Artificial intelligence and machine learning for early detection and diagnosis of colorectal cancer in sub-Saharan Africa

Akbar K Waljee ,[1,2,3,4] Eileen M Weinheimer-Haus ,[2,3,4] Amina Abubakar,[5] Anthony K Ngugi,[6] Geoffrey H Siwo ,[2,3,7,8] Gifty Kwakye,[9] Amit G Singal ,[10,11] Arvind Rao ,[4,12,13,14] Sameer D Saini,[1,2] Andrew J Read ,[2,4] Jessica A Baker,[1,3,4] Ulysses Balis ,[15] Christopher K Opio ,[16] Ji Zhu,[3,4,17] Mansoor N Saleh[18,19]

INTRODUCTION

Colorectal cancer (CRC) was once considered a rare disease in sub-Saharan Africa (SSA), but decades of globalisation has changed this narrative. Currently, CRC is the fifth most common cancer in SSA, and while CRC incidence and mortality are decreasing in some high-income countries, rates in SSA are on the rise.[1] Because CRC develops from a benign precursor polyp over several years, early detection is critical to either prevent malignancy or detect it at an early stage when it is highly curable. Moreover, curative surgery has been shown to improve survival in a SSA setting.[2] Unfortunately, more than 60% of patients in SSA present with stage 4 CRC with a <1% 5 year survival rate.[3–5] In contrast, almost 40% of patients in the USA present with stage 1 CRC, resulting in a 5-year survival rate of 90%.[6 7] Widespread population-based CRC screening programmes and tools (eg, faecal immunochemical test (FIT), colonoscopy) have improved early detection in high-income countries, but SSA-specific data, tools and screening programmes are currently lacking. There is an urgent need to develop more efficient approaches to CRC screening and early detection that do not rely heavily on trained healthcare personnel or specialised resources (eg, endoscopy, pathology), which are often scarce in low- and middle-income countries (LMICs).

Recent technological advances and developments in artificial intelligence (AI) and machine learning (ML) methods have the potential to transform global health, particularly for early detection and diagnosis of CRC in SSA. Researchers are collecting enormous volumes of data, and while data science applications are largely underdeveloped in Africa, many enabling factors are already in place. Developments in cloud computing, substantial investments in digitising health information, and robust mobile phone penetration have poised many places in SSA with the necessary basics to initiate meaningful AI/ML applications.[8] Businesses in SSA have already embraced technological change, leapfrogging high-income countries in the proliferation of mobile banking (eg, M-PESA - one of the first mobile banking system for those with limited access or no access to banks in Africa.).[9] Furthermore, intergovernmental agencies have convened high-profile meetings discussing the development and democratisation of AI solutions to address specific global challenges.[10 11] The United Nations has highlighted the centrality of AI to achieve its Sustainable Development Goals.[2] The National Institutes of Health in the United States has invested about US$74.5 million over 5 years to advance data science, catalyse innovation and spur health discoveries across Africa under its new Harnessing Data Science for Health Discovery and Innovation in Africa (DS-I Africa) programme.[11] Given these resources and investments, the impact of AI/ML applications on healthcare in SSA is imminent.

Herein, we discuss how AI/ML tools could be leveraged to conduct population-based surveillance and improve the early diagnosis and prognosis of CRC in SSA. We highlight limitations to the currently available CRC screening programmes and tools in the SSA setting and provide two examples of potential AI/ML approaches: (1) Multianalyte Assays with Algorithmic Analysis (MAAA) for population-based surveillance and early detection and (2) pattern recognition and computer vision algorithms to guide diagnostic recommendations and prognosis. While CRC is the use case, we also discuss how current initiatives around data science capacity in Africa offer a platform to scale such AI-based solutions to other potential high impact areas such as maternal, newborn, and child health and the growing burden of non-communicable diseases (eg, other cancers, diabetes, cardiovascular disease) in Africa. Lastly, we highlight how these innovative solutions have the potential to impact health outcomes in high-income countries through reciprocal innovation.[12–15]

LIMITATIONS TO CURRENT CRC SCREENING TOOLS IN SSA

Screening programmes and policies around CRC prevention and detection are lacking in SSA. Furthermore, data on disease aetiology and prevalence are sparse, leaving practitioners with a limited knowledge base on the disease in their communities and inadequate

[1]Veterans Affairs Center for Clinical Management Research, Ann Arbor, Michigan, USA
[2]Department of Internal Medicine, Division of Gastroenterology, University of Michigan, Ann Arbor, Michigan, USA
[3]Center for Global Health Equity, University of Michigan, Ann Arbor, Michigan, USA
[4]Michigan Integrated Center for Health Analytics and Medical Prediction (MiCHAMP), University of Michigan, Ann Arbor, Michigan, USA
[5]Institute for Human Development, The Aga Khan University, Nairobi, Kenya
[6]Department of Population Health, The Aga Khan University, Nairobi, Kenya
[7]Eck Institute for Global Health, University of Notre Dame, South Bend, Indiana, USA
[8]Center for Research Computing, University of Notre Dame, South Bend, Indiana, USA
[9]Department of Surgery, Division of Colorectal Surgery, University of Michigan, Ann Arbor, Michigan, USA
[10]Harold C. Simmons Comprehensive Cancer Center, The University of Texas Southwestern Medical Center, Dallas, Texas, USA
[11]Department of Internal Medicine, Division of Digestive and Liver Diseases, The University of Texas Southwestern Medical Center, Dallas, Texas, USA
[12]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan, USA
[13]Department of Biomedical Engineering, University of Michigan, Ann Arbor, Michigan, USA
[14]Department of Radiation Oncology, University of Michigan, Ann Arbor, Michigan, USA
[15]Department of Pathology, University of Michigan Health System, Ann Arbor, Michigan, USA
[16]Department of Medicine, Aga Khan University Hospital Nairobi, Nairobi, Kenya
[17]Department of Statistics, University of Michigan, Ann Arbor, Michigan, USA
[18]O'Neal Comprehensive Cancer Center, The University of Alabama at Birmingham, Birmingham, Alabama, USA
[19]Department of Hematology-Oncology, Aga Khan University Hospital Nairobi, Nairobi, Kenya

