Impact of Big Data Analytics on People's Health: Overview of Systematic Reviews and Recommendations for Future Studies

Israel Júnior Borges do Nascimento[1,2], ClinPath, PharmB; Milena Soriano Marcolino[3,4], MD, MSc, PhD; Hebatullah Mohamed Abdulazeem[5], MBBS; Ishanka Weerasekara[6,7], PhD; Natasha Azzopardi-Muscat[8], MD, MPH, MSc, PhD; Marcos André Gonçalves[9], PhD; David Novillo-Ortiz[8], MLIS, MSc, PhD

[1]School of Medicine, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

[2]Department of Medicine, School of Medicine, Medical College of Wisconsin, Wauwatosa, WI, United States

[3]Department of Internal Medicine, University Hospital, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

[4]School of Medicine and Telehealth Center, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

[5]Department of Sport and Health Sciences, Technical University Munich, Munich, Germany

[6]School of Health Sciences, Faculty of Health and Medicine, The University of Newcastle, Callaghan, Australia

[7]Department of Physiotherapy, Faculty of Allied Health Sciences, University of Peradeniya, Peradeniya, Sri Lanka

[8]Division of Country Health Policies and Systems, World Health Organization, Regional Office for Europe, Copenhagen, Denmark

[9]Department of Computer Science, Institute of Exact Sciences, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

Corresponding Author:
David Novillo-Ortiz, MLIS, MSc, PhD
Division of Country Health Policies and Systems
World Health Organization, Regional Office for Europe
Marmorej 51
Copenhagen, 2100
Denmark
Phone: 45 61614868
Email: dnovillo@who.int

Abstract

Background: Although the potential of big data analytics for health care is well recognized, evidence is lacking on its effects on public health.

Objective: The aim of this study was to assess the impact of the use of big data analytics on people's health based on the health indicators and core priorities in the World Health Organization (WHO) General Programme of Work 2019/2023 and the European Programme of Work (EPW), approved and adopted by its Member States, in addition to SARS-CoV-2–related studies. Furthermore, we sought to identify the most relevant challenges and opportunities of these tools with respect to people's health.

Methods: Six databases (MEDLINE, Embase, Cochrane Database of Systematic Reviews via Cochrane Library, Web of Science, Scopus, and Epistemonikos) were searched from the inception date to September 21, 2020. Systematic reviews assessing the effects of big data analytics on health indicators were included. Two authors independently performed screening, selection, data extraction, and quality assessment using the AMSTAR-2 (A Measurement Tool to Assess Systematic Reviews 2) checklist.

Results: The literature search initially yielded 185 records, 35 of which met the inclusion criteria, involving more than 5,000,000 patients. Most of the included studies used patient data collected from electronic health records, hospital information systems, private patient databases, and imaging datasets, and involved the use of big data analytics for noncommunicable diseases. "Probability of dying from any of cardiovascular, cancer, diabetes or chronic renal disease" and "suicide mortality rate" were the most commonly assessed health indicators and core priorities within the WHO General Programme of Work 2019/2023 and the EPW 2020/2025. Big data analytics have shown moderate to high accuracy for the diagnosis and prediction of complications of diabetes mellitus as well as for the diagnosis and classification of mental disorders; prediction of suicide attempts and behaviors; and the diagnosis, treatment, and prediction of important clinical outcomes of several chronic diseases. Confidence in the results was rated as "critically low" for 25 reviews, as "low" for 7 reviews, and as "moderate" for 3 reviews. The most frequently identified challenges were establishment of a well-designed and structured data source, and a secure, transparent, and standardized database for patient data.

Conclusions: Although the overall quality of included studies was limited, big data analytics has shown moderate to high accuracy for the diagnosis of certain diseases, improvement in managing chronic diseases, and support for prompt and real-time analyses of large sets of varied input data to diagnose and predict disease outcomes.

