we find a significant effort towards the development of machine-learning applications for *health care* (almost 480 publications), *forecasting* (almost 320), *decision-making* (almost 150), *risk assessment* (almost 80), *health system* (almost 60), *biology* (almost 60), *health monitoring* (almost 60), *e-learning* (almost 50), *behavioral research* (almost 40), *sustainable development* (almost 30), *population statistics* (almost 30), and *education* (almost 30). Works on *behavioral research*, *e-learning* (in the context of medical education), *sustainable development*, and *health system* are surprisingly dominated by North African scientists although machine-learning applications for the social aspects of medicine are mainly developed in Sub-Saharan Africa as shown before. The domination of North Africa in machine learning research on *e-learning* (Brika et al., 2022; Jia et al., 2022), *behavioral research* (Kim et al., 2021), *sustainable development* (Bracarense et al., 2022) and *health system* preparedness Sweileh (2021) is mainly caused by robust research collaborations between Egypt and Saudi Arabia. By contrast, Sub-Saharan Africa mostly dominates machine learning applications for computational *biology* in the context of medical practice. This is confirmed by Figure 10 where molecular biology research is slightly dominated by Sub-Saharan African nations. This is mainly explained by the existence of research consortia led by South Africa such as *H3ABioNet* that

 doi:10.20944/preprints202302.0010.v2

spreads Bioinformatics across Sub-Saharan Africa with research training and support (Mulder et al., 2015). Domain-aware works use metrics like *mortality rate*, standards like the ones of the *World Health Organization*, questionnaires, reproducibility, predictive analytics, comparison, semantics, follow-ups, and performance evaluation among other techniques. Works that are emphasizing the assessment of machine-learning techniques through reproducibility and performance evaluation are receiving better citation impact and are more published as open-access journal articles, particularly since 2020. This goes in line with the worldwide efforts to ensure the trustworthiness and verification of the efficiency of the medical applications of machine learning within the framework of Evidence-Based Medicine (McDermott et al., 2021). Predictive analysis has largely grown since 2020 due to the beginning of the COVID-19 pandemic that imposed the development of intelligent systems, particularly in North Africa, for forecasting the evolution of the infection for public health monitoring (Figure 24) (Phoobane et al., 2022).

Figures 25 and 26 present the co-occurrence of the most fifty keywords in all the considered publications. According to Figure 25, we notice that the keywords "deep learning", "machine learning", "classification" and "magnetic resonance imaging" are close to each other and positioned in the center of the network. This means that those keywords co-exist in many publications at the same time. We also notice that "medical diagnosis" is linked to "deep learning" and "machine learning" meaning that publications are dealing with the use of machine learning techniques in medical diagnosis. The "Breast cancer" keyword is also linked to "machine learning" which reflects that publications are interested in employing machine learning to deal with breast cancer issues. The Heatmap presented in Figure 26 confirms those results. In fact, "medical diagnosis" in the Heatmap is very tied to "machine learning" and "deep learning". We notice also the occurrence of "magnetic resonance imaging", "machine learning" and "deep learning" in the Heatmap. When comparing the study results with other studies about machine learning in healthcare, we notice that many authors confirm that machine learning techniques are used in healthcare

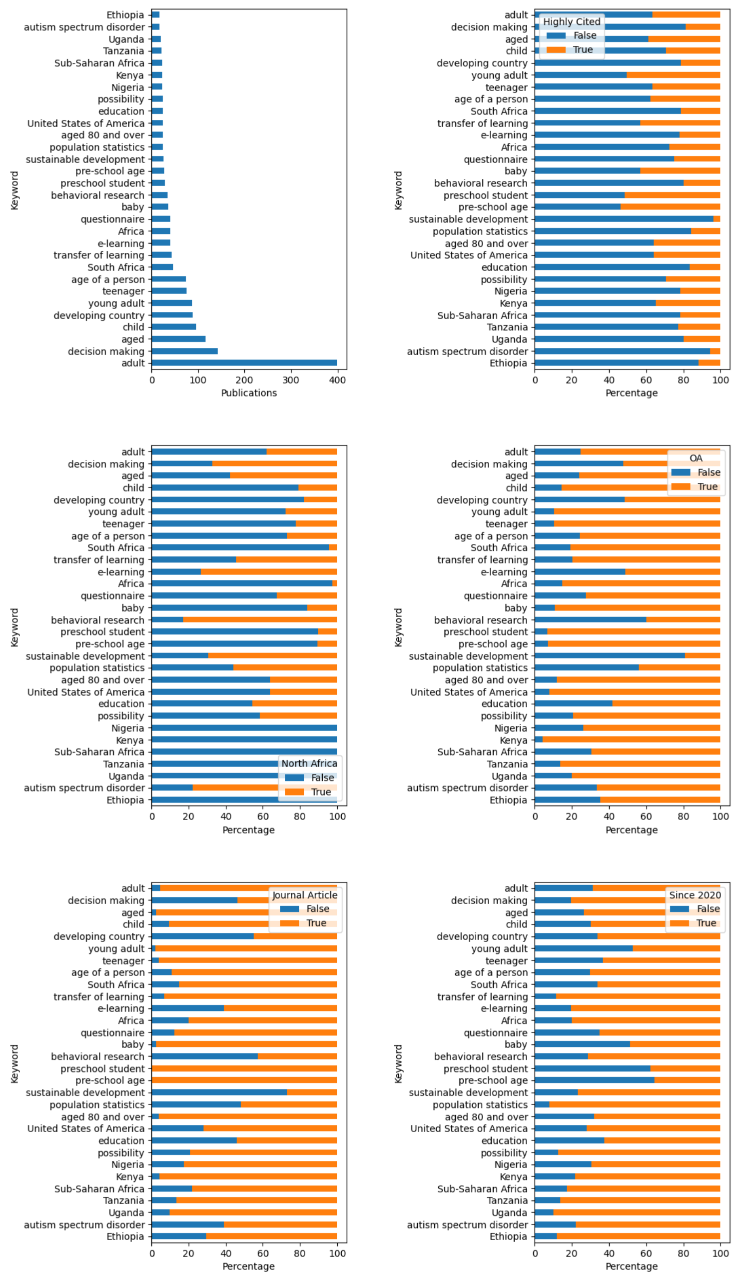

Figure 23: Keyword mentions for social science terms: Overall number and rate of their availability in highly-cited publications, North African research production, open-access outputs, journal articles, and works published since 2020.

