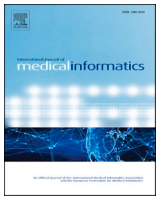
International Journal of Medical Informatics 179 (2023) 105244

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf



Review article

Machine learning prediction models for clinical management of blood-borne viral infections: a systematic review of current applications and future impact

Busayo I. Ajuwon [a,b,*], Oluwatosin N. Awotundun [c], Alice Richardson [d], Katrina Roper [a], Meru Sheel [e], Nurudeen Rahman [f], Abideen Salako [g], Brett A. Lidbury [a]

[a] National Centre for Epidemiology and Population Health, ANU College of Health and Medicine, The Australian National University, Acton, Australian Capital Territory, Australia
[b] Department of Biosciences and Biotechnology, Faculty of Pure and Applied Sciences, Kwara State University, Malete, Nigeria
[c] Department of Epidemiology, Biostatistics and Occupational Health, Faculty of Medicine and Health Sciences, McGill University, Montreal, Canada
[d] Statistical Support Network, The Australian National University, Acton, ACT, Australia
[e] Sydney School of Public Health, Faculty of Medicine and Health, The University of Sydney, New South Wales, Australia
[f] Department of Medical Parasitology and Infection Biology, Swiss Tropical and Public Health Institute, Basel, Switzerland
[g] Department of Clinical Sciences, Nigerian Institute of Medical Research, Yaba, Lagos State, Nigeria

ARTICLE INFO

Keywords:
Machine learning
Clinical-decision making
Hepatitis B virus
Hepatitis C virus
Human immunodeficiency virus
Blood-borne viral infections

ABSTRACT

Background: Machine learning (ML) prediction models to support clinical management of blood-borne viral infections are becoming increasingly abundant in medical literature, with a number of competing models being developed for the same outcome or target population. However, evidence on the quality of these ML prediction models are limited.
Objective: This study aimed to evaluate the development and quality of reporting of ML prediction models that could facilitate timely clinical management of blood-borne viral infections.
Methods: We conducted narrative evidence synthesis following the synthesis without meta-analysis guidelines. We searched PubMed and Cochrane Central Register of Controlled Trials for all studies applying ML models for predicting clinical outcomes associated with hepatitis B virus (HBV), human immunodeficiency virus (HIV), or hepatitis C virus (HCV).
Results: We found 33 unique ML prediction models aiming to support clinical decision making. Overall, 12 (36.4%) focused on HBV, 10 (30.3%) on HCV, 10 on HIV (30.3%) and two (6.1%) on co-infection. Among these, six (18.2%) addressed the diagnosis of infection, 16 (48.5%) the prognosis of infection, eight (24.2%) the prediction of treatment response, two (6.1%) progression through a cascade of care, and one (3.03%) focused on the choice of antiretroviral therapy (ART). Nineteen prediction models (57.6%) were developed using data from high-income countries. Evaluation of prediction models was limited to measures of performance. Detailed information on software code accessibility was often missing. Independent validation on new datasets and/or in other institutions was rarely done.
Conclusion: Promising approaches for ML prediction models in blood-borne viral infections were identified, but the lack of robust validation, interpretability/explainability, and poor quality of reporting hampered their clinical relevance. Our findings highlight important considerations that can inform standard reporting guidelines for ML prediction models in the future (e.g., TRIPOD-AI), and provides critical data to inform robust evaluation of the models.

* Corresponding author at: National Centre for Epidemiology and Population Health, ANU College of Health and Medicine, The Australian National University, Acton, Australian Capital Territory, Australia.
E-mail address: busayo.ajuwon@anu.edu.au (B.I. Ajuwon).

https://doi.org/10.1016/j.ijmedinf.2023.105244
Received 21 March 2023; Received in revised form 8 September 2023; Accepted 3 October 2023
Available online 4 October 2023

1. Introduction

Infectious diseases include a wide range of bacterial, fungal, parasitic, and viral diseases that pose a significant threat to global health security [1]. One major category of infectious diseases is blood-borne viruses, a group of viruses that are transmitted through blood or bodily fluids, including the primary pathogens of concern: human immunodeficiency virus (HIV), hepatitis B virus (HBV), and hepatitis C virus (HCV).

The overall burden of these viruses is significant. Global HIV infection caused 47.63 million disability-adjusted life years in 2019, presenting a 1.28-fold increase from 1990 to 2019 [2]. At the end of 2020, there were an estimated 37.7 million [30.2–45.1 million] people living with HIV, over two-thirds of whom (25.4 million) are in the WHO African Region [3]. While significant progress has been made in reducing HIV mortality by 51% during the last 20 years[4], deaths due to viral hepatitis remain high, far from the 2030 target of 65% reduction [5]. WHO estimated that 1.1 million people died from HBV and HCV in 2019, mostly from associated clinical complications of cirrhosis and liver cancer [6], with the number of deaths outstripping that from HIV, tuberculosis, and malaria [7]. Early detection and epidemiological surveillance of these viruses, through machine learning (ML) innovations, can impact the overall burden of disease, where a delay in clinical management may have significant implications on patient health outcomes.

ML offers significant potential for driving innovation in public health research. An important feature that underpins the significance of ML in health and medicine is its capacity to analyse large and complex data structures to develop prediction models, based on train-test functions. There is a growing body of evidence demonstrating that ML prediction models can support clinicians in delivering personalised patient care [8,9]. An illustrative instance is the autonomous FDA-approved ML system, IDs-DR EyeArt®, which effectively detects diabetic retinopathy in retinal fundus photographs and has shown improved patient outcomes across multiple settings [10–12]. ML prediction models to support clinical decision-making are therefore becoming increasingly abundant in medical literature, with a number of competing models being developed for the same outcome or target population [13,14]. For example, there are over 232 models proposed for predicting the diagnosis of COVID-19 in patients with suspected infection, and the prognosis of infected patients [15].

In recent times, ML prediction models have also been proposed to inform clinical decision-making for HBV, HCV, and HIV infections. One of such ML decision-making tools was developed by Kim *et al.* [16] for individualised risk prediction of hepatocellular carcinoma (HCC) in chronic hepatitis B (CHB) patients. However, these tools are yet to be deployed for use in clinical practice, despite their potential for improving patient outcomes. There are several possible reasons for this lack of implementation in clinical workflow, but the most notable is the need for robust critical appraisal and evidence synthesis on the development, validation, and quality of reporting of ML prediction models to guide healthcare providers, medical researchers, guideline developers, and policymakers in making evidence-based decision on the safe implementation and subsequent adoption of these tools in routine clinical practice [17–19].

Complete and transparent reporting leads to more comprehensive understanding, conduct, analysis, risk of bias assessment and potential usefulness of published prediction model studies. This ensures that subsequent researchers and users can further study the models to guide health care decisions. However, incomplete reporting hinders reproducibility and impedes validation by independent researchers, ultimately leading to research waste. Complete and informative reporting of the key aspects of prediction model studies in the light of existing evidence is therefore vital.

Previous systematic reviews have highlighted suboptimal reporting quality in prediction model studies across various clinical domains. The existing research has predominantly centred on models developed using traditional statistical techniques like logistic and Cox regression [20–22]. To address the need for enhanced transparency and reporting of prediction model studies, the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) guideline was formulated, comprising a checklist of 22 items (https://www.tripod-statement.org/) [23,24]. Whilst initially developed to support regression-based prediction model approaches, many aspects of the TRIPOD guideline can be suitably applied to studies involving prediction models that utilise ML methods [24])— as there are no specific reporting guideline tailored for ML-based prediction model studies. Ongoing efforts are still underway to develop a TRIPOD-AI version specific to ML interventions in different health research domains [25–27].

In this study, we aimed to evaluate the development and completeness of reporting of ML prediction models that could facilitate timely clinical management of blood-borne viral infections. This will provide the foundation for a broader research program focused on developing novel ML clinical decision support systems for HBV care in vulnerable populations, and also generate evidence to inform future reporting guidelines for ML prediction models, recognising that complete and transparent reporting via validated ML models are required to fulfil the promise of ML innovation.

2. Methods

This systematic review was conducted in accordance with the updated Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines [28], and the protocol (CRD42022332122) was pre-registered with PROSPERO; a database of systematic review protocols, maintained by the Centre for Reviews and Dissemination at the University of York. Ethics approval was not required for this study, as this was a systematic review of published articles.