Correspondence to Dr Akbar K Waljee, Veterans Affairs Center for Clinical Management Research, Ann Arbor, Michigan, USA; awaljee@med.umich.edu

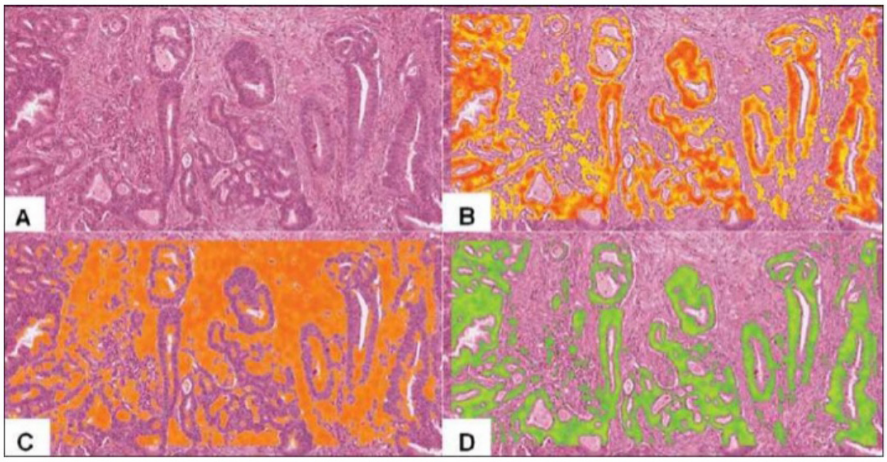

Figure 1 Identification of colon cancer in a digital H&E-stained tissue section of colonic adenocarcinoma. (A) image of colon cancer from a digital slide. (B) a vector was created to identify only the malignant glands, and (C) an additional vector was created to recognise only the stroma. (D) Boolean logic was used to determine the malignant glands, and the stroma was subtracted out. This approach could assist pathologists in identifying small foci of invasive glands or small foci of tumour present in blood and lymphatic vessels, which might be otherwise overlooked. Figure copyright Hipp *et al*,[50] licensed under CC-BY 2.0 (https://creativecommons.org/licenses/by/2.0/).

access to evidence-based tools for screening and early detection. These limitations are understandable given the burden of infectious diseases that has historically afflicted SSA. However, as SSA experiences the epidemiological shift from infectious diseases to non-communicable diseases, such as CRC, the aetiology of the disease and the solutions to address the emerging CRC epidemic require SSA-specific data and approaches. Extrapolation of cancer screening recommendations from high-income countries to SSA is often inappropriate due to differences in demographics, disease epidemiology and resources. For example, the average risk screening for CRC is typically recommended at age 50; however, the US Preventive Services Task Force, the American Cancer Society and the US Multisociety Task Force on Colorectal Cancer have recently recommended lowering it to 45 years.[16–18] In SSA, estimates from available data indicate that 19%–38% of CRC diagnoses are in persons <40 years of age—a stark contrast to the 1%–7% reported in high-income countries.[19–21] The higher risk of the early development of CRC, coupled with the recent lowering in screening age, highlight the evolving epidemiology of CRC in younger adults and the need to tailor screening approaches to capture this cohort, particularly in SSA. There is urgency to address this need given that Africa's population is projected to double by 2050, reaching nearly 2.5 billion (23% of the global population) with more than half of its population <25 years of age.[22]

Currently, several modalities exist for CRC screening and early detection. Colonoscopy can be used for CRC detection and intervention (eg, polyp removal), but SSA has limited endoscopic services. Recent data from Mwachiro *et al*[23] reported an overall endoscopy capacity in East Africa of 1.2 endoscopists, 1.2 gastroscopes and 0.9 colonoscopes per 100 000 people—values 1% to 10% of that of resource-rich countries. Non-invasive screening tests include faecal occult blood testing, FIT and stool-based DNA tests[17 24]; however, widespread adoption of stool-based approaches remains suboptimal in both high-income countries as well as SSA.[25–28] In addition, questions about the impact of high ambient temperature and endemic parasitic infection as well as the practicality and cost-effectiveness of these approaches in SSA remain.[29–31] Regardless, endoscopy is still needed for diagnosis and prognosis. Thus, early detection strategies that target those at the highest risk benefit from these limited services are paramount. With growing investments in technologies (eg, electronic health records and cloud computing) in SSA, the existing and expanding infrastructure can be leveraged to employ novel AI/ML methods to develop and validate surveillance tools that identify populations at highest risk for CRC in a more individualised or precise manner, as described below.

AI AND ML APPROACHES

MAAA as a population-based surveillance and early detection tool

Laboratory studies, such as complete blood counts (CBC) and comprehensive metabolic panels (CMP), are standard diagnostic tests ordered by clinicians, even in LMICs. These tests often contain subtle diagnostic clues; however, interpretation of laboratory studies is routinely subject to human error. Presymptomatic longitudinal CBC patterns may be imperceptible to clinicians but would be readily detectable by statistical algorithms or 'prediction models,' often referred to as Multianalyte Assays with Algorithmic Analysis (MAAA).[32] Currently, proprietary MAAA exist that were built and validated in high-income countries; these MAAA use CBC and demographic data to identify patients at high risk of CRC.[33–36] Similarly, we have developed a MAAA prediction model in a US cohort using longitudinal and single timepoint laboratory studies and patient characteristics (accepted to Digestive Disease Week 2022). Initially, we set out to develop and compare multiple MAAA to predict luminal GI tract cancers in a retrospective cohort of patients (n=148 158 with 1025 GI tract cancers) who had at least 2 CBCs within 2 years. Predictor variables included age, sex, race, body mass index, individual components of the CBC and the CMP. To incorporate longitudinal features, summary statistics were calculated for each subject's particular part of the CBC (ie, maximum, minimum, slope and total variation). Data were split into 70% training and 30% validation sets for analysis. For the 3-year prediction of GI tract cancers, the longitudinal random forest model performed the best with an area under the receiver operator curve (AUROC) of 0.750 (95% CI 0.729 to 0.771) and Brier score of 0.116, compared with the longitudinal logistic regression with an AUROC of 0.735 (95% CI 0.713 to 0.757) and Brier score of 0.205. The longitudinal logistic regression and random forest models outperformed the single timepoint logistic regression at 3 years, with an AUROC of 0.683 (95% CI 0.665 to 0.701). These findings are limited in that the MAAA predicts GI tract cancer, not CRC specifically, although just over half of patients with GI tract cancers had CRC (53.5%, n=548/1025). To date, this approach has not been validated in a low resource setting or SSA, where demographics and disease aetiology may differ, and longitudinal laboratory studies may not be readily available. In addition, CBC and CMP baselines likely vary across