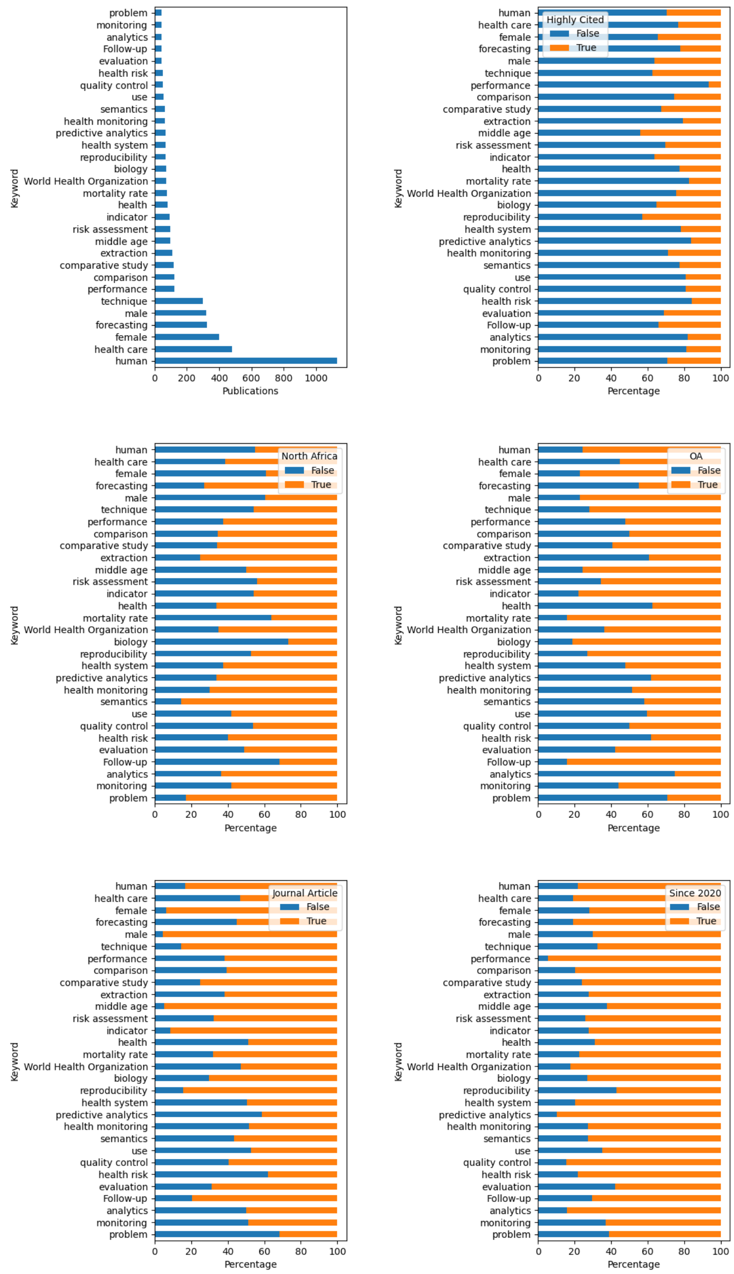

Figure 24: Keyword mentions for other terms: Overall number and rate of their availability in highly-cited publications, North African research production, open-access outputs, journal articles, and works published since 2020.

domains for many reasons such as dealing with massive volumes of data (Hain et al., 2023). In fact, according to Figure 21, many recent North African publications in our data set concern big data domains. According to Hain et al. (2023), publications about the application of AI techniques in healthcare issues are mainly interested in disease diagnosis which is confirmed in Figure 22 where most publications from the category "biomedical terms" concerns "medical diagnosis" (more than 800 publications). Another study (Li et al., 2020) concerning a bibliometric analysis of deep learning confirms that one of the applications of deep learning is related to domains such as "computer-aided diagnosis" which is a topic found in Figure 21. It is also noticed that many studies about machine learning for healthcare are interested in classification algorithms and mainly use "support vector machine" algorithm (Rincon-Patino et al., 2018) which is found in Figures 21, 25 and 26. Medical image recognition and analysis is another important application of deep learning (Hain et al., 2023). In fact, according to results presented in Figures 21, 22, 25, and 26, topics like "medical imaging", "image processing", and "image classification" are of interest to many publications.

5. Conclusion

In this research paper, we conducted a bibliometric study of African contributions to research publications related to machine learning for healthcare (AML4H) as indexed in *Scopus*. We analyzed bibliographic data related to research venues, publication years, publication types, funding, authorship, publishers, open access, and main research areas for every publication. In particular, we investigated publishing patterns over four time periods: *Until 2013*, *2014-2016*, *2017-2019*, and *2020-2022*.

5.1. Key findings

We identified 3,772 AML4H publications between 1993 and 2022 that have at least one contributing author from an African research institution. We found

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 May 2023 doi:10.20944/preprints202302.0010.v2

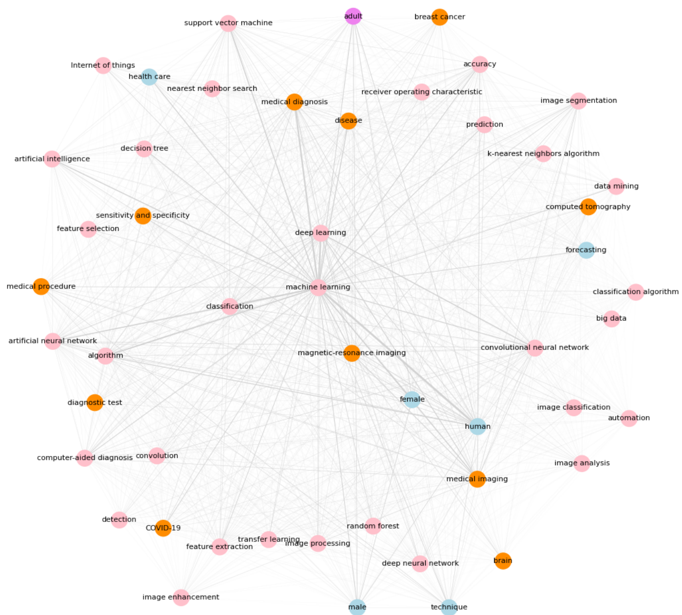

Figure 25: Co-occurrence network for the most common fifty keywords in all the considered research publications: Nodes are unweighted and colored according to their respective classes: *STEM* (Pink), *Biomedicine* (Orange), *Social Sciences* (Purple), and *Other* (Light Blue). Thicker edges correspond to more common keyword associations.

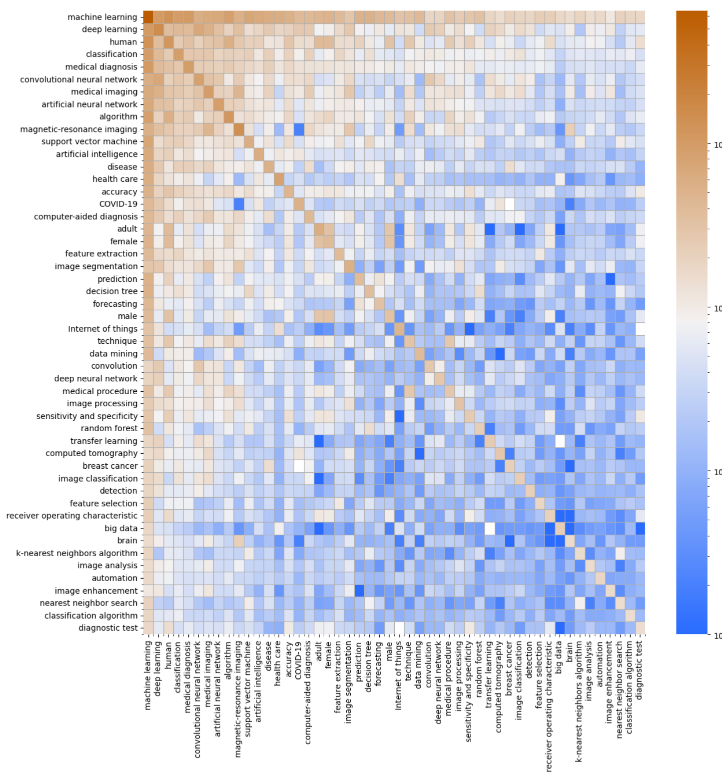

Figure 26: Heatmap visualization of the co-occurrence matrix of the most common fifty keywords in all the considered research publications.

 doi:10.20944/preprints202302.0010.v2

an increasing impetus to publish in the field of machine learning for healthcare from African institutes, likely spurred by the recent growth of machine learning and a global pandemic. Regionally, North African countries contributed 64.5% of AML4H publications on average. This ratio has been shifting, however, from a high of 80.5% in 2012-2013 to 61.5% in 2020-2022.

There is a correlation between international funding for African healthcare research and the rise of AML4H publications. This is particularly evident in Sub-Saharan Africa, which has seen accelerated international funding in recent years. In general, international entities contribute the majority of funds for AML4H publications arising from African institutes, with locally-funded publications garnering less attention.

Advanced deep learning methods are still new in AML4H, especially within critical care systems, but also in research publications. Instead, traditional methods like support vector machines, decision trees, and k-nearest neighbors algorithms are more widely researched and implemented. Furthermore, AML4H technologies are primarily focused on computer-aided diagnosis as opposed to treatment, although this is not unique to the African context. Research has mainly concentrated on lethal diseases such as breast cancer and brain tumors, widespread diseases like COVID-19, and chronic illnesses like Alzheimer's disease and heart disease.

In conclusion, our analysis of AML4H publications involving African researchers and institutions has shown a growing interest in the field and an increasing interest in machine learning algorithms. The rise in AML4H publications from African institutes correlates with international funding for healthcare research, particularly in Sub-Saharan Africa. As the field continues to evolve and mature, it is hoped that these efforts will lead to improved healthcare outcomes for the African population.