2.1. Search strategy

We conducted a comprehensive search of PubMed and Cochrane Central Register of Controlled Trials (CENTRAL) for articles published from the inception of each database up until 11:59 PM, 8 May 2022, without restrictions on language, publication date or study setting. PubMed and CENTRAL databases were chosen to limit the search to literature that is most likely to be accessed and applied by medical and health professionals. The search was conducted using keywords related to ML and blood-borne viral infections. We did not include search terms related to decision support to allow for a broader search strategy that takes into account experimental study designs that might not yet focus on direct translation to clinical domain. This is important to capture and evaluate different methodological approaches in the early phase of ML applications in the clinical management of blood-borne viral infections. The search was further supplemented through citation searches and reference lists of relevant articles. The details of the search strategy are in the Supplementary file 1.

2.2. Eligibility criteria

In this systematic review, articles were considered for inclusion if the title and/or abstract indicated the report of primary research focused on ML prediction models for clinical diagnosis and management of blood-borne viral infections. Studies that used only conventional statistical methods (e.g., Cox regression, logistic regression) without any ML application, and studies with ML-based models that use non-routine data not readily available in clinical practice (e.g., whole-genome sequence data) were excluded, as were those with a non-clinical outcome (e.g., development of a biomarker). Electronic publications, including available abstracts of all articles retrieved from the search, were

independently screened by pairs of review authors (BIA, ONA, NR and AS) to select articles for full-text review. After removing duplicates from the initial search, the full-texts of all potentially relevant studies were reviewed to determine eligibility for inclusion. Any disagreements were resolved through consensus or by recourse to the senior author (BAL), if necessary. All articles identified in the searches were imported into the Covidence systematic review software [29], through which title, abstract screening, and full-text review were performed.

2.3. Data extraction

Data were extracted using a structured data collection form. Data fields of interest related to the following: general study characteristics (authors, publication year, country under study); study population (source of data, single or multi-centre, sample size, feature size); data pre-processing methods (handling missing data and unbalanced outcomes, other data pre-processing steps); clinical outcomes; analytical methods (statistical models, ML algorithms, validation methods, performance measures); results (best performing model). Details of the data extraction form are in Supplementary file 2. Data extraction was completed by one reviewer (BIA) and checked for completeness, accuracy, and consistency by a second reviewer (ONA or NR).

2.4. Assessing the quality of reporting in included studies

Risk of bias and quality of the studies were not assessed because the objective of this paper was descriptive, and not to draw conclusions about the validity of the prediction models and their performance estimates. However, the TRIPOD guideline was used to assess the quality of current reporting [23,24], as there are no reporting guidelines for ML-based clinical prediction models. As stated in the TRIPOD documentation, most terms in the TRIPOD checklist apply equally to ML methods developed, validated, or updated as prediction tools [24]. Therefore, the terms in the TRIPOD reporting checklist, relevant to both prediction model development and validation, were adapted for assessment, to understand the completeness of current reporting of the ML prediction models in the literature. Three terms were slightly adjusted to fit standard ML methods. Supplementary file 3 detailed the adjusted TRIPOD checklist. Notations of common ML terms are also presented in Table 1.

2.5. Evidence synthesis

Studies identified were heterogeneous, using different ML techniques and evaluating the algorithms with varying measures of performance and outcome predictions. A meta-analysis was therefore not considered meaningful; instead, we conducted a narrative synthesis of evidence following the synthesis without meta-analysis (SWIM) in systematic review reporting guidelines [30] to explore, describe, and interpret key findings.

3. Results

3.1. Study selection

Fig. 1 shows the PRISMA flow chart of the systematic review. The searches identified 574 records in total. Following the initial duplicate removal, 572 records were screened, and 395 were excluded after title and abstract review, primarily due to studies reporting on other disease conditions outside of the viral infections of interest. After abstract review, 177 full-text were deemed potentially relevant, from which a total of 33 studies were eligible for data extraction and subsequently included in the study, while 144 were excluded. Primary reasons for exclusion included ineligible outcomes (reporting on the use of ML for non-clinical outcomes, e.g., the development of a novel biomarker), ineligible study design (reporting only on conventional statistical methods), ineligible data type (studies using variables that are unavailable in routine clinical practice, e.g., whole-genome sequencing data, free-text data using natural language processing, and other non-routine data types), and ineligible patient population (reporting on diseased populations outside the scope of the systematic review).

Table 1
Description of some key ML terms.

Term	Basic explanation
Adaptive boosting (Adaboost)	An algorithm that is typically used to combine weak base learners to produce a more accurate model
Bootstrapping	A resampling method that uses random sampling with replacement
Decision tree (DT)	A tree-based algorithm that models decisions, outcomes, and predictions using a flowchart-like structure
External validation	Evaluation of the performance of prediction model in the face of an external dataset
Features	The input variables given to ML models
Gradient boosting (GB)	A tree-based algorithm that generates a final model from a series of decision trees by decreasing prediction errors from the individual model
Hyper-parameters	Parameters whose values are set before a learning algorithm is trained to control the learning process
Imbalanced data	Refers to those types of datasets where the target class/ outcome has uneven class distributions
K-nearest neighbours (KNN)	An algorithm that classifies new data points based on the similarity measure of the earlier stored data points
Naïve Bayes (NB)	An algorithm based on Bayes' theorem that returns a probabilistic prediction, with the assumption that all features are independent
Neural networks (NN)	A type of ML algorithm that uses interconnected nodes or neurons in a layered structure in a way that resembles the human brain
Over-fitting	A model fits exactly against its training data, but fails to generalise and perform accurately against unseen data
Random forest (RF)	An ensemble learning algorithm, where a large number of decision trees are aggregated using bootstrap resampling
Stacked ensemble	An ensemble algorithm that combines the predictions from two or more base algorithms on the same dataset
Supervised learning	A sub-group of ML models that learns from well-labelled training data to predict outcomes for unforeseen data
Support vector machine (SVM)	A supervised classifier that finds a hyperplane in N-dimensional space to distinctly classify the data points
Testing	A process where the performance of a fully trained model is evaluated on a testing set
Training	A primary step in ML, in which a model is fed with sufficient training data to learn from
Unsupervised learning	A sub-group of ML models that learns directly from data without any labels or predefined outcome of interest

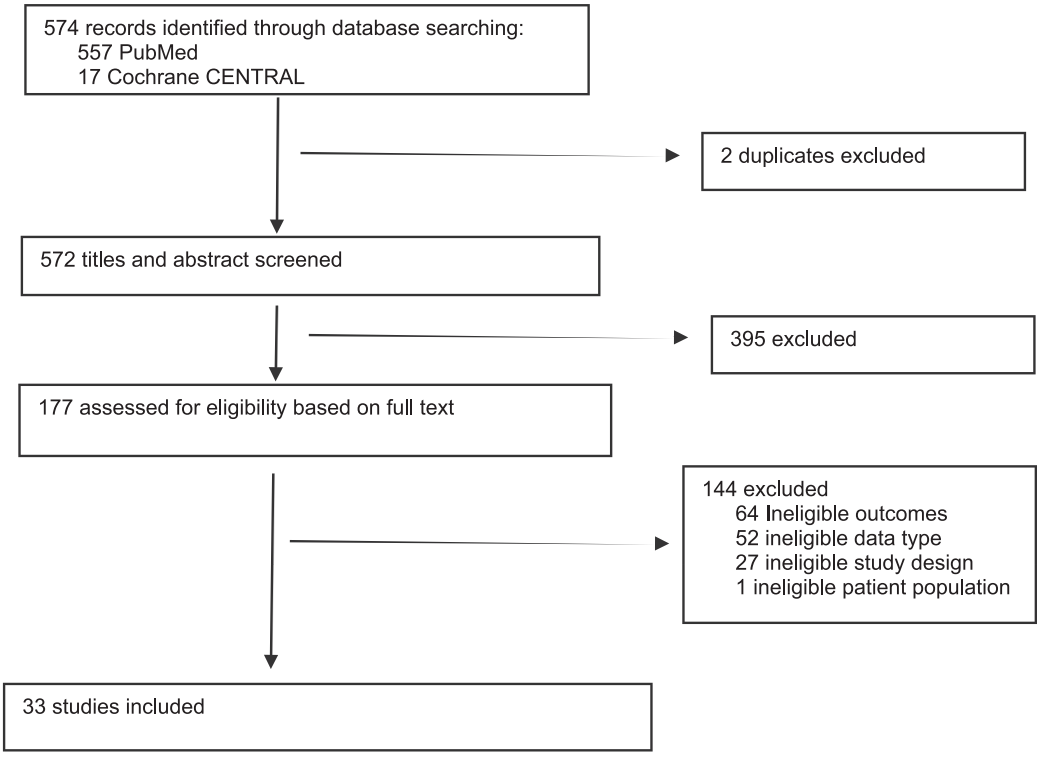


Fig. 1. PRISMA flow chart to select included studies CENTRAL; Central Register of Controlled Trials; PRISMA; Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

3.2. Scope of the evidence included

Fig. 2 represents the geographical locations of the region under study. The 33 included studies covered 14 countries across Asia, Africa, North America, Europe, and Oceania. Of the included studies, 12 (36.4%) focused on HBV, 10 (30.3%) on HCV, 10 (30.3%) on HIV, and two (6.1%) on co-infection. In one study [31], the authors included both HBV and HCV patient data. The sample size of all included studies ranged from 32 to 1,155,966 with a median of 1856.5. The median number of features included was 22.5, with values ranging from 4 to 284. Summary statistics of the identified studies per research area is shown in Table 2.