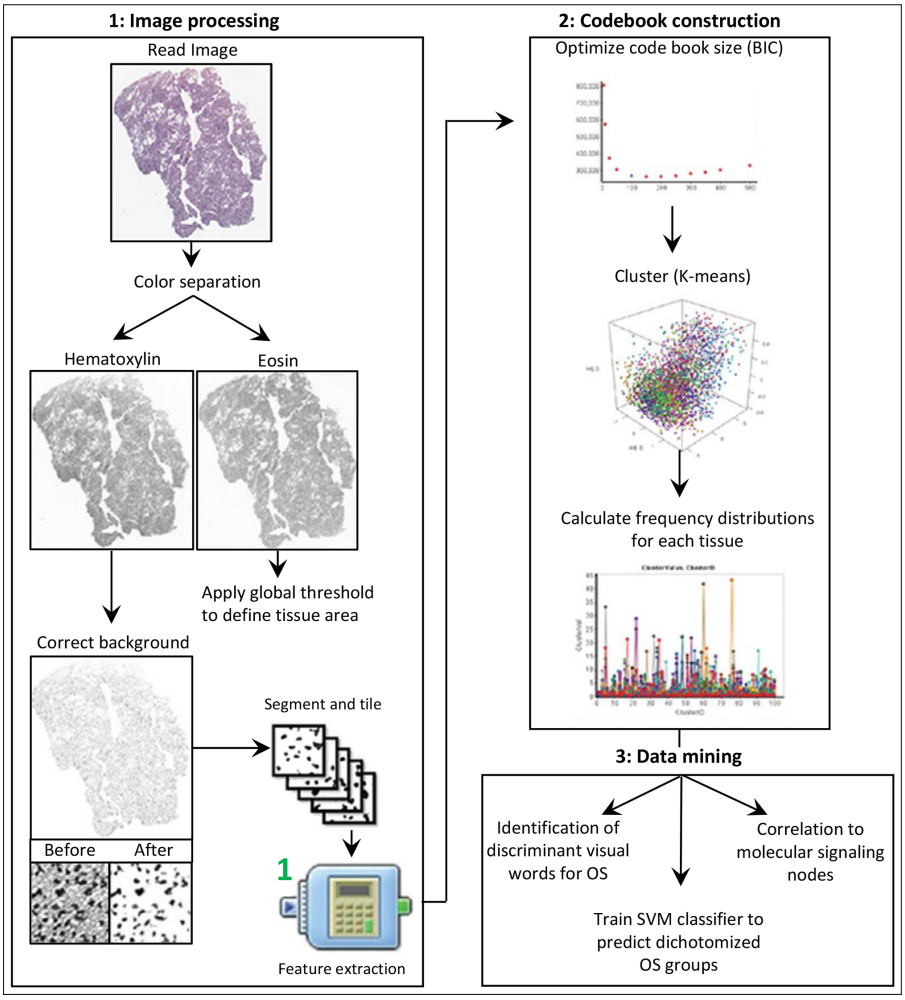

Figure 2 Summary of analysis workflow for identifying histological determinants of malignant transformation and disease grade. Step 3 uses a support vector machine classifier, but any classifier can be used (eg, random forest). Figure copyright Powell *et al*,[52] licensed under CC-BY 2.0 (https://creativecommons.org/licenses/by/2.0/). BIC, Bayesian information criterion; OS, overall survival; SVM, support vector machine.

genetically diverse populations and can be influenced by the prevalence of infectious and chronic conditions, including other malignancies and genetic conditions such as sickle-cell disease that have distinct prevalence in different populations. While these previous studies provide proof of concept for the development of MAAA for CRC screening in SSA, it would be essential to develop and compare models that incorporate longitudinal and cross-sectional laboratory data to determine the performance and optimal specificity or sensitivity for the target populations.

CRC is particularly amenable to MAAA-guided early detection strategies for multiple reasons. First, CRC is highly curable when diagnosed at an early stage.[6] Second, CRC is highly vascular and can produce very subtle chronic occult blood loss, which could be detected before symptoms develop using data from routine longitudinal CBCs and CMP and ML-based methods.[34 35 37] Indeed, patients in SSA tend to present with late-stage CRC, being diagnosed after clinical presentation with symptoms.[6 7] Third, MAAA can be tailored to the local needs. For example, positive predicted values and negative predicted values of MAAAs can be adjusted to maximise sensitivity or specificity based on target populations (eg, age groups), resource availability, or sequential testing approaches (eg, MAAA and then FIT). Similar work has been done in other settings where resources were limited, particularly during COVID-19, where FIT-based quantitative screening thresholds were used to direct patients for endoscopic services.[38 39] Finally, the costs and resources required to the patient and healthcare facility/provider are significantly less since it uses routine labs collected in various clinical settings. This is particularly relevant given the significant improvements in laboratory medicine in Africa driven by efforts to combat HIV/AIDS. In addition to developing a competent workforce and innovative quality improvement programmes that saw more than 1100 laboratories enrolled and 44 accredited to international standards, several regional laboratory networks have also been established to support programme scale-up and disease surveillance.[40] This infrastructure can support robust healthcare systems and combat emerging continental and global health threats, like CRC and other cancers. Although, despite available diagnostic testing, studies have shown that they are not optimally used in managing patient care, and tools to bridge the diagnostic–treatment divide are needed.[41 42] MAAAs offer one approach to help bridge this gap and can be coupled with simple paper-based tools (eg, nomograms) to more complex mobile app-based tools or lightweight, field-deployed (cloud-based) Laboratory Information Systems designed for use in LMICs.[43 44] In addition to the use of MAAAs as a tool for CRC diagnosis, the approach could be adapted for the prediction of CRC prognosis and treatment outcomes as both the CBC and CMP profiles of patients have been associated with disease stage, metastasis and treatment outcomes.[45–47]