5.2. Limitations and future work

The study lacks a thorough assessment of the stakeholder interactions that led to the awakening of the African community within this field during the last

 doi:10.20944/preprints202302.0010.v2

few years. As a future direction, we aim to perform a network analysis of African research publications within the field of machine learning for healthcare. We also envision developing a literature survey of African scholarly publications on biomedical machine learning to study the main topics, findings, and features of the research.

Acknowledgements

We thank *Masakhane*, the African community for Natural Language Processing, for their support in the development of this research paper. We thank Lamia Benhiba (Mohamed V University of Rabat, Morocco), Yawo Kobara (University of Toronto, Canada), and Sakinat Oluwabukonla Folorunso (Olabisi Onabanjo University, Nigeria) for their useful comments and discussion regarding this research manuscript.

Funding

This research work is funded by Wikimedia Foundation (San Francisco, United States of America) within the framework of the *Adapting Wikidata to support clinical practice using Data Science, Semantic Web and Machine Learning* project (`https://w.wiki/5iU2`).

Data Availability

For reproducibility purposes, all the data, visualizations, and source codes that were used to develop this research paper are made available under the MIT License at `https://github.com/SisonkeBiotik-Africa/AfriBioML`.

Conflicts of Interest

All the co-authors are members of *SisonkeBiotik*, a community of African enthusiasts of machine learning for healthcare.

References

Abramo, G., Cicero, T., & D'Angelo, C. A. (2014). Are the authors of highly cited articles also the most productive ones? *Journal of Informetrics*, *8*, 89–97. doi:10.1016/j.joi.2013.10.011.

Abramo, G., D'Angelo, C. A., & Costa, F. D. (2016). The effect of a country's name in the title of a publication on its visibility and citability. *Scientometrics*, *109*, 1895–1909. doi:10.1007/s11192-016-2120-1.

Adunlin, G., Diaby, V., & Xiao, H. (2014). Application of multicriteria decision analysis in health care: a systematic review and bibliometric analysis. *Health Expectations*, *18*, 1894–1905. doi:10.1111/hex.12287.

Akpanudo, S. (2022). Application of Artificial Intelligence Systems to Improve Healthcare Delivery in Africa. *Primary Health Care: Open Access*, *12*, 1–4. doi:10.4172/2167-1079.22.12.1.1000422.

AlRyalat, S. A. S., Malkawi, L. W., & Momani, S. M. (2019). Comparing Bibliometric Analysis Using PubMed, Scopus, and Web of Science Databases. *Journal of Visualized Experiments*, . doi:10.3791/58494.

Alyasseri, Z. A. A., Al-Betar, M. A., Doush, I. A., Awadallah, M. A., Abasi, A. K., Makhadmeh, S. N., Alomari, O. A., Abdulkareem, K. H., Adam, A., Damasevicius, R., Mohammed, M. A., & Zitar, R. A. (2021). Review on COVID-19 diagnosis models based on machine learning and deep learning approaches. *Expert Systems*, *39*. doi:10.1111/exsy.12759.

Arvanitis, R., Mouton, J., & Néron, A. (2022). Funding Research in Africa: Landscapes of Re-institutionalisation. *Science, Technology and Society*, *27*, 351–367. doi:10.1177/09717218221078235.

Asubiaro, T. V., & Shaik, H. (2021). Sub-Saharan African Countries' COVID-19 Research: An analysis of the External and Internal Contributions, Collaboration Patterns and Funding Sources. *Open Information Science*, *5*, 263–277. doi:10.1515/opis-2020-0125.

Bakken, S. (2020). Informatics is a critical strategy in combating the COVID-19 pandemic. *Journal of the American Medical Informatics Association*, *27*, 843–844. doi:10.1093/jamia/ocaa101.

Bartol, T., & Stopar, K. (2015). Nano language and distribution of article title terms according to power laws. *Scientometrics*, *103*, 435–451. doi:10.1007/s11192-015-1546-1.

Baxter, M. S., White, A., Lahti, M., Murto, T., & Evans, J. (2021). Machine learning in a time of COVID-19-Can machine learning support Community Health Workers (CHWs) in low and middle income countries (LMICs) in the new normal? *Journal of Global Health*, *11*. doi:10.7189/jogh.11.03017.

Bayoudhi, L., Sassi, N., & Jaziri, W. (2022). How Latest Computer Science Research Copes with COVID-19? In A. Abraham, N. Gandhi, T. Hanne, T.-P. Hong, T. Nogueira Rios, & W. Ding (Eds.), *Intelligent Systems Design and Applications* (pp. 1207–1215). Cham: Springer International Publishing. doi:10.1007/978-3-030-96308-8_112.

Bellemo, V., Lim, Z. W., Lim, G., Nguyen, Q. D., Xie, Y., Yip, M. Y. T., Hamzah, H., Ho, J., Lee, X. Q., Hsu, W., Lee, M. L., Musonda, L., Chandran, M., Chipalo-Mutati, G., Muma, M., Tan, G. S. W., Sivaprasad, S., Menon, G., Wong, T. Y., & Ting, D. S. W. (2019). Artificial intelligence using deep learning to screen for referable and vision-threatening diabetic retinopathy in Africa: a clinical validation study. *The Lancet Digital Health*, *1*, e35–e44. doi:10.1016/s2589-7500(19)30004-4.

Benos, L., Tagarakis, A. C., Dolias, G., Berruto, R., Kateris, D., & Bochtis, D. (2021). Machine learning in agriculture: A comprehensive updated review. *Sensors*, *21*, 3758. doi:10.3390/s21113758.

Bernstam, E. V., Smith, J. W., & Johnson, T. R. (2010). What is biomedical informatics? *Journal of Biomedical Informatics*, *43*, 104–110. doi:10.1016/j.jbi.2009.08.006.

Björk, B.-C. (2012). The hybrid model for open access publication of scholarly articles: A failed experiment? *Journal of the American Society for Information Science and Technology*, *63*, 1496–1504. doi:10.1002/asi.22709.

Blom, A., Lan, G., & Adil, M. (2015). *Sub-Saharan African Science, Technology, Engineering, and Mathematics Research: A Decade of Development*. Washington, DC: World Bank. doi:10.1596/978-1-4648-0700-8.

Boniol, M., Kunjumen, T., Nair, T. S., Siyam, A., Campbell, J., & Diallo, K. (2022). The global health workforce stock and distribution in 2020 and 2030: a threat to equity and 'universal' health coverage? *BMJ Global Health*, *7*, e009316. doi:10.1136/bmjgh-2022-009316.

Booth, R., Strudwick, G., McMurray, J., Chan, R., Cotton, K., & Cooke, S. (2021). The future of nursing informatics in a digitally-enabled world. In *Health Informatics* (pp. 395–417). Springer International Publishing. doi:10.1007/978-3-030-58740-6_16.

Bracarense, N., Bawack, R. E., Fosso Wamba, S., & Carillo, K. D. A. (2022). Artificial Intelligence and Sustainability: A Bibliometric Analysis and Future Research Directions. *Pacific Asia Journal of the Association for Information Systems*, *14*, 9. doi:10.17705/1pais.14209.

Bragazzi, N. L. (2019). Nanomedicine: Insights from a Bibliometrics-Based Analysis of Emerging Publishing and Research Trends. *Medicina*, *55*, 785. doi:10.3390/medicina55120785.

Brika, S. K. M., Chergui, K., Algamdi, A., Musa, A. A., & Zouaghi, R. (2022). E-Learning Research Trends in Higher Education in Light of COVID-19: A Bibliometric Analysis. *Frontiers in Psychology*, *12*, 762819. doi:10.3389/fpsyg.2021.762819.

Brunskill, E., & Lesh, N. (2010). Routing for rural health: Optimizing community health worker visit schedules. In *AAAI Spring Symposium Se-*

ries. URL: https://www.aaai.org/ocs/index.php/SSS/SSS10/paper/view/1139.

Busari, S., & Adebayo, B. (2018). Nigerian Girls Win Silicon Valley Contest for App that Spots Fake Drugs. *CNN*, . URL: https://www.cnn.com/2018/08/17/africa/nigerian-girls-win-silicon-valley-contest/index.html.

Byass, P. (1987). Computers in Africa: Appropriate Technology? *Comput. Bull.*, *3*, 17. doi:10.5555/27926.27931.