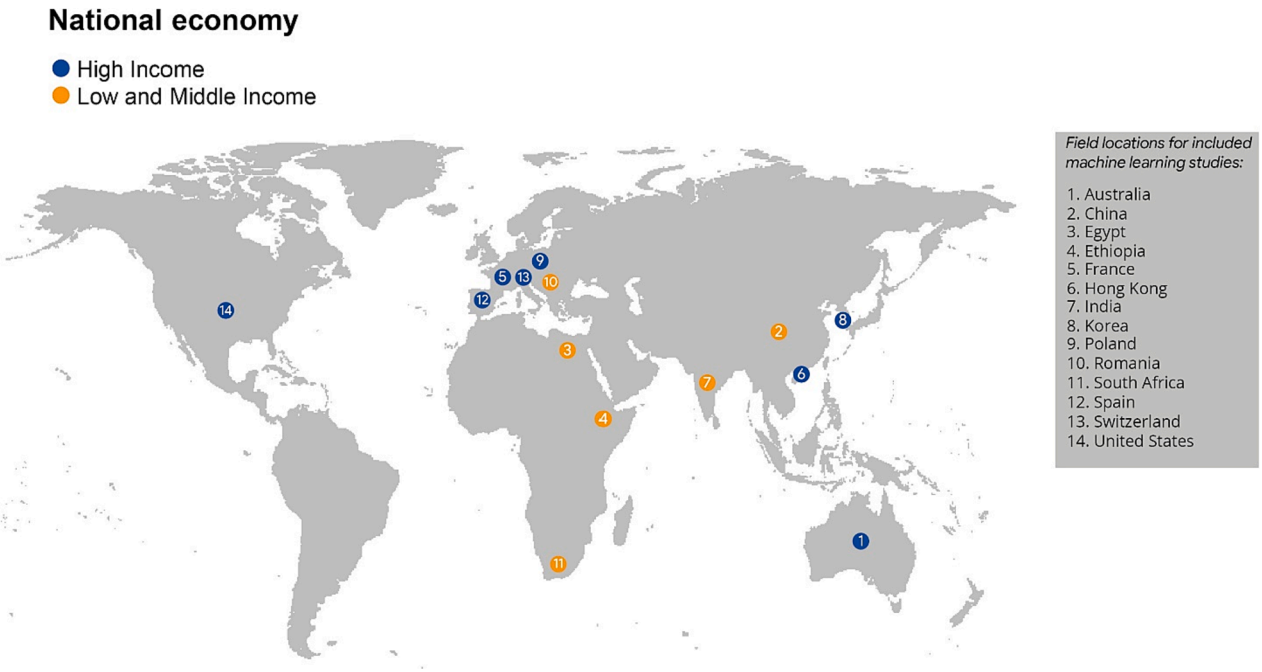


Fig. 2. Field locations for included papers. The national economy classification for each country represented in the study is presented in two different colours. Country income classification was based on the gross national income per capita data of the World Bank.

Table 2
Summary statistics per research area.

Research area	*Number of studies*	Features (n) Median (range)	Study size (n) Median (range)
HBV‡	12	17 (4–98)	655.5(141–60688)
HCV‡	10	25.5 (9–284)	5474 (442–120023)
HIV	10	35 (12–190)	2291(32–1155966)
Co-infection	2	32 (18–46)	68237.5 (1007–124006)

HBV; hepatitis B virus; HCV; hepatitis C virus; HIV; human immunodeficiency virus. ‡ One article included both HBV and HCV patient data.

3.3. Characteristics of included studies

The key characteristics of included studies are summarised in Supplementary file 4, and data for class imbalance and performance measures in supplementary file 5. Included studies ranged in geographical scope from single-centre studies to multi-centre studies. Of the 33 included studies, 57.6% (19/33) were single-centre, and 42.4% (14/33) were multi-centre.

An increase in the application of ML to fundamental blood-borne viral infections from 5 publications in 2013 to 10 publications in 2020 was observed as shown in Fig. 3. All included studies were published after 2012, with 19 (57.6%) published after 2019. While laboratory data was the primary source of data in included studies, a gradual diversification was observed, with treatment data now becoming relatively common. Overall, laboratory data was the most common source (28/33 studies, 84.8%), followed by clinical data (22/33 studies, 66.7%), treatment data (13/33 studies, 39.4%), and other sources (4/33 studies, 12.1%). However, in most of the studies, laboratory, clinical, and sometimes treatment variables were utilised in a complementary manner as model inputs. Nineteen studies (57.6%) exclusively utilised data from high-income countries (HICs) for model development.

The majority of the prediction models were developed in the United States (30.3 %, 10/33) and China (27.3%, 9/33), both representing 57.6% of the entire sample, while other 14 countries make up the other 43% (Fig. 4).

Adherence to the TRIPOD reporting guidelines was typically mixed (Supplementary file 6), with no clear pattern of improvement since TRIPOD was published in 2015. Consistently poor reporting remained for methods of blinding of outcomes or predictors, how performance measures were reported, reporting of the full prediction models and their functionalities, and code accessibility. For example, of the 30 studies (90.9%) reporting on the software used for modelling, 27 (90%) used open-source programming languages such as R or Python. However, only 15.1% (5/33) of included studies made their code publicly available[16,32–35].

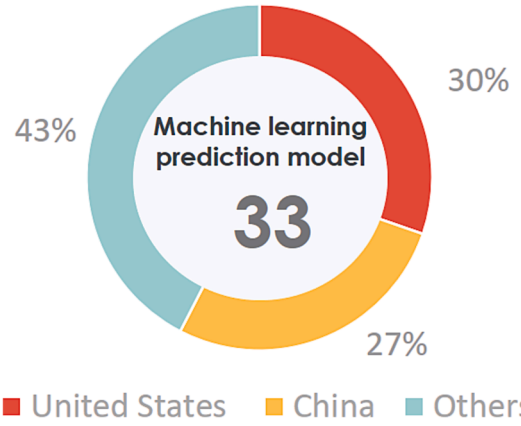


Fig. 4. The national concentration of the published ML prediction models.

Pre-processing of data was reported to varying degrees. Missing data handling was not described in 48.5% (16/33) of studies. When reported, different strategies were applied, including complete case analysis in 12 studies, multiple imputation in three studies, and single imputation in one study. One study combined the use of both complete case analysis and multiple imputation in handling missing data [36]. Similarly, class imbalance of the labelled outcome variable was observed in 42.4% (14/33) of the studies (Supplementary file 5). Among which seven studies addressed the imbalanced class distributions using different techniques, including undersampling [31,37,38], oversampling [39], random oversampling (ROSE) [40], and synthetic minority oversampling (SMOTE) [36,41]. Hyper-parameter optimisation was not described in 54.5% (18/33) of studies.

3.4. ML algorithms in use

Different ML algorithms were used in the included studies. The most commonly applied ML algorithm was RF (21 studies, 63.6%), DT (18 studies, 54.5%), NN (10 studies, 30.3%), GB (seven studies, 21.2%), and SVM (five studies, 15.2%). Other algorithms, including NB, Adaboost, KNN, and stacked ensemble were used in three (9.1%), two (6.1%), two (6.1%), and one (3.0%) studies, respectively. A number of studies employed more than one ML technique. DT and NN were the primary algorithms used in earlier publications, but were gradually overtaken by RF (Fig. 5).