Trial Registration: International Prospective Register of Systematic Reviews (PROSPERO) CRD42020214048; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=214048

Introduction

Big data analytics tools handle complex datasets that traditional data processing systems cannot efficiently and economically store, manage, or process. Through the application of artificial intelligence (AI) algorithms and machine learning (ML), big data analytics has potential to revolutionize health care, supporting clinicians, providers, and policymakers for planning or implementing interventions [1], faster disease detection, therapeutic decision support, outcome prediction, and increased personalized medicine, resulting in lower-cost, higher-quality care with better outcomes [1,2].

In 2018, the World Health Organization (WHO) proposed the expedited 13th General Programme of Work (GPW13), which was approved and adopted by its 194 Member States, focusing on measurable impacts on people's health at the state level to transform public health with three core features: enhanced universal health coverage, health emergencies protection, and better health and well-being [3]. Forty-six outcome target indicators emerged from the GPW13, covering a range of health issues [3]. Big data analytics may help to support health policy decision-making, accelerate the achievement of the GPW13 core priorities and targets, and guide the roadmap for the European region based on the European Programme of Work (EPW) 2020/2025 [4,5].

Therefore, the aim of this study was to provide an overview of systematic reviews that assessed the effects of the use of big data analytics on people's health according to the WHO core features defined in the GPW13 and the EPW. We included complex reviews that assessed multiple interventions, different populations, and differing outcomes resulting from big data analytics on people's health, and identified the challenges, opportunities, and best practices for future research.

Methods

Study Design

This study was designed to provide an overview of systematic reviews in accordance with guidelines from the Cochrane Handbook for Systematic Reviews of Interventions, along with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) and the QUOROM (Quality of Reporting of Meta-analyses) guidelines [6-8]. The study protocol is published on PROSPERO (CRD42020214048).

Search Strategy

To identify records assessing the effect of big data analytics on people's health, aligned with the WHO health indicators defined in the GPW13 (Textbox 1), a comprehensive and systematic search was performed using six multidisciplinary databases from their inception to September 21, 2020. The search strategy was designed in collaboration with a senior librarian and is described in detail in Multimedia Appendix 1.

References were imported into reference management software (EndNote X9) and duplicates were removed. Unique records were uploaded onto the Covidence Platform (Veritas Health Innovation) for screening, data extraction, and quality assessment. A manual search of reference lists was performed to supplement the search.

Textbox 1. List of 46 World Health Organization health indicators defined at the Thirteenth General Programme of Work.

- Number of persons affected by disasters (per 100,000 population)
- Domestic general government health expenditure (% of general government expenditure)
- Prevalence of stunting in children under 5 (%)
- Prevalence of wasting in children under 5 (%)
- Prevalence of overweight in children under 5 (%)
- Maternal mortality ratio (per 100,000 live births)
- Proportion of births attended by skilled health personnel (%)
- Under 5 mortality rate (per 1000 live births)
- Neonatal mortality rate (per 1000 live births)
- New HIV infections (per 1000 uninfected population)
- Tuberculosis incidence (per 100,000 population)
- Malaria incidence (per 1000 population at risk)
- Hepatitis B incidence (measured by surface antigen [HBsAg] prevalence among children under 5 years)
- Number of people requiring interventions against neglected tropical diseases (NTDs)
- Probability of dying from any of cardiovascular disease (CVD), cancer, diabetes, chronic renal disease (CRD) (aged 30-70 years) (%)
- Suicide mortality rate (per 100,000 population)
- Coverage of treatment interventions for substance-use disorders (%)
- Total alcohol per capita consumption in adults aged >15 years (liters of pure alcohol)
- Road traffic mortality rate (per 100,000 population)
- Proportion of women (aged 15-49 years) having need for family planning satisfied with modern methods (%)
- Universal Health Coverage (UHC) Service Coverage Index
- Population with household expenditures on health >10% of total household expenditure or income (%)
- Mortality rate attributed to air pollution (per 100,000 population)
- Mortality rate attributed to exposure to unsafe water, sanitation, and hygiene (WASH) services (per 100,000 population)
- Mortality rate from unintentional poisoning (per 100,000 population)
- Prevalence of tobacco use in adults aged ≥15 years (%)
- Proportion of population covered by all vaccines included in national programs (diphtheria-tetanus-pertussis vaccine, measles-containing-vaccine second dose, pneumococcal conjugated vaccine) (%)
- Proportion of health facilities with essential medicines available and affordable on a sustainable basis (%)
- Density of health workers (doctors, nurse and midwives, pharmacists, dentists per 10,000 population)
- International Health Regulations capacity and health emergency preparedness
- Proportion of bloodstream infections due to antimicrobial-resistant organisms (%)
- Proportion of children under 5 years developmentally on track (health, learning, and psychosocial well-being) (%)
- Proportion of women (aged 15-49 years) subjected to violence by current or former intimate partner (%)
- Proportion of women (aged 15-49 years) who make their own decisions regarding sexual relations, contraceptive use, and reproductive health care (%)
- Proportion of population using safely managed drinking-water services (%)
- Proportion of population using safely managed sanitation services and hand-washing facilities (%)
- Proportion of population with primary reliance on clean fuels (%)
- Annual mean concentrations of fine particulate matter (PM2.5) in urban areas ($\mu g/m^3$)
- Proportion of children (aged 1-17 years) experiencing physical or psychological aggression (%)
- Vaccine coverage for epidemic-prone diseases
- Proportion of vulnerable people in fragile settings provided with essential health services (%)