AI-based algorithms in pathology for early diagnosis and prognosis

After screening, accurate and timely diagnosis is critical to identifying appropriate treatment plans in cancer management. While CRC is diagnosed via clinicopathological assessment by a pathologist, the availability of such expertise and resources in SSA are minimal. A 2012 survey of 33 African countries found that 31 (94%) had fewer than one pathologist for every 500 000 people, and many had fewer than one pathologist for every 1 million people.[48] These values are 10% in high-income countries, like the USA, which had one pathologist for every 20 600 people in 2010. In addition to the lack of trained pathologists, access to immunohistochemical (IHC) reagents required for accurate and definitive diagnosis remains a significant hurdle. Unlike in infectious diseases, H&E-stained slides do not often suffice to make a precise diagnosis. Thus, a lack of efficient and reliable pathology services leads to delays and inaccurate reporting of results, which contributes to patients receiving inappropriate treatment. Patients may be prescribed medications that are expensive yet ineffective and sometimes even harmful in treating their

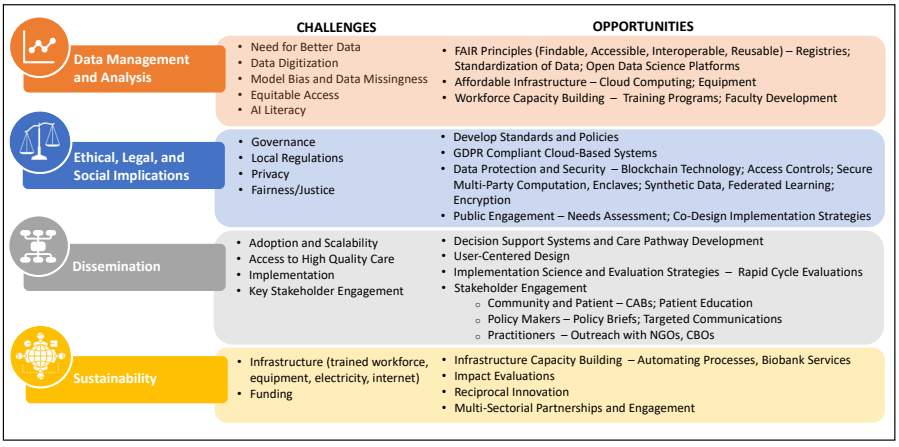

Figure 3 Examples of challenges and opportunities for leveraging AI-based approaches in sub-Saharan Africa. AI, artificial intelligence; CAB, community advisory board; CBO, community-based organisation; NGO, non-governmental organisations.

cancer type. Recent advances in AI-based computer vision and pattern recognition algorithms that use routine H&E-stained whole slide imaging offer transformative tools well suited for early cancer diagnosis and prognosis in SSA.

Pattern recognition algorithms aim to detect abnormalities in cell and tissue samples faster, more accurately and more consistently. In clinical care, these tools can assist pathologists in diagnostic recommendations by pre-screening an image and identifying potentially problematic areas, including subtle features that may not be readily apparent to the eye. For example, the VIPR (Vectorising spatially-Invariant Pattern Recognition) algorithm and software is a fully operational application suite developed by the Data Visualisation Core of the National Institute of Diabetes and Digestive and Kidney Diseases' Kidney Precision Medicine Project.[49–51] VIPR uses semisupervised and unsupervised, pixel-level classification of digital whole slide image content, which allows for extremely high-throughput analysis of entire libraries of whole slide imagery. VIPR differs from conventional pattern recognition software by basing its core search on a series of concentric, pattern-matching rings rather than the more typical rectangular or square blocks. This approach takes advantage of the continuous symmetry of the rings, allowing for the recognition of features independent of rotation. By making use of massively parallel computational platforms to realise necessary speed and performance, VIPR performs direct integration of vectorised image data with other classes of patient data (eg, lab values, clinical phenotypic features, clinical course, outcomes), thus allowing for a more global assessment of health status and biological potential of any given malignancy. The pixel-level precision and consistency for whole slide image classification exceeds what is possible using subject matter expertise alone. Moreover, it has demonstrated high reproducibility across different fields of view of a single slide, different slides in the same case, and different cases entirely.[49–51]

The VIPR tool was initially developed to interrogate tissue from patients with acute kidney injury or chronic kidney disease to define disease subgroups and identify critical cells, pathways, and targets for novel therapies. It has since proved to be highly effective for cancer detection and classification in colon cancer (figure 1) as well as haematology, breast cancer and lymphoma.[49–51] Because VIPR has been designed as a turn-key system for automated objective assessment of H&E slides for disease diagnosis, it is suitable for deployment in settings where pathologists alone can effectively incorporate the tool into clinical workflow, without the need for the immediate response from an image analysis expert. Once histologically distinct regions are identified or 'prescreened,' image analysis algorithms can then be used to mine individual regions and aggregate them to predict malignant transformation, as described below.

Following disease detection from histopathology, disease grading and IHC classification is critical to classifying various subtypes of cancer and thus determining appropriate treatment. Another rapidly advancing area is the use of computer vision and deep learning to digitally phenotype histological slides to better understand treatment response and survival.[52] These algorithms can complement the clinical interpretation of diseased tissue in which the underlying diagnosis has already been made. This approach was employed in an image analysis and data mining pipeline to identify histological features capable of differentiating between cancer and non-cancer lesions and the malignant transformation-potential in gliomas (figure 2).[52] Using whole slide imaging data from the Cancer Genome Atlas and companion clinical data for these specimens, we assessed the prognostic relevance of these histological discriminants.[53 54] Histopathology image-derived measurements, such as cell morphologies, spatial patterns of cellular organisation, in combination with a bag-of-words (BoW) approach[53 55] was used to identify tissue subregions that have visually distinct properties (eg, nuclear morphology, patterns of spatial organisation) and were associated with time-to-malignant transformation. The BoW approach is akin to clustering image subregions (ie, patches) derived from the whole slide image of the tissue. Importantly, this dictionary achieved an AUROC (through cross-validation) of 0.76 to discriminate surrogates of malignant transformation. While this approach was developed in glioma, it offers one potential strategy to incorporate image features derived from routine H&E-stained slides into prognostic, predictive models of other cancers, such as CRC. In addition to the above approach, deep learning algorithms leveraging popular architectures, such as Reset, VGGNet, and Inception, are also being adopted in the context of cancer prognosis,[56 57] providing a path to a 'non-feature-engineering' approach to image recognition and content mining. In tandem with recently developed methods around feature interpretability,[58] these tools can be incorporated into clinical workflows. It is worth noting that modern computer vision techniques aim to adjust for multiple biases in data acquisition, image staining and related artefacts, contributing to the development and delivery of robust decision support algorithms.