Logistic regression was used in 16 studies (48.5%) as a comparative reference to reflect the traditional statistical approach. Of the 18 studies (54.5%) comparing the performance of traditional statistical approach

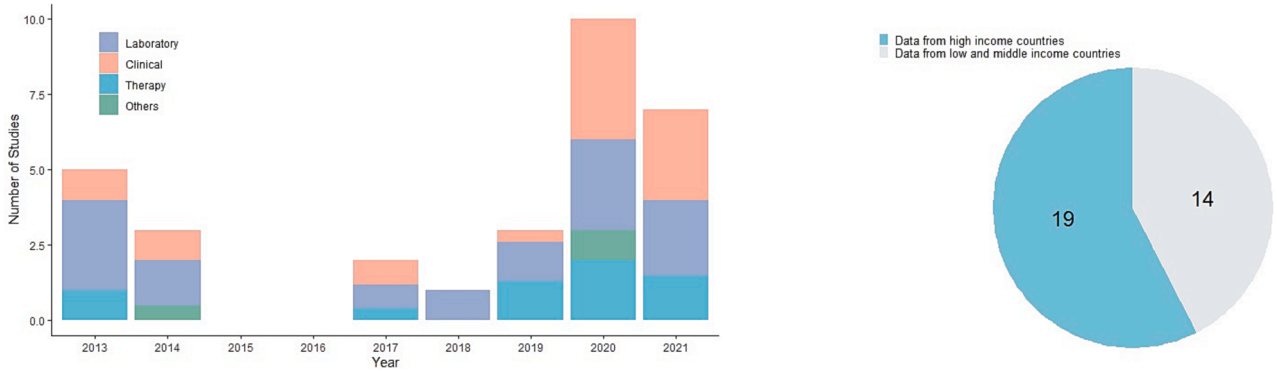


Fig. 3. Number of studies from each data source used and the type of settings in the included studies. (A) The number of included studies using each routine data type. (B) The proportion of included studies using data from different geographical settings. The total from 2022 was excluded from Fig. 3A because data were unavailable for the full year, so a visual comparison with other years would not be accurate. The total studies for 2022 (1 Jan 2022 to 8 May 2022) were two. The data source bar represents the proportion of studies that used a particular data source within the total number of studies published in that year.

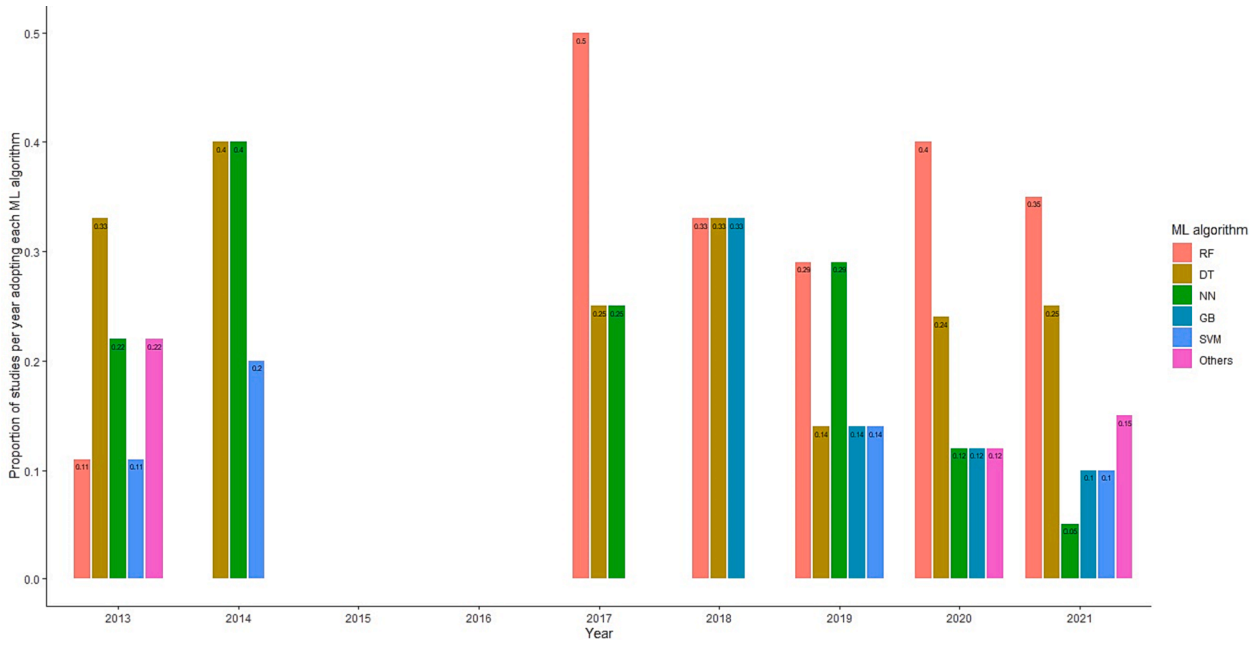


Fig. 5. Annual proportion of studies that used a particular ML algorithm. Studies for 2022 were excluded because the full-year data were unavailable, so visual comparison with other years would not be accurate.

with ML algorithms, 14 (77.8%) reported that ML algorithms outperformed the conventional statistical models, including standard logistic regression, LASSO regression and ridge regression. The best performing ML algorithms relative to the comparative reference, included RF in four studies [34,38,42,43], NN in four studies [44–47], and GB in three studies [33,35,48], while DT [49], Adaboost [36], and stacked ensemble [37] outperformed the competing conventional statistical models in one study each. In 23 (69.7%) studies, an AUROC was reported as a discrimination performance measure. Other commonly reported discrimination and performance measures included sensitivity, specificity, accuracy, and precision (supplementary file 5). Model calibration was reported in only four studies [16,33,42,50].

3.5. ML prediction models for clinical decision support

Of the 33 unique ML prediction models reported in all the included studies, six (18.2%) addressed the diagnosis of infection, 16 (48.5%) addressed the prognosis of infection— identifying a patient's risk of developing complications associated with chronic viral hepatitis, including progressive liver fibrosis, cirrhosis and life-threatening clinical complications of end-stage HCC, eight (24.2%) addressed the prediction of treatment response, two (6.1%) addressed progression through a cascade of care including predicting patient retention, and one (3.03%) focused on the choice of antiretroviral therapy (ART). The majority of the models were limited to internal validation (69.7%, 23/33), and only 10 were assessed for wide-scale performance and generalisability through external validation on independent patient cohorts [16,35,42,43,49–54]. The evaluation of the prediction models was limited to measures of performance, and none was evaluated for clinical impact and adoption in real-world settings.

3.5.1. HBV

Table 3 provides a summary of the ML prediction models for clinical decision support of HBV infection. Twelve prediction models addressed HBV infection, with two focused on treatment response and outcome, seven on prognosis of HBV, and three predicted HBV diagnosis to help clinicians make a quicker decision on patient infection status. The sample size of patient data used for developing HBV model of infection ranged from 141 to 60688, with a median of 655.5. The median number of features included was 17, with values ranging from 4 to 49.

Table 3
ML prediction models for HBV.

First author, year	Country	Predicted clinical outcome	Evaluation
Diagnosis			
Ramrakhiani, 2021 [38]	United States	HBsAg status	Performance
Richardson, 2013 [31] ‡	Australia	HBsAg status	Performance
Shang, 2013 [58]	China	HBsAg status	Performance
Prognosis			
Cao, 2013 [55]	China	Liver cirrhosis	Performance
Cao, 2013 [56]	China	Liver cirrhosis	Performance
Kim, 2022 [16]	Korea	HCC occurrence	Automated model deployment as a web-tool
Wang, 2014 [57]	China	Liver fibrosis	Performance
Wei, 2018 [35]	China	Hepatic fibrosis and liver cirrhosis	Automated model deployment as a web-tool
Xie, 2020 [43]	China	Liver fibrosis	Performance
Zheng, 2014 [47]	Hong-Kong	HBsAg seroclearance	Performance
Treatment outcome			
Tian, 2019 [48]	China	HBsAg seroclearance after treatment	Performance
Wei, 2019 [46]	China	Response to chronic hepatitis B therapy	Performance

HBsAg; hepatitis B surface antigen; HCC; hepatocellular carcinoma. ‡ One article included both hepatitis B and hepatitis C patient data.

Two ML prediction models addressed the prediction of treatment outcomes in HBV patients. One predicted hepatitis B surface antigen (HBsAg) seroclearance [48], and the other predicted liver fibrosis reverse at baseline and 1.5 years after initial therapy [46]. They both analysed therapeutic history and laboratory data. Therapeutic history was limited to previous HBV therapies. We found seven ML-based models for the prediction of the prognosis of HBV infection: three predicted HBV-induced liver cirrhosis in Chinese populations using laboratory and clinical data [55,56]; two predicted the occurrence of HCC,

with one using laboratory and treatment data [16], and the other combining laboratory data with metabolomics for personalised prediction of HCC risk [57]; one quantitatively profiled laboratory and serum metabolites to predict liver fibrosis in Chinese patients [43]; and one used clinical and laboratory data to predict HBsAg sero-clearance in hepatitis B e antigen (HBeAg)-negative CHB patients [47]. The three ML prediction models to guide HBV diagnosis focused on primary care settings: one analysed demographic data from a population-based dataset [38], whereas the other two used laboratory data [31,58].