Cahan, P., & Treutlein, B. (2023). A conversation with ChatGPT on the role of computational systems biology in stem cell research. *Stem Cell Reports*, *18*, 1–2. doi:10.1016/j.stemcr.2022.12.009.

Candela, L., Castelli, D., Manghi, P., & Tani, A. (2015). Data journals: A survey. *Journal of the Association for Information Science and Technology*, *66*, 1747–1762. doi:10.1002/asi.23358.

Carpenter, J. R., Todd, J., Baisley, K., Bradley, J., Tumwesigye, N. M., Musonda, P., & Chirwa, T. (2022). Training and capacity building in medical statistics in Sub-Saharan Africa: Impact of the London School of Hygiene & Tropical Medicine MSc in Medical Statistics, 1969 to 2021. *Statistics in Medicine*, *41*, 838–844. doi:10.1002/sim.9304.

Chantrapornchai, C., & Tunsakul, A. (2021). Information extraction tasks based on BERT and SpaCy on tourism domain. *ECTI Transactions on Computer and Information Technology (ECTI-CIT)*, *15*, 108–122. doi:10.37936/ecti-cit.2021151.228621.

Chen, I. Y., Pierson, E., Rose, S., Joshi, S., Ferryman, K., & Ghassemi, M. (2021a). Ethical machine learning in healthcare. *Annual review of biomedical data science*, *4*, 123–144. doi:10.1146/annurev-biodatasci-092820-114757.

Chen, I. Y., Pierson, E., Rose, S., Joshi, S., Ferryman, K., & Ghassemi, M. (2021b). Secure and Robust Ethical machine learning in healthcare. *Annual review of biomedical data science*, *4*, 123–144. doi:10.1146/annurev-biodatasci-092020-114757.

Cheng, V. C.-C., To, K. K., Tse, H., Hung, I. F., & Yuen, K. Y. (2012). Two Years after Pandemic Influenza A/2009/H1N1: What Have We Learned? *Clinical Microbiology Reviews*, *25*, 223 – 263. doi:10.1128/CMR.05012-11.

Chien, S.-C., Chen, Y.-L., Chien, C.-H., Chin, Y.-P., Yoon, C. H., Chen, C.-Y., Yang, H.-C., & Li, Y.-C. J. (2022). Alerts in Clinical Decision Support Systems (CDSS): A Bibliometric Review and Content Analysis. *Healthcare*, *10*, 601. doi:10.3390/healthcare10040601.

Chuang, K.-Y., Chuang, Y.-C., Ho, M., & Ho, Y.-S. (2011). Bibliometric analysis of public health research in Africa: The overall trend and regional comparisons. *South African Journal of Science*, *107*. doi:10.4102/sajs.v107i5/6.309.

Cokol, M., Ozbay, F., & Rodriguez-Esteban, R. (2008). Retraction rates are on the rise. *EMBO reports*, *9*, 2–2. doi:10.1038/sj.embor.7401143.

Collazo-Reyes, F. (2013). Growth of the number of indexed journals of latin america and the caribbean: the effect on the impact of each country. *Scientometrics*, *98*, 197–209. doi:10.1007/s11192-013-1036-2.

Confraria, H., Blanckenberg, J., & Swart, C. (2019). Which Factors Influence International Research Collaboration in Africa? In *Sustainable Development Goals Series* (pp. 243–255). Springer International Publishing. doi:10.1007/978-3-030-14857-7_23.

Currie-Alder, B., Arvanitis, R., & Hanafi, S. (2017). Research in Arabic-speaking countries: Funding competitions, international collaboration, and career incentives. *Science and Public Policy*, *45*, 74–82. doi:10.1093/scipol/scx048.

Currin, C. B., Khoza, P. N., Antrobus, A. D., Latham, P. E., Vogels, T. P., & Raimondo, J. V. (2019). Think: Theory for Africa. *PLOS Computational Biology*, *15*, 1–5. doi:10.1371/journal.pcbi.1007049.

Dabre, R. (2022). ACL Rolling Review: A New Format For Centralized Peer Review. *Journal of Natural Language Processing*, *29*, 230–236. doi:10.5715/jnlp.29.230.

Dang, W., McInnes, M. D. F., Kielar, A. Z., & Hong, J. (2015). A Comprehensive Analysis of Authorship in Radiology Journals. *PLoS One*, *10*, e0139005. doi:10.1371/journal.pone.0139005.

Davies, N. G., Klepac, P., Liu, Y., Prem, K., Jit, M., Pearson, C. A. B., Quilty, B. J., Kucharski, A. J., Gibbs, H., Clifford, S., Gimma, A., van Zandvoort, K., Munday, J. D., Diamond, C., Edmunds, W. J., Houben, R. M. G. J., Hellewell, J., Russell, T. W., Abbott, S., Funk, S., Bosse, N. I., Sun, Y. F., Flasche, S., Rosello, A., Jarvis, C. I., & and, R. M. E. (2020). Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nature Medicine*, *26*, 1205–1211. doi:10.1038/s41591-020-0962-9.

De Felice, F., & Polimeni, A. (2020). Coronavirus Disease (COVID-19): A Machine Learning Bibliometric Analysis. *In Vivo*, *34*, 1613–1617. doi:10.21873/invivo.11951.

Deo, R. C. (2015). Machine learning in medicine. *Circulation*, *132*, 1920–1930. doi:10.1161/circulationaha.115.001593.

Dhombres, F., Bonnard, J., Bailly, K., Maurice, P., Papageorghiou, A. T., & Jouannic, J.-M. (2022). Contributions of Artificial Intelligence Reported in Obstetrics and Gynecology Journals: Systematic Review. *Journal of Medical Internet Research*, *24*, e35465. doi:10.2196/35465.

Dix, A. (2017). Human–computer interaction, foundations and new paradigms. *Journal of Visual Languages & Computing*, *42*, 122–134. doi:10.1016/j.jvlc.2016.04.001.

Dodoo, J. E., Al-Samarraie, H., & Alzahrani, A. I. (2021). Telemedicine use in Sub-Saharan Africa: Barriers and policy recommendations for Covid-19 and beyond. *International Journal of Medical Informatics*, *151*, 104467. doi:10.1016/j.ijmedinf.2021.104467.

Eichbaum, Q. G., Adams, L. V., Evert, J., Ho, M.-J., Semali, I. A., & van Schalkwyk, S. C. (2021). Decolonizing global health education: rethinking institutional partnerships and approaches. *Academic Medicine*, *96.3*, 329–335. doi:10.1097/ACM.0000000000003473.

Elliott, A., Nerima, B., Bagaya, B., Kambugu, A., Joloba, M., Cose, S., Pantaleo, G., Yazdanbakhsh, M., Mabey, D., Dunne, D., Moffett, A., Katunguka Rwakishaya, E., Kaleebu, P., & Katongole Mbidde, E. (2015). Capacity for science in sub-Saharan Africa. *The Lancet*, *385*, 2435–2437. doi:10.1016/s0140-6736(15)61111-4.

Estreguil, C., & Buschke, F. (2022). *The evolving role of the European Commission in research on Africa*. Luxembourg: Publications Office of the European Union. doi:10.2760/38335.

Fadlelmola, F. M., Zass, L., Chaouch, M., Samtal, C., Ras, V., Kumuthini, J., Panji, S., & Mulder, N. (2021). Data Management Plans in the genomics research revolution of Africa: Challenges and recommendations. *Journal of Biomedical Informatics*, *122*, 103900. doi:10.1016/j.jbi.2021.103900.

Fauci, A. S., Touchette, N. A., & Folkers, G. K. (2005). Emerging Infectious Diseases: a 10-Year Perspective from the National Institute of Allergy and Infectious Diseases. *Emerging Infectious Diseases*, *11*, 519–525. doi:10.3201/eid1104.041167.

Fiala, D., & Tutoky, G. (2017). Computer Science Papers in Web of Science: A Bibliometric Analysis. *Publications*, *5*, 23. doi:10.3390/publications5040023.

Fonkou, M. D. M., Bragazzi, N. L., Tsinda, E. K., Bouba, Y., Mmbando, G. S., & Kong, J. D. (2021). COVID-19 Pandemic Related Research in Africa: Bibliometric Analysis of Scholarly Output, Collaborations and Scientific Leadership. *International Journal of Environmental Research and Public Health*, *18*, 7273. doi:10.3390/ijerph18147273.