Digital pathology lab systems and infrastructure are becoming more obtainable in SSA. For example, the VIPR Software is open-source, and microscopes that are small and fully remote-operable, capable of high-resolution images have become more affordable. Also, while these technologies can be computationally expensive (ie, requiring graphical processing unit and storage for gigapixel histopathological

scans), the emergence of cloud computing in SSA can transform innovation and efficiency around how data are used. Taken together, one could envision an analytical pipeline that couples operational pattern recognition tools with image analysis algorithms for automated and democratised identification and prediction of CRC from routine H&E histology images that is scalable. The current development of data science collaboratives in Africa could also facilitate the adoption and deployment of these tools, as well as MAAA-guided models for early detection and diagnosis of CRC as outlined below.

FUTURE DIRECTIONS

In the era of value-based healthcare, AI/ML provides opportunities to improve access to care, reduce wastage, optimise resource utilisation and provide a mechanism for quality assurance of healthcare regarding CRC screening, diagnosis and management. Funding agencies (government, donors or commercial) are more likely to invest in a system whose outputs are easily measured and can be bench marked against available resources. This is particularly important in SSA were data driven management of healthcare delivery is still a challenge. Routine use of AI/ML tools and their dissemination remains rare in high-income countries, not to mention LMICs. Advances in model performance characteristics have accelerated, but despite performing well in silo using retrospective data in a research setting, prediction models (ie, using logistic regression or AI/ML-based methods) rarely leave the exploratory domain for use in the clinical or community settings. The development and deployment of AI/ML-based tools in SSA require addressing existing limitations in computing infrastructures and a lack of local data needed to support the creation of effective models. However, solving these problems will not automatically lead to widespread adoption. If we do not directly address the challenges of dissemination and adoption of these prediction models in a way that supports social justice and health equity, data science approaches will have minimal impact on the health of individuals and populations. The issues surrounding the development, deployment and adoption of AI/ML-based tools in LMICs, and SSA, have been extensively described elsewhere.[59–61] Examples of some of the challenges and opportunities for leveraging AI-based approaches in SSA are provided in figure 3.

To address these challenges and increase the capacity to use and develop data science approaches in health research and innovation in Africa, the National Institutes of Health (NIH) recently launched a new Common Fund Programme: Harnessing Data Science for Health Discovery and Innovation in Africa (DS-I Africa).[11] DS-I Africa builds on prior investments by the NIH Common Fund and its partners in the Medical Education Partnership Initiatives and the Human Health and Heredity in Africa (H3Africa) consortium to form a unique continental ecosystem that could be transformative, leveraging existing expertise to develop data tools and applications that can be shared, adopted, and harmonised globally. Creating a robust network of partnerships across the African continent and in the USA, including numerous national health ministries, non-governmental organisations, corporations and other academic institutions, the DS-I consortium includes seven research hubs (all of which are led by African institutions), seven research training programmes, four projects focused on ethical, legal and social implications of data science, and an open data science platform and coordinating centre.

Figure 4 depicts the synergistic initiatives within the DS-I Africa Consortium and highlights one of the research hubs to demonstrate how the hub aims to function as a scalable and sustainable data science platform in Kenya and within the greater DS-I consortium. The exemplar hub, UtiliZing Health Information for Meaningful Impact in East Africa Through Data Science (UZIMA-DS), will address three critical needs across the translational