Of the 12 ML prediction models, eight (66.7%) were developed in China, and the others were developed in Korea, United States, Australia, and Hong Kong, with none available for clinical decision support of HBV in African populations, including Nigeria. While all the 12 prediction models were evaluated with measures of performance, none progressed from pilot study to evaluation of clinical impact in real-world settings. However, two of the models were translated into an automated web-analytic, point-of-care decision-making tool for future clinical evaluation. These two models, LiveBoost (https://metabolomics.cc.hawaii.edu/software/LiveBoost/)[35] and PLAN B (https://planbhcc.com/) [16] were developed in Korean and Chinese populations for individualised risk prediction of HCC and advanced hepatic fibrosis, respectively.

3.5.2. HCV

Ten ML prediction models addressed HCV infection, with five focused on prognosis, two on treatment outcomes, two on diagnosis, and one predicted progression through a cascade of care. The information is summarised in Table 4. The median sample size for the HCV model of infection was 5474 and the median number of features included was 25.5, with values ranging from 9 to 284.

Among the five ML prediction models for the prognosis of HCV, three predicted the occurrence of HCC [42,45,49], and two predicted liver fibrosis progression [44,52]. Their use of patient data was homogenous, with the prediction models analysing clinical and laboratory data. Therapeutic history of patient data was further used as additional input variables in two of the prediction models. Two ML prediction models addressed the prediction of treatment outcomes in HCV patients using clinical, laboratory, and HCV treatment data; both predicting direct-acting antivirals (DAA) treatment failure in a large Polish multicentre cohort [39] and HCV-TARGET consortium in the United States [33], respectively. We found two prediction models for the diagnosis of HCV: one used medical claims data [37], and the other analysed routine laboratory data [31]. One ML prediction model to guide progression through the cascade of care focused on three care outcomes: linkage to care, initiation of antiviral treatment, and virologic cure [59].

Of the 10 HCV models of infection, five (50%) were developed in the United States, and the remaining five were developed in Egypt, France, Poland, and Australia. However, none of the prediction models were deployed and evaluated for clinical impact in real-world settings.

3.5.3. HIV

Ten ML prediction models addressed HIV infection. Among these prediction models, four focused on predicting HIV treatment response, three on the prognosis of HIV, one on ART therapy selection, one on care outcome, and one predicted the diagnosis of HIV. The information is summarised in Table 5. The sample size used for developing HIV model of infection ranged from 32 to 1155966, with a median of 2291. The median number of features included was 35, with values ranging from 12 to 190.

Four ML prediction models addressed the prediction of treatment outcomes in HIV patients; one predicted the probability of viral control after ART treatment interruptions in HIV research trials [51], one combined medical adherence and clinical data to predict RNA viral rebound [60], one analysed the predictors of CD4 count changes among patients on ART [41], and one predicted virological response to ART [53]. Only one ML prediction model did not use therapeutic history of patients [51], in addition to the common laboratory data. We found three ML models for the prediction of the prognosis of HIV infection: one predicted HIV-induced neurocognitive impairments in Spanish populations using laboratory and clinical data [61]; one examined the relationships between circulating biomarkers and HIV symptoms using laboratory data and clinical data in combination with serum metabolites [40], and the other predicted the burden of comorbidity among people living with HIV using laboratory data and clinical data [62].

Of the ten ML prediction models for HIV infection, four (40%) were developed in the United States, two (20%) in Ethiopia, two (20%) in Spain, one in Switzerland, and another in a multi-institutional study with data from South Africa, Romania and India. However, none of these

Table 4
ML prediction models for HCV.

First author, year	Country	Predicted clinical outcome	Evaluation
Diagnosis			
Doyle, 2020 [37]	United States	HCV status	Performance
Richardson, 2013 [31] ‡	Australia	HCV status	Performance
Prognosis			
Audureau, 2020 [42]	France	HCC occurrence	Performance
Emu, 2020 [44]	Egypt	Liver fibrosis progression	Performance
Hashem, 2020 [49]	Egypt	HCC occurrence	Performance
Ioannou, 2020 [45]	United States	HCC occurrence	Performance
Konerman, 2017 [52]	United States	Liver fibrosis progression	Performance
Treatment outcome			
Janczewska, 2021 [39]	Poland	DAA treatment failure	Performance
Park, 2021 [33]	United States	DAA treatment failure	Performance
Progression through a cascade of care			
Nakayama, 2021 [32]	United States	Linkage to care, initiation of antiviral therapyand virologic cure	Performance

HCV; hepatitis C virus; HCC; hepatocellular carcinoma; DAA; direct-acting antivirals. ‡ One article included both hepatitis B and hepatitis C patient data.

Table 5
ML prediction models for HIV.

First author, year	Country	Predicted clinical outcome	Evaluation
Diagnosis			
Krakower, 2019 [50]	United States	HIV acquisition	Performance
Prognosis			
Munoz-Moreno, 2014 [61]	Spain	Neurocognitive impairment	Performance
Yang, 2021 [62]	United States	Burden of comorbidity	Performance
Zuñiga, 2020 [40]	United States	Prevalent HIV symptoms	Performance
Treatment outcome			
Feher, 2020 [51]	Spain	Viral control after ART treatment interruptions	Performance
Kamal, 2020 [60]	Switzerland	Viral rebound	Performance
Kebede, 2017 [41]	Ethiopia	CD4 count changes	Performance
Revell, 2013 [53]	South Africa, Romania, India	Response to ART	Performance
ART therapy selection			
Nemomsa, 2021 [63]	Ethiopia	ART regimen	Performance
Retention in care			
Ramachandra, 2020 [34]	United States	Risk of falling out of care	Performance

HIV; human immunodeficiency virus; ART; antiretroviral therapy.

were deployed and adopted for clinical use in real-world settings.

3.5.4. Co-infection

Two ML prediction models addressed co-infection. The first addressed the diagnosis of patients co-infected with HBV and HIV in China [36], and the second, which was developed in Hong Kong, addressed the risk prediction of HCC occurrence in patients with CHB and CHC [54] (Table 6). The sample size for developing the model for HBV/HIV co-infection and HBV/HCV co-infection ranged from 1007 to 124006, with a median of 68237.5. The median number of features included was 32, with values ranging from 18 to 46.

4. Discussion

We evaluated ML-based prediction models designed for the clinical management of blood-borne viral infections. A total of 33 clinically-relevant studies met the inclusion criteria; however, there was a wide variation in outcome measures and reporting quality. Our study revealed a growing trend in publication, with 19 of the included studies published after 2019, demonstrating the increasing relevance of ML for predicting clinical outcomes associated with blood-borne viral infections. We identified 14 ML prediction models that are adapted to LMICs, with nine of them specifically developed in Chinese populations. The clinical relevance of most existing ML prediction models for blood-borne viral infections is yet to fully emerge. This can be attributed to the paucity of external validation, the heterogeneity and complexity of variables required for analysis, and considerable variation in the studied populations. Our results highlight specific reporting areas that can inform the formulation of standardised reporting guidelines for ML prediction models in the future (e.g., TRIPOD-AI).

Although HBV, HCV, and HIV stand as three of the most prevalent blood-borne viral infections encountered in clinical practice, early prediction of their clinical outcomes remains a challenge. In order to steer future advancements in the field and facilitate the integration of prediction models into clinical workflows, we have identified important areas of concern. First, it is worth noting that some of the included studies used relatively small datasets, which could impede the full potential of ML in public health. For instance, two studies predicted HIV outcomes using datasets with fewer than 40 patients [40,51]. With such a small dataset, it becomes impracticable to divide data into distinct segments for training, testing, and validation. Consequently, this might lead to inflated accuracy of the outcome response due to overfitting, potentially undermining the clinical utility of the prediction model and raising questions about the quality of the underlying data. Second, the ML community's adage, "garbage in, garbage out," underscores the significance of data quality when training prediction models. Our study revealed a wide variation in the reported datasets, particularly around the source and type of dataset used for model development. This variation was particularly evident in frequently used laboratory and clinical datasets, where important patient demographics such as birth year, ethnicity/race, birthplace/residence, and travel history were inconsistently documented. Notably, no ML prediction model considered patient exposure to high-risk environments (i.e., previous travel history). ML prediction models developed with data that inadequately represent the broader patient population in which they are intended to be applied might harbor intrinsic biases, including spectrum bias.