Forster, D. (1992). *Expert systems in health for developing countries: practice, problems, and potential*. IDRC, Ottawa, ON, CA. URL: http://hdl.handle.net/10625/10733.

Gao, J., Yin, Y., Myers, K. R., Lakhani, K. R., & Wang, D. (2021). Potentially long-lasting effects of the pandemic on scientists. *Nature Communications*, *12*, 6188. doi:10.1038/s41467-021-26428-z.

Garousi, V., & Fernandes, J. M. (2016). Highly-cited papers in software engineering: The top-100. *Information and Software Technology*, *71*, 108–128. doi:10.1016/j.infsof.2015.11.003.

Guo, Y., Hao, Z., Zhao, S., Gong, J., & Yang, F. (2020). Artificial Intelligence in Health Care: Bibliometric Analysis. *Journal of Medical Internet Research*, *22*, e18228. doi:10.2196/18228.

Gwagwa, A., Kraemer-Mbula, E., Rizk, N., Rutenberg, I., & de Beer, J. (2020). Artificial Intelligence (AI) Deployments in Africa: Benefits, Challenges and Policy Dimensions. *The African Journal of Information and Communication*, *26*, 1 – 28. doi:10.23962/10539/30361.

Haensly, P. J., Hodges, P. E., & Davenport, S. A. (2008). Acceptance Rates and Journal Quality: An Analysis of Journals in Economics and Finance. *Journal of Business & Finance Librarianship*, *14*, 2–31. doi:10.1080/08963560802176330.

Hagberg, A. A., Schult, D. A., & Swart, P. J. (2008). Exploring network structure, dynamics, and function using networkx. In G. Varoquaux,

T. Vaught, & J. Millman (Eds.), *Proceedings of the 7th Python in Science Conference (SciPy 2008)* (pp. 11 – 15). Pasadena, CA, USA. URL: https://conference.scipy.org/proceedings/scipy2008/paper_2/.

Hain, D., Jurowetzki, R., Lee, S., & Zhou, Y. (2023). Machine learning and artificial intelligence for science, technology, innovation mapping and forecasting: Review, synthesis, and applications. *Scientometrics*, *128*, 1465–1472. doi:10.1007/s11192-022-04628-8.

Halevi, G., Moed, H., & Bar-Ilan, J. (2017). Suitability of google scholar as a source of scientific information and as a source of data for scientific evaluation—review of the literature. *Journal of Informetrics*, *11*, 823–834. doi:10.1016/j.joi.2017.06.005.

Harper, L., Kalfa, N., Beckers, G., Kaefer, M., Nieuwhof-Leppink, A. J., Fossum, M., Herbst, K., & Bagli, D. (2020). The impact of COVID-19 on research. *Journal of Pediatric Urology*, *16*, 715 – 716. doi:10.1016/j.jpurol.2020.07.002.

Hassani, H., & Silva, E. S. (2023). The Role of ChatGPT in Data Science: How AI-Assisted Conversational Interfaces Are Revolutionizing the Field. *Big Data and Cognitive Computing*, *7*, 62. doi:10.3390/bdcc7020062.

Heidari, A., Navimipour, N. J., Unal, M., & Toumaj, S. (2022). Machine learning applications for COVID-19 outbreak management. *Neural Computing and Applications*, *34*, 15313–15348. URL: https://doi.org/10.1007/s00521-022-07424-w. doi:10.1007/s00521-022-07424-w.

Heldring, L., & Robinson, J. A. (2012). *Colonialism and economic development in Africa*. Technical Report National Bureau of Economic Research. doi:10.3386/w18566.

Herbert, R. (2020). Accept Me, Accept Me Not: What Do Journal Acceptance Rates Really Mean? *SSRN Electronic Journal*, . doi:10.2139/ssrn.3526365.

Houssein, E. H., Mohamed, R. E., & Ali, A. A. (2021). Machine learning techniques for biomedical natural language processing: A comprehensive review. *IEEE Access*, *9*, 140628–140653. doi:10.1109/access.2021.3119621.

Hu, Z., Tian, W., Guo, J., & Wang, X. (2020). Mapping research collaborations in different countries and regions: 1980–2019. *Scientometrics*, *124*, 729–745. doi:10.1007/s11192-020-03484-8.

Hunter, J., Cookson, J., & Wyatt, J. (Eds.) (1989). *AIME 89: Second European Conference on Artificial Intelligence in Medicine, London, August 29th-31st 1989*. Springer Berlin Heidelberg. doi:10.1007/978-3-642-93437-7.

Hyde, K. K., Novack, M. N., LaHaye, N., Parlett-Pelleriti, C., Anden, R., Dixon, D. R., & Linstead, E. (2019). Applications of Supervised Machine Learning in Autism Spectrum Disorder Research: a Review. *Review Journal of Autism and Developmental Disorders*, *6*, 128–146. doi:10.1007/s40489-019-00158-x.

Ibeneme, S., Okeibunor, J., Muneene, D., Husain, I., Bento, P., Gaju, C., Housseynou, B., Chibi, M., Karamagi, H., & Makubalo, L. (2021). Data revolution, health status transformation and the role of artificial intelligence for health and pandemic preparedness in the African context. *BMC Proceedings*, *15*. doi:10.1186/s12919-021-00228-1.

Ioannidis, J. P. A., Klavans, R., & Boyack, K. W. (2018). Thousands of scientists publish a paper every five days. *Nature*, *561*, 167–169. doi:10.1038/d41586-018-06185-8.

Jia, K., Wang, P., Li, Y., Chen, Z., Jiang, X., Lin, C.-L., & Chin, T. (2022). Research Landscape of Artificial Intelligence and e-Learning: A Bibliometric Research. *Frontiers in Psychology*, *13*. doi:10.3389/fpsyg.2022.795039.

Jia, Y., Wang, W., Liang, J., Liu, L., Chen, Z., Zhang, J., Chen, T., & Lei, J. (2018). Trends and characteristics of global medical informatics conferences from 2007 to 2017: A bibliometric comparison of conference publications from chinese, american, european and the global conferences. *Computer Methods and Programs in Biomedicine*, *166*, 19–32. doi:10.1016/j.cmpb.2018.08.017.

Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., & Wang, Y. (2017). Artificial intelligence in healthcare: past, present and future. *Stroke and Vascular Neurology*, *2*, 230–243. doi:10.1136/svn-2017-000101.

Jimma, B. L. (2023). Artificial intelligence in healthcare: A bibliometric analysis. *Telematics and Informatics Reports*, *9*, 100041. doi:10.1016/j.teler.2023.100041.

Joaquin, J. J., & Tan, R. R. (2021). The lost art of short communications in academia. *Scientometrics*, *126*, 9633–9637. doi:10.1007/s11192-021-04192-7.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*, 255–260. doi:10.1126/science.aaa8415.

Jowi, J., Ong'ondo, C. O., Nega, M., Sehoole, C., Alabi, G., Dimé, M., Barasa, P., & Akudolu, L. (2018). *Building PhD Capacity in Sub-Saharan Africa*. German Academic Exchange Service. URL: https://www.britishcouncil.org/sites/default/files/h233_07_synthesis_report_final_web.pdf.

Kastner, J. K., Dawson, C. R., Weiss, S. M., Kern, K. B., & Kulikowski, C. A. (1984). An expert consultation system for frontline health workers in primary eye care. *Journal of Medical Systems*, *8*, 389–397. doi:10.1007/bf02285251.

Kaswa, R., Nair, A., Murphy, S., & Pressentin, K. B. V. (2022). Artificial intelligence: A strategic opportunity for enhancing primary care in South Africa. *South African Family Practice*, *64*. doi:10.4102/safp.v64i1.5596.

Kaur, D., Uslu, S., Rittichier, K. J., & Durresi, A. (2022). Trustworthy Artificial Intelligence: A Review. *ACM Computing Surveys*, *55*, 1–38. doi:10.1145/3491209.

Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., & King., D. (2019). Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine*, *17*, 1–9. doi:10.1186/s12916-019-1426-2.

Kim, J., Lee, D., & Park, E. (2021). Machine Learning for Mental Health in Social Media: Bibliometric Study. *Journal of Medical Internet Research*, *23*, e24870. doi:10.2196/24870.