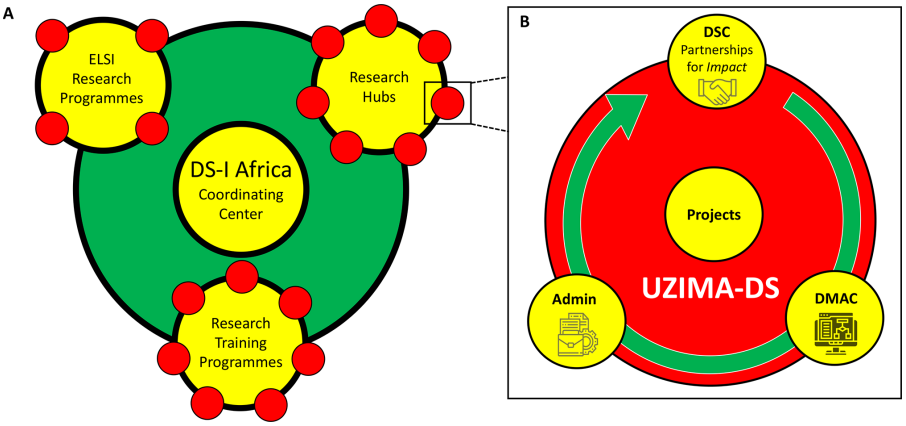

Figure 4 Depiction of the Harnessing Data Science for Health Discovery and Innovation in Africa (DS-I Africa) programme and exemplar research hub. (A) The four main initiatives are: (1) Research hubs will apply novel approaches to data analysis and artificial intelligence to address critical health issues in Africa. (2) Open data science platform and coordinating centre will provide a flexible, scalable platform for the DS-I Africa researchers to find and access data, select tools and workflows, and run analyses through collaborative workspaces. It will also deliver the organisational framework for the direction and management of the initiative's common activities; (3) Research training programmes will create multi-tiered curricula to build skills in foundational health data science, with options ranging from master's and doctoral degree tracks to postdoctoral training and faculty development; and (4) The ethical, legal and social implications (ELSI) projects will address data science issues that present challenges in Africa such as data privacy and ownership, cybersecurity and sensitivities concerning the use of geospatial information for research or public health surveillance. (B) Led by the Aga Khan University—East Africa, Kenya Medical Research Institute-Wellcome Trust Research Programme, and the University of Michigan, the research hub will implement two research projects around maternal, newborn and child health as well as mental health, which will be supported by three cores: Admin core, Data Management and Analysis Core (DMAC) and Dissemination and Sustainability Core (DSC). The Admin Core will lead the UZIMA-DS researchhub, fostering synergy and integration of all hub components and partnerships and facilitating participation in DS-I cross-consortium activities. The DMAC will employ FAIR (Findable, Accessible, Interoperable, Reusable) principles to support the hub's data ecosystem through data governance, facilitating data analytics within the projects, and fostering data sharing and interoperability throughout the greater DS-I Africa consortium. The DSC will promote engagement with stakeholders to identify sustainable model dissemination pathways into target communities. Through multisectoral partnerships with government, healthcare and non-profit sectors, the core will: facilitate the development of best practices and policies with stakeholders using data-driven approaches to inform guidelines; and promote engagement with private sectors to explore sustainable commercialisation opportunities and pathways. UZIMA-DS, UtiliZing Health Information for Meaningful Impact in East Africa Through Data Science.

spectrum of data science: (1) harmonisation of multimodal data sources; (2) leveraging temporal patterns of data to identify trajectories through prediction modelling using AI/ML-based methods; and (3) engaging with key stakeholders to identify pathways for dissemination and sustainability of these models in target communities. While the initial health domains of UZIMA-DS address critical health issues in maternal, newborn and child health and mental health, the hub can serve as a model that can be scaled to other countries and health domains within the greater DS-I consortium.

Lastly, while global health research was traditionally characterised by a unidirectional exchange of innovation and expertise from high-income countries to LMICs, it is now well-recognised that these collaborations have 'reciprocal value'. Because necessity often drives innovation, health tools that have been researched, developed, and implemented in LMICs can be adapted and adopted to address similar challenges in the USA and other high-income countries through 'reverse innovation'.[13–15] While empirically this is a nascent field, some early successes have been highlighted in areas such as antiretroviral treatment for HIV, cognitive impairment in older adults and mental health.[62–65] Given the growing investments in data science infrastructure, the demonstrated openness to embracing technological change (ie, mobile banking proliferation), and the urgent need to develop more efficient approaches to cancer screening and early detection that do not rely heavily on trained healthcare personnel or specialised resources (eg, endoscopy, pathology), SSA is well poised to drive innovative AI-based solutions to augment the utilisation of specialised resources across the globe.

SUMMARY

With the growing resources and investments in AI/ML-based tools in SSA, one could envision a CRC surveillance and diagnosis pipeline that employs MAAA for population-based surveillance and pattern recognition and computer vision algorithms to guide diagnostic recommendations and prognosis. These tools will need to be tailored to local needs based on available resources and testing approaches (eg, sequential testing with MAAA and then FIT) and key stakeholders will need to engage in the codesign of widespread implementation strategies (eg, community-based screening programmes, practitioner education, health policies). Future studies are required to compare the efficacy of these tools to existing CRC surveillance and diagnosis tools (eg, FIT) in SSA populations. Furthermore, these innovative solutions provide opportunities for the adaption and adoption of these approaches in high-income countries. While CRC was used as the use case, these tools could be expanded to other prevalent and emergent cancers (eg, liver, breast and cervical) or other non-communicable diseases that would benefit from lab-based MAAA and computer vision AI-based methods for automated objective assessment of disease diagnosis and prognosis.

Twitter Akbar K Waljee @AkbarWaljee and Ulysses Balis @ulyssesbalis

Contributors All authors were involved in manuscript writing and gave approval of the final version.

Funding Research reported in this publication was supported by the Office Of The Director, National Institutes Of Health (OD), the National Institute Of Biomedical Imaging And Bioengineering (NIBIB), the National Institute Of Mental Health (NIMH) and the Fogarty International Centre (FIC) of the National Institutes of Health under award number U54TW012089 (AA and AW).

Competing interests AGS has consulted for and received research funding from Exact Sciences. AR serves as member for Voxel Analytics and consults for Genophyll and Pact&Health. GHS is a founder of Anza Biotechnologies.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

AKW and EMW-H are joint first authors.

To cite Waljee AK, Weinheimer-Haus EM, Abubakar A, *et al*. *Gut* 2022;**71**:1259–1265.

Received 17 February 2022
Accepted 17 March 2022
Published Online First 13 April 2022

Gut 2022;**71**:1259–1265.
doi:10.1136/gutjnl-2022-327211

ORCID iDs
Akbar K Waljee http://orcid.org/0000-0003-1964-8790
Eileen M Weinheimer-Haus http://orcid.org/0000-0001-5408-0024
Geoffrey H Siwo http://orcid.org/0000-0003-0726-997X
Amit G Singal http://orcid.org/0000-0002-1172-3971
Arvind Rao http://orcid.org/0000-0002-9613-426X
Andrew J Read http://orcid.org/0000-0001-7336-8496
Ulysses Balis http://orcid.org/0000-0002-4168-1477
Christopher K Opio http://orcid.org/0000-0001-8898-6350