Third, the dearth of external validation studies and suboptimal reporting quality constitute significant barriers to the successful implementation of ML prediction models in clinical settings. Our study revealed that only 10 (30.3%) prediction models were validated externally in an independent patient cohort, which is reflective of a broader issue with the transportability and generalisability of ML clinical prediction models [64,65]. Similarly, we found only five studies that made their code openly accessible — this further amplifies the ongoing discourse on reproducibility [66], and echoes the widespread critique of ML tools as black boxes [67]. Additionally, the TRIPOD guidelines detailed a series of important checkpoint for ML research [24]; however, we found only five studies that acknowledged adherence to reporting guidelines, such as TRIPOD. Fourth, the clinical evaluation of ML prediction models in real-world settings is still lacking. None of the prediction models has been evaluated for clinical impact, and their adoption in real-world clinical settings is yet to be exploited. Albeit, two of the prediction models were translated into an automated web-tool [16,35], the implementation, interface integration, and subsequent adoption of these tools by clinicians are required. This will involve conducting extensive pilot research to test the clinical utility of the tools, followed by randomised clinical trials to evaluate patient outcomes, process improvement and cost-effectiveness.

This study has several strengths. First, we used a comprehensive search strategy in electronic database and trial registries, without any restrictions on date and language. This approach minimises the likelihood of missing published trials; however, there is a possibility that unpublished or ongoing trials not registered in clinical trial databases might have been missed. Second, we aimed to reduce bias by having at least two review authors work independently on study selection, data extraction, and assessing the quality of reporting. Third, we adhered to the PRISMA guideline - in reporting the study methodology and results. Fourth, all our included studies were published from 2013 upward, reflecting current developments in the field. Whilst ML techniques have existed since the 1960s [68], their development was hindered by computational limitations and data availability constraints. The utilisation of these techniques in clinical contexts started gaining momentum only in the 21st century, spurred by advancements in computing power. The earliest research describing new extensions and developments of ML techniques emerged in the 2000s [69].

The study was limited by substantial variability in the quality of reporting in the included studies, even though majority of the studies were published after the TRIPOD statement. This precluded a meta-analysis of the data. Also, many ML technologies developed by commercial entities do not have published data in academic journals, hence, it is likely that some ML innovations in commercial research archives and networks, such as IBM, Apple, and NHS Digital, may not have been identified through systematic searches. Nonetheless, it is worth noting that the field of research is swiftly evolving and will require subsequent updates.

Previous studies focusing on ML-assisted melanoma diagnosis have highlighted the challenges in comparing ML prediction models. These challenges stem from the use of undisclosed training methods [70], lack of prospective studies, concerns about bias and overfitting [71], and issues related to model generalisability. Our study aligns with these findings. Obermayer and Topol [72] emphasised the significance of diverse training data for ML prediction models to ensure unbiased predictions across gender, sociodemographic status, and ethnicity [73]. In this systematic review, we found that many of the included studies heavily relied on datasets from the United States and China. Research has shown that data availability often correlates with healthcare access, which can be more limited for vulnerable populations with fragmented care systems [74]. Consequently, the development of ML clinical prediction models is often constrained in resource-limited settings, where

Table 6
ML prediction models for HBV/HIV and HBV/HCV co-infection.

First author, year	Country	Predicted clinical outcome	Evaluation
Diagnosis			
Yin, 2021 [36]	China	Patients co-infected with HBV and HIV	Performance
Prognosis			
Wong, 2022 [54]	Hong-Kong	HCC occurrence in patientswith chronic HBV and HCV	Performance

HBV; hepatitis B virus; HIV; human immunodeficiency virus; HCC; hepatocellular carcinoma; HCV; hepatitis C virus.

patient data and electronic databases are either scarce or non-existent. This potentially undermines the clinical utility of the prediction models and amplifies healthcare inequalities.

ML prediction models hold promising potential for advancing HBV, HIV, and HCV clinical research and patient care. Nevertheless, this study highlights that research in this domain is still in its early stages. It is, however, encouraging to see a growing trend of studies addressing implementation considerations, including the use of ML to assist clinicians in making timely and accurate decisions on the diagnosis, prognosis, therapy selection, and treatment outcomes of important blood-borne viral infections which constitute a significant public health threat. We identified two good examples of prediction models that have been translated to point-of-care decision-making tools as a proof-of-concept, but robust validation is required before they can be adopted for clinical use. Before clinical adoption, several considerations must also be addressed. This includes conducting interpretability analysis to foster trust among clinicians and patients, evaluating the acceptability of such systems to both groups, and devising strategies for seamless integration into clinical workflows and healthcare systems for improved patient care.

It is important to stress that ML may not replace human experts, as Holzinger [75] noted, interactive ML requiring a human in the loop remains essential. Thus, investing in the capacity building of clinicians and clinical scientists to develop proficiency in creating and utilising ML prediction models becomes crucial. This proficiency must encompass both quantitative modelling and translational skills, effectively bridging the realms of ML and clinical decision-making. Failure to strike a balance between these two critical skills could impede the seamless integration of ML prediction models into routine patient management and clinical care.

Whilst the general public and patients [76] hold optimistic views towards ML innovations, it is imperative to subject these tools to thorough evaluation to ascertain their accuracy, cost-effectiveness, and safety for clinical use. Currently, guidelines are in development for assessing and reporting ML interventions in studies involving diagnostic accuracy, prediction models, and clinical trials [25–27]. To address the concerns highlighted in this study, we have formulated a checklist proposed for developing and evaluating ML prediction models designed to support the clinical management of blood-borne viral infections (see Supplementary File 7). If widely adopted, we anticipate that this checklist will facilitate the clinical relevance of ML prediction models and enhance their successful implementation and adoption in public health practice. While the checklist is tailored to the development of ML prediction models for the clinical management of HBV, HIV, and HCV, the issues raised encompass broader considerations in ML model development and validation, with potential applicability in other domains of infectious disease.

5. Conclusion

Existing ML prediction models designed for the management of blood-borne viral infections encompass a diverse range of clinical outcomes, including diagnostic predictions, prognoses, treatment evaluations, care outcomes, and antiviral therapy selections. However, a significant number of these models do not meet the essential reporting standards for clinical prediction tools. Moreover, there is a lack of evidence regarding their practical application and potential impact in real-world clinical settings. This gap is further compounded by the models' reliance on patient data from HICs, mainly drawn from large open-access datasets. To enhance development in the field, future ML prediction models for blood-borne viral infections should be developed in diverse population settings, including low and middle income countries where representation is currently low. Research endeavours should prioritise model transferability, integrate clinical domain into study design, enhance model interpretability, and improve the quality of reporting. This will streamline clinical implementation and ensure sustainable adoption into clinical workflow. Additionally, registering ML prediction models on public databases will enable adequate evaluations using standardised frameworks, independent of any financial interests. Amidst these efforts, it is crucial to address data privacy concerns and the ethics of responsibility and accountability when adopting ML prediction models. These considerations are critical for enhancing the responsible use of ML prediction tools in routine clinical practice and public health.

CRediT authorship contribution statement

Busayo I. Ajuwon: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Validation, Visualization, Writing – original draft, Writing – review & editing. **Oluwatosin N. Awotundun:** Data curation, Validation. **Alice Richardson:** Writing – review & editing. **Katrina Roper:** Writing – review & editing. **Meru Sheel:** Writing – review & editing. **Nurudeen Rahman:** Data curation, Validation. **Abideen Salako:** Data curation, Validation. **Brett A. Lidbury:** Writing – review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijmedinf.2023.105244.