Korbel, J. O., & Stegle, O. (2020). Effects of the COVID-19 pandemic on life scientists. *Genome Biology*, *21*. doi:10.1186/s13059-020-02031-1.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*, 84–90. doi:10.1145/3065386.

Krüger, D., & Marshall, D. M. (2017). Bite-size research: BMC research notes goes back to its roots. *BMC Research Notes*, *10*. doi:10.1186/s13104-017-2418-y.

Kwiek, M. (2015). The European research elite: a cross-national study of highly productive academics in 11 countries. *Higher Education*, *71*, 379–397. doi:10.1007/s10734-015-9910-x.

Laakso, M. (2013). Green open access policies of scholarly journal publishers: a study of what, when, and where self-archiving is allowed. *Scientometrics*, *99*, 475–494. doi:10.1007/s11192-013-1205-3.

Landini, F., Malerba, F., & Mavilia, R. (2015). The structure and dynamics of networks of scientific collaborations in Northern Africa. *Scientometrics*, *105*, 1787–1807. doi:10.1007/s11192-015-1635-1.

Larbi, D., Sarheim Anthun, K., Nah Asah, F., Debrah, O., & Antypas, K. (2022). Assessing Strategic Priority Factors in eHealth policies of Four African Countries. In *2022 IST-Africa Conference (IST-Africa)*. IEEE. doi:10.23919/ist-africa56635.2022.9845650.

Larivière, V., Haustein, S., & Mongeon, P. (2015). The oligopoly of academic publishers in the digital era. *PLOS ONE*, *10*, e0127502. doi:10.1371/journal.pone.0127502.

Lee, J. J., & Haupt, J. P. (2020). Scientific globalism during a global crisis: research collaboration and open access publications on COVID-19. *Higher Education*, *81*, 949–966. doi:10.1007/s10734-020-00589-0.

Li, Y., Shan, B., Li, B., Liu, X., & Pu, Y. (2021). Literature Review on the Applications of Machine Learning and Blockchain Technology in Smart Healthcare Industry: A Bibliometric Analysis. *Journal of Healthcare Engineering*, *2021*, 1–11. doi:10.1155/2021/9739219.

Li, Y., Xu, Z., Wang, X., & Wang, X. (2020). A bibliometric analysis on deep learning during 2007–2019. *International Journal of Machine Learning and Cybernetics*, *11*, 2807–2826. doi:10.1007/s13042-020-01152-0.

Lidströmer, N., & Ashrafian, H. (Eds.) (2022). *Artificial Intelligence in Medicine*. Springer International Publishing. doi:10.1007/978-3-030-64573-1.

London, S. D., Fontelo, P., Boroumand, S., & Dye, B. A. (2022). COVID-19 provides an opportunity for integration of dentistry into the health informatics system. *The Journal of the American Dental Association*, *153*, 3–8. doi:10.1016/j.adaj.2021.11.003.

Luna, D., Almerares, A., Mayan, J. C., de Quirós, F. G. B., & Otero, C. (2014). Health Informatics in Developing Countries: Going beyond Pilot Practices to Sustainable Implementations: A Review of the Current Challenges. *Healthcare Informatics Research*, *20*, 3. doi:10.4258/hir.2014.20.1.3.

Maina, M. B., Ahmad, U., Ibrahim, H. A., Hamidu, S. K., Nasr, F. E., Salihu, A. T., Abushouk, A. I., Abdurrazak, M., Awadelkareem, M. A., Amin, A., Imam, A., Akinrinade, I. D., Yakubu, A. H., Azeez, I. A., Mohammed, Y. G., Adamu, A. A., Ibrahim, H. B., Bukar, A. M., Yaro, A. U., Goni, B. W., Prieto-Godino, L. L., & Baden, T. (2021). Two decades of neuroscience publication trends in Africa. *Nature Communications*, *12*. doi:10.1038/s41467-021-23784-8.

Martín-Martín, A., Thelwall, M., Orduna-Malea, E., & López-Cózar, E. D. (2020). Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: a multidisciplinary comparison of coverage via citations. *Scientometrics*, *126*, 871–906. doi:10.1007/s11192-020-03690-4.

McDermott, M. B. A., Wang, S., Marinsek, N., Ranganath, R., Foschini, L., & Ghassemi, M. (2021). Reproducibility in machine learning for health research: Still a ways to go. *Science Translational Medicine*, *13*. doi:10.1126/scitranslmed.abb1655.

McKinney, W. (2011). pandas: a foundational python library for data analysis and statistics. In *Python for High Performance and Scientific Computing (PyHPC@SC2011)* (p. 9). URL: https://www.dlr.de/sc/Portaldata/15/Resources/dokumente/pyhpc2011/submissions/pyhpc2011_submission_9.pdf.

Moradi, S., & Asnafi, A. R. (2016). Analysis of citation rate of papers with titles containing a country name. *Webology*, *13*, 35–60. URL: http://eprints.rclis.org/32202/.

Morrison, H., Salhab, J., Calvé-Genest, A., & Horava, T. (2015). Open access article processing charges: DOAJ survey may 2014. *Publications*, *3*, 1–16. doi:10.3390/publications3010001.

Moyo, S., Doan, T. N., Yun, J. A., & Tshuma, N. (2018). Application of machine learning models in predicting length of stay among healthcare workers in underserved communities in South Africa. *Human Resources for Health*, *16*. doi:10.1186/s12960-018-0329-1.

Mulder, N., Abimiku, A., Adebamowo, S. N., de Vries, J., Matimba, A., Olowoyo, P., Ramsay, M., Skelton, M., & Stein, D. J. (2018). H3Africa: current perspectives. *Pharmacogenomics and Personalized Medicine*, *11*, 59–66. URL: https://doi.org/10.2147/pgpm.s141546. doi:10.2147/pgpm.s141546.

Mulder, N. J., Adebiyi, E., Alami, R., Benkahla, A., Brandful, J., Doumbia, S., Everett, D., Fadlelmola, F. M., Gaboun, F., Gaseitsiwe, S., Ghazal, H., Hazelhurst, S., Hide, W., Ibrahimi, A., Fakim, Y. J., Jongeneel, C. V., Joubert, F., Kassim, S., Kayondo, J., Kumuthini, J., Lyantagaye, S., Makani, J., Alzohairy, A. M., Masiga, D., Moussa, A., Nash, O., Oukem-Boyer, O. O. M., Owusu-Dabo, E., Panji, S., Patterton, H., Radouani, F., Sadki, K., Seghrouchni, F., Özlem Tastan Bishop, Tiffin, N., & and, N. U. (2015). H3ABioNet, a sustainable pan-African bioinformatics network for human heredity and health in Africa. *Genome Research*, *26*, 271–277. doi:10.1101/gr.196295.115.

Mullan, F., Frehywot, S., Omaswa, F., Buch, E., Chen, C., Greysen, S. R., Wassermann, T., Abubakr, D. E. E., Awases, M., Boelen, C., Diomande, M. J.-M. I., Dovlo, D., Ferro, J., Haileamlak, A., Iputo, J., Jacobs, M., Koumaré, A. K., Mipando, M., Monekosso, G. L., Olapade-Olaopa, E. O., Rugarabamu, P., Sewankambo, N. K., Ross, H., Ayas, H., Chale, S. B., Cyprien, S., Cohen, J., Haile-Mariam, T., Hamburger, E., Jolley, L., Kolars, J. C., Kombe, G., & Neusy, A.-J. (2011). Medical schools in sub-Saharan Africa. *The Lancet*, *377*, 1113–1121. doi:10.1016/s0140-6736(10)61961-7.

Mutebi, M., Lewison, G., Aggarwal, A., Alatise, O., Booth, C., Cira, M., Grover, S., Ginsburg, O., Gralow, J., Gueye, S. M., Kithaka, B., Kingham, P., Kochbati, L., Moodley, J., Muhammed, S., Mutombo, A., Ndlovu, N., Ntizimira, C., Parham, G., Walter, F., Parkes, J., Shamley, D., Hammad, N., Seeley, J., Torode, J., Sullivan, R., & Vanderpuye, V. (2022). Cancer research across Africa: a comparative bibliometric analysis. *The Lancet Global health*, . doi:10.1136/bmjgh-2022-009849.