REFERENCES

1 Sung H, Ferlay J, Siegel RL, *et al*. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2021;71:209–49.
2 Parker RK, Mwachiro MM, Ranketi SS, *et al*. Curative surgery improves survival for colorectal cancer in rural Kenya. *World J Surg* 2020;44:30–6.
3 Saluja S, Alatise OI, Adewale A, *et al*. A comparison of colorectal cancer in Nigerian and North American patients: is the cancer biology different? *Surgery* 2014;156:305–10.
4 Asombang AW, Madsen R, Simuyandi M, *et al*. Descriptive analysis of colorectal cancer in Zambia, southern Africa using the National cancer disease Hospital database. *Pan Afr Med J* 2018;30:248.
5 Gullickson C, Goodman M, Joko-Fru YW, *et al*. Colorectal cancer survival in sub-Saharan Africa by age, stage at diagnosis and human development index: a population-based registry study. *Int J Cancer* 2021;149:1553–63.
6 Brenner H, Kloor M, Pox CP. Colorectal cancer. *Lancet* 2014;383:1490–502.
7 Rawla P, Sunkara T, Barsouk A. Epidemiology of colorectal cancer: incidence, mortality, survival, and risk factors. *Prz Gastroenterol* 2019;14:89–103.
8 Holst C, Sukums F, Radovanovic D, *et al*. Sub-Saharan Africa-the new breeding ground for global digital health. *Lancet Digit Health* 2020;2:e160–2.
9 Suri T, Jack W. The long-run poverty and gender impacts of mobile money. *Science* 2016;354:1288–92.
10 United nations. Resource guide on artificial intelligence (AI) strategies, 2021. Available: https://sdgs.un.org/sites/default/files/2021-06/Resource%20Guide%20on%20AI%20Strategies_June%202021.pdf
11 NIH awards nearly $75M to catalyze data science research in Africa, 2021. Available: https://www.nih.gov/news-events/news-releases/nih-awards-nearly-75m-catalyze-data-science-research-africa [Accessed 20 Dec 2021].
12 Gouda HN, Charlson F, Sorsdahl K, *et al*. Burden of non-communicable diseases in sub-Saharan Africa, 1990-2017: results from the global burden of disease study 2017. *Lancet Glob Health* 2019;7:e1375–87.
13 Wigle JM, Akseer N, Carbone S, *et al*. Developing a tool to measure the reciprocal benefits that accrue to health professionals involved in global health. *BMJ Glob Health* 2018;3:e000792.
14 Harris M, Dadwal V, Syed SB. Review of the reverse innovation series in globalization and health - where are we and what else is needed? *Global Health* 2020;16:26.
15 Anderson F, Donkor P, de Vries R, *et al*. Creating a charter of collaboration for international university partnerships: the Elmina Declaration for human resources for health. *Acad Med* 2014;89:1125–32.
16 Knudsen AB, Rutter CM, Peterse EFP, *et al*. Colorectal cancer screening: an updated modeling study for the US preventive services Task force. *JAMA* 2021;325:1998–2011.
17 Wolf AMD, Fontham ETH, Church TR, *et al*. Colorectal cancer screening for average-risk adults: 2018 guideline update from the American cancer Society. *CA Cancer J Clin* 2018;68:250–81.
18 Patel SG, May FP, Anderson JC, *et al*. Updates on age to start and stop colorectal cancer screening:

recommendations from the U.S. Multi-Society Task force on colorectal cancer. *Gastroenterology* 2022;162:285–99.
19 Abdulkarim A, Barasa M. Clinicopathological profile of colorectal cancer at a tertiary hospital in Kenya. *Br J Surg* 2021;108:e137.
20 Parker RK, Ranketi SS, McNelly C, *et al*. Colorectal cancer is increasing in rural Kenya: challenges and perspectives. *Gastrointest Endosc* 2019;89:1234–7.
21 Katsidzira L, Gangaidzo I, Thomson S, *et al*. The shifting epidemiology of colorectal cancer in sub-Saharan Africa. *Lancet Gastroenterol Hepatol* 2017;2:377–83.
22 Hajjar B. The children's continent: keeping up with Africa's growth. Available: https://wwwweforumorg/agenda/2020/01/the-children-s-continent/
23 Mwachiro MM, Topazian H, Lenga G, *et al*. Tu1965 gastrointestinal endoscopy capacity in East Africa: a multinational survey. *Gastroenterology* 2020;158:S-1235–S-1236.
24 Ebner DW, Kisiel JB. Stool-Based tests for colorectal cancer screening: performance benchmarks lead to high expected efficacy. *Curr Gastroenterol Rep* 2020;22:32.
25 Siegel RL, Miller KD, Goding Sauer A, *et al*. Colorectal cancer statistics, 2020. *CA Cancer J Clin* 2020;70:145–64.
26 Laiyemo AO, Brawley O, Irabor D, *et al*. Toward colorectal cancer control in Africa. *Int J Cancer* 2016;138:1033–4.
27 Lussiez A, Dualeh SHA, Dally CK, *et al*. Colorectal cancer screening in Ghana: physicians' practices and perceived barriers. *World J Surg* 2021;45:390–403.
28 Murphy CC, Sen A, Watson B, *et al*. A systematic review of repeat fecal occult blood tests for colorectal cancer screening. *Cancer Epidemiol Biomarkers Prev* 2020;29:278–87.
29 Knapp GC, Sharma A, Olopade B, *et al*. An exploratory analysis of fecal immunochemical test performance for colorectal cancer screening in Nigeria. *World J Surg* 2019;43:2674–80.
30 Doubeni CA, Jensen CD, Fedewa SA, *et al*. Fecal immunochemical test (fit) for colon cancer screening: variable performance with ambient temperature. *J Am Board Fam Med* 2016;29:672–81.
31 Knapp GC, Alatise O, Olopade B, *et al*. Feasibility and performance of the fecal immunochemical test (fit) for average-risk colorectal cancer screening in Nigeria. *PLoS One* 2021;16:e0243587.
32 Colón-Franco JM, Bossuyt PMM, Algeciras-Schimnich A, *et al*. Current and emerging multianalyte assays with algorithmic Analyses-Are laboratories ready for clinical adoption? *Clin Chem* 2018;64:885–91.
33 Schneider JL, Layefsky E, Udaltsova N, *et al*. Validation of an algorithm to identify patients at risk for colorectal cancer based on laboratory test and demographic data in diverse, community-based population. *Clin Gastroenterol Hepatol* 2020;18:2734–41.
34 Hornbrook MC, Goshen R, Choman E, *et al*. Early colorectal cancer detected by machine learning model using gender, age, and complete blood count data. *Dig Dis Sci* 2017;62:2719–27.
35 Kinar Y, Kalkstein N, Akiva P, *et al*. Development and validation of a predictive model for detection of colorectal cancer in primary care by analysis of complete blood counts: a binational retrospective study. *J Am Med Inform Assoc* 2016;23:879–90.
36 Ayling RM, Wong A, Cotter F. Use of ColonFlag score for prioritisation of endoscopy in colorectal cancer. *BMJ Open Gastroenterol* 2021;8.
37 Usher-Smith JA, Walter FM, Emery JD, *et al*. Risk prediction models for colorectal cancer: a systematic review. *Cancer Prev Res* 2016;9:13–26.
38 Loveday C, Sud A, Jones ME, *et al*. Prioritisation by fit to mitigate the impact of delays in the 2-week wait colorectal cancer referral pathway during the COVID-19 pandemic: a UK modelling study. *Gut* 2021;70:1053–60.
39 D'Souza N, Georgiou Delisle T, Chen M, *et al*. Faecal immunochemical test is superior to symptoms in predicting pathology in patients with suspected colorectal cancer symptoms referred on a 2WW pathway: a diagnostic accuracy study. *Gut* 2021;70:1130–8.
40 Nkengasong JN, Mbopi-Keou F-X, Peeling RW, *et al*. Laboratory medicine in Africa since 2008: then, now, and the future. *Lancet Infect Dis* 2018;18:e362–7.
41 Gibb J, Chitsulo J, Chipungu C, *et al*. Supporting quality data systems: lessons learned from early implementation of routine viral load monitoring at a large clinic in Lilongwe, Malawi. *J Clin Res HIV AIDS Prev* 2017;3. doi:10.14302/issn.2324-7339.jcrhap-17-1468. [Epub ahead of print: 14 03 2017].
42 Nkengasong JN. The diagnostic-clinical chasm: work in progress? *Afr J Lab Med* 2016;5:586.
43 Choo M, Hoy GE, Dugan SP, *et al*. Imputing HbA1c from capillary blood glucose levels in patients with type 2 diabetes in Sri Lanka: a cross-sectional study. *BMJ Open* 2020;10:e038148.
44 Saha AK, Gunaratnam N, Patil R, *et al*. A new model for diabetes-focused capacity building - lessons from Sri Lanka. *Clin Diabetes Endocrinol* 2018;4:22.
45 Galizia G, Lieto E, Zamboli A, *et al*. Neutrophil to lymphocyte ratio is a strong predictor of tumor recurrence in early colon cancers: a propensity score-matched analysis. *Surgery* 2015;158:112–20.
46 Song Y, Huang Z, Kang Y, *et al*. Clinical usefulness and prognostic value of red cell distribution width in colorectal cancer. *Biomed Res Int* 2018;2018:9858943.
47 Jia W, Yuan L, Ni H, *et al*. Prognostic value of platelet-to-lymphocyte ratio, neutrophil-to-lymphocyte ratio, and Lymphocyte-to-White blood cell ratio in colorectal cancer patients who received neoadjuvant chemotherapy. *Technol Cancer Res Treat* 2021;20:15330338211034291.
48 Fleming K. Pathology and cancer in Africa. *Ecancermedicalscience* 2019;13:945.
49 Hipp JD, Cheng J, Hanson JC, *et al*. SIVQ-LCM protocol for the ArcturusXT instrument. *J Vis Exp* 2014. doi:10.3791/51662. [Epub ahead of print: 23 Jul 2014].
50 Hipp JD, Cheng JY, Toner M, *et al*. Spatially invariant vector quantization: a pattern matching algorithm for multiple classes of image subject matter including pathology. *J Pathol Inform* 2011;2:13.
51 El-Achkar TM, Eadon MT, Menon R, *et al*. A multimodal and integrated approach to interrogate human kidney biopsies with rigor and reproducibility: guidelines from the kidney precision medicine project. *Physiol Genomics* 2021;53:1–11.
52 Powell RT, Olar A, Narang S, *et al*. Identification of Histological Correlates of Overall Survival in Lower Grade Gliomas Using a Bag-of-words Paradigm: A Preliminary Analysis Based on Hematoxylin & Eosin Stained Slides from the Lower Grade Glioma Cohort of The Cancer Genome Atlas. *J Pathol Inform* 2017;8:9.
53 Li X, Monga V, Rao UKA, . Analysis-Synthesis learning with shared features: algorithms for histology image classification. *IEEE Trans Biomed Eng* 2020;67:1061–73.
54 Mousavi HS, Monga V, Rao G, *et al*. Automated discrimination of lower and higher grade gliomas based on histopathological image analysis. *J Pathol Inform* 2015;6:15.
55 Özdemir E, Sökmensüer C, G-D Ç. Histopathological image classification with the bag of words model. 2011 IEEE 19th Signal Processing and Communications Applications Conference (SIU), 2011:634–7.
56 Kleppe A, Skrede O-J, De Raedt S, *et al*. Designing deep learning studies in cancer diagnostics. *Nat Rev Cancer* 2021;21:199–211.
57 Zhu W, Xie L, Han J, *et al*. The application of deep learning in cancer prognosis prediction. *Cancers* 2020;12:603.
58 Diao JA, Wang JK, Chui WF, *et al*. Human-interpretable image features derived from densely mapped cancer pathology slides predict diverse molecular phenotypes. *Nat Commun* 2021;12:1613.
59 Hosny A, Aerts HJWL. Artificial intelligence for global health. *Science* 2019;366:955–6.
60 The Lancet. Artificial intelligence in global health: a brave new world. *Lancet* 2019;393:1478.
61 USAID center for innovation and impact (CII) series. Artificial intelligence in global health: defining a collective path forward, 2022. USAID. Available: https://www.usaid.gov/sites/default/files/documents/1864/AI-in-Global-Health_webFinal_508.pdf
62 Chibanda D. Reducing the treatment gap for mental, neurological and substance use disorders in Africa: lessons from the Friendship bench in Zimbabwe. *Epidemiol Psychiatr Sci* 2017;26:342–7.
63 Malik R, Weiss EF, Gottesman R, *et al*. Picture-Based memory impairment screen: effective cognitive screen in ethnically diverse populations. *J Am Geriatr Soc* 2018;66:1598–602.
64 Rao D, Desmond M, Andrasik M, *et al*. Feasibility, acceptability, and preliminary efficacy of the unity workshop: an internalized stigma reduction intervention for African American women living with HIV. *AIDS Patient Care STDS* 2012;26:614–20.
65 NIH Fogarty International center. Tech designed for Africa helps us fight disease, save money, 2017. Available: https://www.fic.nih.gov/News/GlobalHealthMatters/september-october-2017/Pages/emocha-mobile-app.aspx [Accessed 02 Dec 2022].