References

[1] D.E. Bloom, D. Cadarette, Infectious disease threats in the twenty-first century: Strengthening the global response, Front. Immunol. 10 (2019) 549.
[2] J. Wu, T. Lai, H. Han, J. Liu, S. Wang, J. Lyu, Global, regional and national disability-adjusted life years due to hiv from 1990 to 2019: Findings from the global burden of disease study 2019, Trop. Med. Int. Health. 26 (6) (2021) 610–620.
[3] World health organization. Hiv/aids key facts, 2021. Https://www.Who.Int/news-room/fact-sheets/detail/hiv-aids accessed 28 july 2022.
[4] World health organization. Global health estimates, 2000-2019. Https://www.Who.Int/data/global-health-estimates accessed 31 july 2022.
[5] World health organization. Interim guidance for country validation of viral hepatitis elimination, 2021. Https://www.Who.Int/publications/i/item/9789240028395.
[6] World health organization. Global progress report on hiv, viral hepatitis and sexually transmitted infections, 2021. Https://www.Who.Int/publications/i/item/9789240027077 accessed 20 july 2022.
[7] Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: A systematic analysis for the global burden of disease study 2019. Lancet. 2020;396(10258):1204-22.
[8] N.M. Patel, V.V. Michelini, J.M. Snell, S. Balu, A.P. Hoyle, J.S. Parker, et al., Enhancing next-generation sequencing-guided cancer care through cognitive computing, Oncologist. 23 (2) (2018) 179–185.
[9] A. Rajkomar, E. Oren, K. Chen, A.M. Dai, N. Hajaj, M. Hardt, et al., Scalable and accurate deep learning with electronic health records, NPJ Digit. Med. 1 (2018) 18.
[10] E. Ipp, D. Liljenquist, B. Bode, V.N. Shah, S. Silverstein, C.D. Regillo, et al., Pivotal evaluation of an artificial intelligence system for autonomous detection of referrable and vision-threatening diabetic retinopathy, JAMA Netw. Open. 4 (11) (2021) e2134254.
[11] M. Bhaskaranand, C. Ramachandra, S. Bhat, J. Cuadros, M.G. Nittala, S.R. Sadda, et al., The value of automated diabetic retinopathy screening with the eyeart system: A study of more than 100,000 consecutive encounters from people with diabetes, Diabetes. Technol. Ther. 21 (11) (2019) 635–643.
[12] Businesswire. Eyenuk announces fda clearance for eyeart autonomous ai system for diabetic retinopathy screening. Https://www.Businesswire.Com/news/home/20200805005495/en/eyenuk-announces-fda-clearance-eyeart-autonomous-ai accessed 20 july 2022.
[13] C. Krittanawong, H.U.H. Virk, S. Bangalore, Z. Wang, K.W. Johnson, R. Pinotti, et al., Machine learning prediction in cardiovascular diseases: A meta-analysis, Sci. Rep. 10 (1) (2020) 16057.
[14] W. Wang, M. Kiik, N. Peek, V. Curcin, I.J. Marshall, A.G. Rudd, et al., A systematic review of machine learning models for predicting outcomes of stroke with structured data, PLoS One. 15 (6) (2020) e0234722.

[15] L. Wynants, B. Van Calster, G.S. Collins, R.D. Riley, G. Heinze, E. Schuit, et al., Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal, Bmj. 369 (2020), m1328.

[16] H.Y. Kim, P. Lampertico, J.Y. Nam, H.C. Lee, S.U. Kim, D.H. Sinn, et al., An artificial intelligence model to predict hepatocellular carcinoma risk in korean and caucasian patients with chronic hepatitis b, J. Hepatol. 76 (2) (2022) 311–338.

[17] R. Challen, J. Denny, M. Pitt, L. Gompels, T. Edwards, K. Tsaneva-Atanasova, Artificial intelligence, bias and clinical safety, BMJ. Qual. Saf. 28 (3) (2019) 231–327.

[18] Royal college of general practitioners. Artificial intelligence and primary care. Https://www.Racgp.Org.Au/advocacy/position-statements/view-all-position-statements/clinical-and-practice-management/artificial-intelligence-in-primary-care accessed 28 july 2022.

[19] The topol review: Preparing the healthcare workforce to deliver the digital future, 20Https://topol.Hee.Nhs.Uk/wp-content/uploads/hee-topol-review-2019.Pdf accessed 28 july 2022.

[20] P. Heus, J. Damen, R. Pajouheshnia, R. Scholten, J.B. Reitsma, G.S. Collins, et al., Poor reporting of multivariable prediction model studies: Towards a targeted implementation strategy of the tripod statement, BMC Med. 16 (1) (2018) 120.

[21] A.H. Zamanipoor Najafabadi, C.L. Ramspek, F.W. Dekker, P. Heus, L. Hooft, K.G. M. Moons, et al., Tripod statement: A preliminary pre-post analysis of reporting and methods of prediction models, BMJ Open. 10 (9) (2020) e041537.

[22] J.A. Damen, L. Hooft, E. Schuit, T.P. Debray, G.S. Collins, I. Tzoulaki, et al., Prediction models for cardiovascular disease risk in the general population: Systematic review, BMJ 353 (2016), i2416.

[23] K.G. Moons, D.G. Altman, J.B. Reitsma, J.P. Ioannidis, P. Macaskill, E. W. Steyerberg, et al., Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (tripod): Explanation and elaboration, Ann. Intern. Med. 162 (1) (2015) W1–W.

[24] G.S. Collins, J.B. Reitsma, D.G. Altman, K.G.M. Moons, Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (tripod): The tripod statement, BMC. Medicine. 13 (1) (2015) 1.

[25] G.S. Collins, K.G.M. Moons, Reporting of artificial intelligence prediction models, Lancet. 2019;393(10181):1577-79.

[26] X. Liu, S.C. Rivera, L. Faes, L.F. di Ruffano, C. Yau, P.A. Keane, et al., Reporting guidelines for clinical trials evaluating artificial intelligence interventions are needed, Nat. Med. 25 (10) (2019) 1467–1468.

[27] V. Sounderajah, H. Ashrafian, R. Aggarwal, J. De Fauw, A.K. Denniston, F. Greaves, et al., Developing specific reporting guidelines for diagnostic accuracy studies assessing ai interventions: The stard-ai steering group, Nat. Med. 26 (6) (2020) 807–808.

[28] M.J. Page, J.E. McKenzie, P.M. Bossuyt, I. Boutron, T.C. Hoffmann, C.D. Mulrow, et al., The prisma 2020 statement: An updated guideline for reporting systematic reviews, Bmj. 372 (2021), n71.

[29] Veritas health innovation. Covidence systematic review software. 2018. Www. Covidence.Org accessed 8 may 2022.

[30] M. Campbell, J.E. McKenzie, A. Sowden, S.V. Katikireddi, S.E. Brennan, S. Ellis, et al., Synthesis without meta-analysis (swim) in systematic reviews: Reporting guideline, Bmj. 368 (2020), l6890.

[31] A.M. Richardson, B.A. Lidbury, Infection status outcome, machine learning method and virus type interact to affect the optimised prediction of hepatitis virus immunoassay results from routine pathology laboratory assays in unbalanced data, BMC Bioinformatics. 14 (2013) 206.

[32] J.Y. Nakayama, J. Ho, E. Cartwright, R. Simpson, V.S. Hertzberg, Predictors of progression through the cascade of care to a cure for hepatitis c patients using decision trees and random forests, Comput. Biol. Med. 134 (2021), 104461.

[33] H. Park, W.H. Lo-Ciganic, J. Huang, Y. Wu, L. Henry, J. Peter, et al., Machine learning algorithms for predicting direct-acting antiviral treatment failure in chronic hepatitis c: An hcv-target analysis, Hepatology (2022).

[34] A. Ramachandran, A. Kumar, H. Koenig, A. De Unanue, C. Sung, J. Walsh, et al., Predictive analytics for retention in care in an urban hiv clinic, Sci. Rep. 10 (1) (2020) 6421.

[35] R. Wei, J. Wang, X. Wang, G. Xie, Y. Wang, H. Zhang, et al., Clinical prediction of hbv and hcv related hepatic fibrosis using machine learning, EBioMedicine. 35 (2018) 124–132.

[36] Y. Yin, M. Xue, L. Shi, T. Qiu, D. Xia, G. Fu, et al., A noninvasive prediction model for hepatitis b virus disease in patients with hiv: Based on the population of jiangsu, china, Biomed. Res. Int. 2021 (2021) 6696041.

[37] O.M. Doyle, N. Leavitt, J.A. Rigg, Finding undiagnosed patients with hepatitis c infection: An application of artificial intelligence to patient claims data, Sci. Rep. 10 (1) (2020) 10521.

[38] N.S. Ramrakhiani, V.L. Chen, M. Le, Y.H. Yeo, S.D. Barnett, A.K. Waljee, et al., Optimizing hepatitis b virus screening in the united states using a simple demographics-based model, Hepatology. 75 (2) (2022) 430–437.

[39] E. Janczewska, M.F. Kołek, B. Lorenc, J. Klapaczyński, M. Tudrujek-Zdunek, M. Sitko, et al., Factors influencing the failure of interferon-free therapy for chronic hepatitis c: Data from the polish epiter-2 cohort study, World. J. Gastroenterol. 27 (18) (2021) 2177–2192.

[40] J.A. Zuñiga, M.L. Harrison, A. Henneghan, A.A. García, S. Kesler, Biomarkers panels can predict fatigue, depression and pain in persons living with hiv: A pilot study, Appl. Nurs. Res. 52 (2020), 151224.

[41] M. Kebede, D.T. Zegeye, B.M. Zeleke, Predicting cd4 count changes among patients on antiretroviral treatment: Application of data mining techniques, Comput. Methods. Programs. Biomed. 152 (2017) 149–157.

[42] E. Audureau, F. Carrat, R. Layese, C. Cagnot, T. Asselah, D. Guyader, et al., Personalized surveillance for hepatocellular carcinoma in cirrhosis - using machine learning adapted to hcv status, J. Hepatol. 73 (6) (2020) 1434–1445.