Ocheni, S., & Nwankwo, B. C. (2012). Analysis of colonialism and its impact in Africa. *Cross-Cultural Communication*, *8*, 46–54. doi:10.3968/j.ccc.1923670020120803.1189.

Ondari-Okemwa, E. (2007). Scholarly publishing in sub-saharan africa in the twenty-first century: Challenges and opportunities. *First Monday*, *12*. doi:10.5210/fm.v12i10.1966.

Onu, C. C., Lebensold, J., Hamilton, W. L., & Precup, D. (2019). Neural Transfer Learning for Cry-Based Diagnosis of Perinatal Asphyxia. In *Interspeech 2019*. ISCA. doi:10.21437/interspeech.2019-2340.

Owoyemi, A., Owoyemi, J., Osiyemi, A., & Boyd, A. (2020). Artificial Intelligence for Healthcare in Africa. *Frontiers in Digital Health*, *2*, 6. doi:10.3389/fdgth.2020.00006.

Palmatier, R. W., Houston, M. B., & Hulland, J. (2017). Review articles: purpose, process, and structure. *Journal of the Academy of Marketing Science*, *46*, 1–5. doi:10.1007/s11747-017-0563-4.

Phoobane, P., Masinde, M., & Mabhaudhi, T. (2022). Predicting Infectious Diseases: A Bibliometric Review on Africa. *International Journal of Environmental Research and Public Health*, *19*, 1893. doi:10.3390/ijerph19031893.

Pouris, A. (2017). *Bibliometric Analyses on EU-Africa Research Co-publications in Health*. CAAST-NET and Department of Science and Innovation, Pretoria.

Pouris, A., & Ho, Y.-S. (2014). Research emphasis and collaboration in Africa. *Scientometrics*, *98*, 2169–2184. doi:10.1007/s11192-013-1156-8.

Pugsley, M. K., Bekele, B., Griessel, H., de Korte, T., Authier, S., Grobler, A. F., Markgraf, C. G., & Curtis, M. J. (2020). Twenty years of safety pharmacology model validation and the wider implications of this to drug discovery. *Journal of Pharmacological and Toxicological Methods*, *105*, 106912. doi:10.1016/j.vascn.2020.106912.

Qayyum, A., Qadir, J., Bilal, M., & Al-Fuqaha, A. (2021). Secure and robust machine learning for healthcare: A survey. *IEEE Reviews in Biomedical Engineering*, *14*, 156–180. doi:10.1109/RBME.2020.3013489.

Rahman, M., & Fukui, T. (2001). Factors Related to Biomedical Research Productivity in Asian Countries. *Journal of Epidemiology*, *11*, 199–202. doi:10.2188/jea.11.199.

Raman, R., Singh, P., Singh, V. K., Vinuesa, R., & Nedungadi, P. (2022). Understanding the Bibliometric Patterns of Publications in IEEE Access. *IEEE Access*, *10*, 35561–35577. doi:10.1109/access.2022.3161639.

Reuter, E. (2022). 5 takeaways from the FDA's list of AI-enabled medical devices. https://www.medtechdive.com/news/FDA-AI-ML-medical-devices-5-takeaways/635908/.

Rincon-Patino, J., Ramirez-Gonzalez, G., & Corrales, J. C. (2018). Exploring machine learning: A bibliometric general approach using SciMAT. *F1000Research*, *7*, 1210. doi:10.12688/f1000research.15620.1.

Roscoe, J. (2022). The need for accelerated change in diversity, equity and inclusion in publishing and learned societies. *Learned Publishing*, *35*, 481–488. doi:10.1002/leap.1457.

dos Santos, B. S., Arns Steiner, M. T., Trojan Fenerich, A., & Palma Lima, R. H. (2019). Data mining and machine learning techniques applied to public health problems: A bibliometric analysis from 2009 to 2018. *Computers & Industrial Engineering*, *138*, 106120. doi:10.1016/j.cie.2019.106120.

Sarwar, R., & Hassan, S.-U. (2015). A bibliometric assessment of scientific productivity and international collaboration of the Islamic World in science and technology (s&t) areas. *Scientometrics*, *105*, 1059–1077. doi:10.1007/s11192-015-1718-z.

Scellato, G., Franzoni, C., & Stephan, P. (2015). Migrant scientists and international networks. *Research Policy*, *44*, 108–120. doi:10.1016/j.respol.2014.07.014.

Schwarz Rodrigues, R., Abadal, E., & Kricheldorf Hermes de Araújo, B. (2020). Open access publishers: The new players. *PLOS ONE*, *15*, e0233432. doi:10.1371/journal.pone.0233432.

SCImago (2022). Scimago journal & country rank [portal]. URL: https://www.scimagojr.com/journalrank.php.

Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.-F., & Dennison, D. (2015). Hidden Technical Debt in Machine Learning Systems. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems*. Curran Associates, Inc. volume 28. URL: https://proceedings.neurips.cc/paper_files/paper/2015/file/86df7dcfd896fcaf2674f757a2463eba-Paper.pdf.

Shaffer, J. G., Mather, F. J., Wele, M., Li, J., Tangara, C. O., Kassogue, Y., Srivastav, S. K., Thiero, O., Diakite, M., Sangare, M., Dabitao, D., Toure, M., Djimde, A. A., Traore, S., Diakite, B., Coulibaly, M. B., Liu, Y., Lacey, M., Lefante, J. J., Koita, O., Schieffelin, J. S., Krogstad, D. J., & Doumbia, S. O. (2019). Expanding Research Capacity in Sub-Saharan Africa Through Informatics, Bioinformatics, and Data Science Training Programs in Mali. *Frontiers in Genetics*, *10*, 331. doi:10.3389/fgene.2019.00331.

Shailaja, K., Seetharamulu, B., & Jabbar, M. A. (2018). Machine learning in healthcare: A review. In *2018 Second International Conference on Elec-*

tronics, Communication and Aerospace Technology (ICECA) (p. 910–914). doi:10.1109/ICECA.2018.8474918.

Shortliffe, E. H. (2010). Biomedical informatics in the education of physicians. *JAMA*, *304*, 1227. doi:10.1001/jama.2010.1262.

Simard, M.-A., Ghiasi, G., Mongeon, P., & Larivière, V. (2022). National differences in dissemination and use of open access literature. *PLoS One*, *17*, e0272730. doi:10.1371/journal.pone.0272730.

Singh, V. K., Piryani, R., & Srichandan, S. S. (2020). The case of significant variations in gold–green and black open access: evidence from indian research output. *Scientometrics*, *124*, 515–531. doi:10.1007/s11192-020-03472-y.

Smithakin, K., Bearden, C., Pittenger, S., & Bernstam, E. (2007). Toward a veterinary informatics research agenda: An analysis of the PubMed-indexed literature. *International Journal of Medical Informatics*, *76*, 306–312. doi:10.1016/j.ijmedinf.2006.02.009.

Sooryamoorthy, R. (2018). The production of science in Africa: an analysis of publications in the science disciplines, 2000–2015. *Scientometrics*, *115*, 317–349. doi:10.1007/s11192-018-2675-0.

Sooryamoorthy, R. (2022). Science in Africa: Contemporary Trends in Research. *Journal of Scientometric Research*, *10*, 366–372. doi:10.5530/jscires.10.3.54.

Sweileh, W. M. (2021). Global Research Activity on Health System Preparedness Against Viral Infectious Disease Outbreaks. *Disaster Medicine and Public Health Preparedness*, *16*, 1959–1965. doi:10.1017/dmp.2021.205.

Tahamtan, I., Afshar, A. S., & Ahamdzadeh, K. (2016). Factors affecting number of citations: a comprehensive review of the literature. *Scientometrics*, *107*, 1195–1225. doi:10.1007/s11192-016-1889-2.

Tapera, R., & Singh, Y. (2021). A Bibliometric Analysis of Medical Informatics and Telemedicine in Sub-Saharan Africa and BRICS Nations. *Journal of Public Health Research*, *10*, jphr.2021.1903. doi:10.4081/jphr.2021.1903.

Tchuitcheu, G. K., Mostert, N., Ndlovu, K., Oluoch, T., da Costa Vroom, F., Wanyee, S., Nanann, I., & Wright, G. (2020). Pan African Health Informatics Association (HELINA). *Yearbook of Medical Informatics*, *29*, 284–285. doi:10.1055/s-0040-1701967.

Tijssen, R. J. W., & Winnink, J. (2016). Twenty-first century macro-trends in the institutional fabric of science: bibliometric monitoring and analysis. *Scientometrics*, *109*, 2181–2194. doi:10.1007/s11192-016-2041-z.

Toivanen, H., & Ponomariov, B. (2011). African regional innovation systems: bibliometric analysis of research collaboration patterns 2005–2009. *Scientometrics*, *88*, 471–493. doi:10.1007/s11192-011-0390-1.

Tran, B., Vu, G., Ha, G., Vuong, Q.-H., Ho, M.-T., Vuong, T.-T., La, V.-P., Ho, M.-T., Nghiem, K.-C., Nguyen, H., Latkin, C., Tam, W., Cheung, N.-M., Nguyen, H.-K., Ho, C., & Ho, R. (2019). Global Evolution of Research in Artificial Intelligence in Health and Medicine: A Bibliometric Study. *Journal of Clinical Medicine*, *8*, 360. doi:10.3390/jcm8030360.

Turki, H., Ben Aouicha, M., & Hadj Taieb, M. A. (2019). Discussing Arab Spring's effect on scientific productivity and research performance in Arab countries. *Scientometrics*, *120*, 337–339. doi:10.1007/s11192-019-03127-7.

Turki, H., Dossou, B. F. P., Emezue, C. C., Hadj Taieb, M. A., Ben Aouicha, M., Ben Hassen, H., & Masmoudi, A. (2022a). Mesh2matrix: Machine learning-driven biomedical relation classification based on the mesh keywords of pubmed scholarly publications. In *Proceedings of the 12th International Workshop on Bibliometric-enhanced Information Retrieval co-located with the 44th European Conference on Information Retrieval (ECIR 2022)* (pp. 45–60). CEUR-WS.org. URL: https://ceur-ws.org/Vol-3230/paper-07.pdf.

Turki, H., Hadj Taieb, M. A., & Ben Aouicha, M. (2018). The value of letters to the editor. *Scientometrics*, *117*, 1285–1287. doi:10.1007/s11192-018-2906-4.

Turki, H., Hadj Taieb, M. A., & Ben Aouicha, M. (2021). Knowledge-Based Construction of Confusion Matrices for Multi-Label Classification Algorithms using Semantic Similarity Measures. In *Proceedings of the Workshop on Data meets Applied Ontologies in Explainable AI (DAO-XAI 2021)* (pp. 6:1–6:11). CEUR-WS. doi:10.48550/ARXIV.2011.00109.

Turki, H., Hadj Taieb, M. A., Ben Aouicha, M., & Pouris, A. (2023). Infectious epidemics and the research output of nations: A data-driven analysis. *Journal of Information Science*, *49*, 411–436. doi:10.1177/01655515211006605.

Turki, H., Priskorn, D., Hadj Taieb, M. A., Ben Aouicha, M., & Piad-Morffis, A. (2022b). Enhancing multilingual and biomedical named entity recognition using wikidata semantic relations. In *Wiki Workshop (Non-Archival Report)*. Zenodo. doi:10.5281/ZENODO.6923550.

United Nations Economic Commission for Africa (2021). African Statistical Yearbook. doi:10.18356/57dc2605-en-fr.

Uwizeye, D., Karimi, F., Thiong'o, C., Syonguvi, J., Ochieng, V., Kiroro, F., Gateri, A., Khisa, A. M., & Wao, H. (2022). Factors associated with research productivity in higher education institutions in Africa: a systematic review. *AAS Open Research*, *4*, 26. doi:10.12688/aasopenres.13211.2.

Valenti, A., Fortuna, G., Barillari, C., Cannone, E., Boccuni, V., & Iavicoli, S. (2021). The future of scientific conferences in the era of the COVID-19 pandemic: Critical analysis and future perspectives. *Industrial Health*, *59*, 334–339. doi:10.2486/indhealth.2021-0102.

van Velthoven, M. H., Mastellos, N., Majeed, A., O'Donoghue, J., & Car, J. (2016). Feasibility of extracting data from electronic medical records for research: an international comparative study. *BMC Medical Informatics and Decision Making*, *16*. doi:10.1186/s12911-016-0332-1.

Vernon, D. (2019). Robotics and Artificial Intelligence in Africa [Regional]. *IEEE Robotics & Automation Magazine*, *26*, 131–135. doi:10.1109/mra.2019.2946107.

Vogel, A. L., Puricelli Perin, D. M., Lu, Y.-L., & Taplin, S. H. (2019). Understanding the Value of International Research Networks: An Evaluation of the International Cancer Screening Network of the US National Cancer Institute. *Journal of Global Oncology*, (pp. 1–12). doi:10.1200/jgo.19.00197.

Wainer, J., Xavier, E. C., & Bezerra, F. (2009). Scientific production in computer science: A comparative study of brazil and other countries. *Scientometrics*, *81*, 535–547. doi:10.1007/s11192-008-2156-y.

Waskom, M. (2021). seaborn: statistical data visualization. *Journal of Open Source Software*, *6*, 3021. doi:10.21105/joss.03021.

Wijesooriya, N. R., Mishra, V., Brand, P. L., & Rubin, B. K. (2020). COVID-19 and telehealth, education, and research adaptations. *Paediatric Respiratory Reviews*, *35*, 38–42. doi:10.1016/j.prrv.2020.06.009.

Winks, S., Woodland, J. G., Pillai, G. C., & Chibale, K. (2022). Fostering drug discovery and development in Africa. *Nature Medicine*, *28*, 1523–1526. doi:10.1038/s41591-022-01885-1.

Wonkam, A. (2021). Sequence three million genomes across Africa. *Nature*, *590*, 209–211. doi:10.1038/d41586-021-00313-7.

Yang, S., Xing, X., Qi, F., & Grácio, M. C. C. (2021). Comparison of academic book impact from a disciplinary perspective: an analysis of citations and altmetric indicators. *Scientometrics*, *126*, 1101–1123. doi:10.1007/s11192-020-03808-8.

Zhang, K., Ma, B., Hu, K., Yuan, B., Sun, X., Song, X., Tang, Z., Lin, H., Zhu, X., Zheng, Y., García, A. J., Mikos, A. G., Anderson, J. M., & Zhang, X. (2022). Evidence-based biomaterials research. *Bioactive Materials*, *15*, 495–503. doi:10.1016/j.bioactmat.2022.04.014.

Zhang, Y., Xiao, F., Lu, S., Song, J., Zhang, C., Li, J., Gu, K., Lan, A., Lv, B., Zhang, R., Mo, F., Jiang, G., Zhang, X., & Yang, X. (2016). Research trends and perspectives of male infertility: a bibliometric analysis of 20 years of scientific literature. *Andrology*, *4*, 990–1001. doi:10.1111/andr.12204.

Zhao, C., Joglekar, G., Jain, A., Venkatasubramanian, V., & Reklaitis, G. (2005). Pharmaceutical informatics: A novel paradigm for pharmaceutical product development and manufacture. In *Computer Aided Chemical Engineering* (pp. 1561–1566). Elsevier. doi:10.1016/s1570-7946(05)80102-6.

Zhao, W., Zhang, L., Wang, J., & Wang, L. (2022). How has academia responded to the urgent needs created by COVID-19? a multi-level global, regional and national analysis. *Journal of Information Science*, (p. 01655515221084646). doi:10.1177/01655515221084646.

Zhou, P., Cai, X., & Lyu, X. (2020). An in-depth analysis of government funding and international collaboration in scientific research. *Scientometrics*, *125*, 1331–1347. doi:10.1007/s11192-020-03595-2.

Ziltener, P., & Künzler, D. (2013). Impacts of Colonialism: A Research Survey. *Journal of World-Systems Research*, *19*, 290–311. doi:10.5195/jwsr.2013.507.

Zyoud, S. H. (2021). The Arab region's contribution to global COVID-19 research: Bibliometric and visualization analysis. *Globalization and Health*, *17*. doi:10.1186/s12992-021-00690-8.