[43] G. Xie, X. Wang, R. Wei, J. Wang, A. Zhao, T. Chen, et al., Serum metabolite profiles are associated with the presence of advanced liver fibrosis in chinese patients with chronic hepatitis b viral infection, BMC Med. 18 (1) (2020) 144.

[44] M. Emu, F.B. Kamal, S. Choudhury, T.E. Alves de Oliveira, Assisting the non-invasive diagnosis of liver fibrosis stages using machine learning methods, in: Annu Int Conf IEEE Eng Med Biol. Soc. 2020, 2020, pp. 5382–5387.

[45] G.N. Ioannou, W. Tang, L.A. Beste, M.A. Tincopa, G.L. Su, T. Van, et al., Assessment of a deep learning model to predict hepatocellular carcinoma in patients with hepatitis c cirrhosis, JAMA Netw. Open. 3 (9) (2020) e2015626.

[46] W. Wei, X. Wu, J. Zhou, Y. Sun, Y. Kong, X. Yang, Noninvasive evaluation of liver fibrosis reverse using artificial neural network model for chronic hepatitis b patients, Comput. Math. Methods. Med. 2019 (2019) 7239780.

[47] M.H. Zheng, W.K. Seto, K.Q. Shi, D.K. Wong, J. Fung, I.F. Hung, et al., Artificial neural network accurately predicts hepatitis b surface antigen seroclearance, PLoS. One. 9 (6) (2014) e99422.

[48] X. Tian, Y. Chong, Y. Huang, P. Guo, M. Li, W. Zhang, et al., Using machine learning algorithms to predict hepatitis b surface antigen seroclearance, Comput. Math. Methods. Med. 2019 (2019) 6915850.

[49] S. Hashem, M. ElHefnawi, S. Habashy, M. El-Adawy, G. Esmat, W. Elakel, et al., Machine learning prediction models for diagnosing hepatocellular carcinoma with hcv-related chronic liver disease, Comput. Methods. Programs. Biomed. 196 (2020), 105551.

[50] D.S. Krakower, S. Gruber, K. Hsu, J.T. Menchaca, J.C. Maro, B.A. Kruskal, et al., Development and validation of an automated hiv prediction algorithm to identify candidates for pre-exposure prophylaxis: A modelling study, Lancet. HIV. 6 (10) (2019) e696–e704.

[51] C. Fehér, M. Plana, A. Crespo Guardo, N. Climent, L. Leal, A. Ugarte, et al., A classifier to predict viral control after antiretroviral treatment interruption in chronic hiv-1-infected patients, J. Acquir. Immune. Defic. Syndr. 83 (5) (2020) 479–485.

[52] M.A. Konerman, D. Lu, Y. Zhang, M. Thomson, J. Zhu, A. Verma, et al., Assessing risk of fibrosis progression and liver-related clinical outcomes among patients with both early stage and advanced chronic hepatitis c, PLoS. One. 12 (11) (2017) e0187344.

[53] A.D. Revell, D. Wang, R. Wood, C. Morrow, H. Tempelman, R.L. Hamers, et al., Computational models can predict response to hiv therapy without a genotype and may reduce treatment failure in different resource-limited settings, J. Antimicrob. Chemother. 68 (6) (2013) 1406–1414.

[54] G.L. Wong, V.W. Hui, Q. Tan, J. Xu, H.W. Lee, T.C. Yip, et al., Novel machine learning models outperform risk scores in predicting hepatocellular carcinoma in patients with chronic viral hepatitis, JHEP. Rep. 4 (3) (2022), 100441.

[55] Y. Cao, K. He, M. Cheng, H.Y. Si, H.L. Zhang, W. Song, et al., Two classifiers based on serum peptide pattern for prediction of hbv-induced liver cirrhosis using maldi-tof ms, Biomed. Res. Int. 2013 (2013), 814876.

[56] Y. Cao, Z.D. Hu, X.F. Liu, A.M. Deng, C.J. Hu, An mlp classifier for prediction of hbv-induced liver cirrhosis using routinely available clinical parameters, Dis. Markers. 35 (6) (2013) 653–660.

[57] N. Wang, Y. Cao, W. Song, K. He, T. Li, J. Wang, et al., Serum peptide pattern that differentially diagnoses hepatitis b virus-related hepatocellular carcinoma from liver cirrhosis, J. Gastroenterol. Hepatol. 29 (7) (2014) 1544–1550.

[58] G. Shang, A. Richardson, M.E. Gahan, S. Easteal, S. Ohms, B.A. Lidbury, Predicting the presence of hepatitis b virus surface antigen in chinese patients by pathology data mining, J. Med. Virol. 85 (8) (2013) 1334–1339.

[59] R.M. Nance, H.M. Crane, C. Ritchings, L. Rosenblatt, M. Budoff, S.R. Heckbert, et al., Differentiation of type 1 and type 2 myocardial infarctions among hiv-infected patients requires adjudication due to overlap in risk factors, AIDS Res. Hum. Retroviruses. 34 (11) (2018) 916–921.

[60] S. Kamal, J. Urata, M. Cavassini, H. Liu, R. Kouyos, O. Bugnon, et al., Random forest machine learning algorithm predicts virologic outcomes among hiv infected adults in lausanne, switzerland using electronically monitored combined antiretroviral treatment adherence, AIDS Care. 33 (4) (2021) 530–536.

[61] J.A. Muñoz-Moreno, N. Pérez-Álvarez, A. Muñoz-Murillo, A. Prats, M. Garolera, M. Jurado, et al., Classification models for neurocognitive impairment in hiv infection based on demographic and clinical variables, PLoS. One. 9 (9) (2014) e107625.

[62] X. Yang, J. Zhang, S. Chen, S. Weissman, B. Olatosi, X. Li, Utilizing electronic health record data to understand comorbidity burden among people living with hiv: A machine learning approach, Aids. 35 (Suppl 1) (2021) S39–S51.

[63] G. Nemomsa, M. Azath, Designing a predictive model for antiretroviral regimen at the antiretroviral therapy center in chiro hospital, ethiopia, J. Healthc. Eng. 2021 (2021) 1161923.

[64] I.R. König, J.D. Malley, C. Weimar, H.C. Diener, A. Ziegler, Practical experiences on the necessity of external validation, Stat. Med. 26 (30) (2007) 5499–5511.

[65] G.S. Collins, J.A. de Groot, S. Dutton, O. Omar, M. Shanyinde, A. Tajar, et al., External validation of multivariable prediction models: A systematic review of methodological conduct and reporting, BMC Med. Res. Methodol. 14 (2014) 40.

[66] M. Baker, 1,500 scientists lift the lid on reproducibility, Nature. 533 (7604) (2016) 452–454.

[67] F. Cabitza, R. Rasoini, G.F. Gensini, Unintended consequences of machine learning in medicine, Jama. 318 (6) (2017) 517–518.

[68] House of lords. Ai in the uk: Ready, willing and able? Https://publications. Parliament.Uk/pa/ld201719/ldselect/ldai/100/100.Pdf accessed 28 july 2022. 2018.

[69] J. Carthy, M.L. Minsky, N. Rochester, C.E. Shannon, A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955, AI magazine 27 (4) (2006) 12.
[70] T.J. Brinker, A. Hekler, J.S. Utikal, N. Grabe, D. Schadendorf, J. Klode, et al., Skin cancer classification using convolutional neural networks: Systematic review, J. Med. Internet. Res. 20 (10) (2018) e11936.
[71] V. Dick, C. Sinz, M. Mittlböck, H. Kittler, P. Tschandl, Accuracy of computer-aided diagnosis of melanoma: A meta-analysis, JAMA Dermatol. 155 (11) (2019) 1291–1299.
[72] Z. Obermeyer, E.J. Topol, Artificial intelligence, bias, and patients' perspectives, Lancet. 397 (10289) (2021) 2038.
[73] H. Ibrahim, X. Liu, N. Zariffa, A.D. Morris, A.K. Denniston, Health data poverty: An assailable barrier to equitable digital health care, Lancet. Digit. Health. 3 (4) (2021) e260–e265.
[74] M.A. Gianfrancesco, S. Tamang, J. Yazdany, G. Schmajuk, Potential biases in machine learning algorithms using electronic health record data, JAMA Intern. Med. 178 (11) (2018) 1544–1547.
[75] A. Holzinger, Interactive machine learning for health informatics: When do we need the human-in-the-loop? Brain. Inform. 3 (2) (2016) 119–131.
[76] A.T. Young, D. Amara, A. Bhattacharya, M.L. Wei, Patient and general public attitudes towards clinical artificial intelligence: A mixed methods systematic review, Lancet. Digit. Health. 3 (9) (2021) e599–e611.