- Prevalence of raised blood pressure in adults aged ≥18 years

- Effective policy/regulation for industrially produced trans-fatty acids

- Prevalence of obesity (%)

- Number of cases of poliomyelitis caused by wild poliovirus

- Patterns of antibiotic consumption at the national level

Study Selection

Peer-reviewed publications categorized as systematic reviews assessing the effects of big data analytics on any of the GPW13 and EPW health indicators and core priorities were included, regardless of language and study design. We only considered studies in which the search was performed in at least two databases, and included a description of the search strategy and the methodology used for study selection and data extraction. We only included studies that evaluated concrete relationships between the use of big data analytics and its effect on people's lives, according to the WHO strategic priorities and indicators. Along with the 46 indicators listed in Textbox 1, we also included studies evaluating the use of big data during the COVID-19 pandemic. To identify gaps, we included reviews focusing on challenges, best practices, and short- and long-term opportunities related to big data technologies. Nonsystematic reviews, primary studies, opinions, short communications, nonscientific articles, conference abstracts, and reviews with big data inappropriately defined were excluded.

Although big data analysis is capable of handling large volumes of data, rather than focusing on the data volume/size, we focused on the process that defines big data analytics, which includes the following phases [9]: (1) data selection, (2) data preprocessing, (3) data transformation, (4) AI/expert systems, and (5) understanding/assessment. The first three phases include subtasks such as: (i) feature selection and extraction, (ii) data cleaning, and (iii) data integration from multiple sources. The included studies covered all phases of the process. Title, abstract, and full-text screening were independently performed by two reviewers using the inclusion criteria. Any disagreements were resolved by a third independent investigator.

Data Extraction

The following data were extracted from the retrieved articles: publication information, journal name and impact factor, study characteristics, big data characteristics, outcomes, lessons and barriers for implementation, and main limitations. Data were individually extracted by team members and cross-checked for accuracy by a second investigator.

Assessment of Methodological Quality of Included Reviews

Two researchers independently assessed the studies using the AMSTAR 2 (A Measurement Tool to Assess Systematic Reviews 2) checklist, which includes the following critical domains, assessed in 16 items: protocol registered prior to review, adequacy of literature search, justification for excluded studies, risk of bias in included studies, appropriateness of meta-analytic methods, consideration of bias risk when interpreting results, and assessing the presence and likely impact of publication bias [10]. Appropriateness to each appraisal feature was rated as yes, no, partial yes, not applicable, or unclear. Any conflict was resolved by a third party. Studies with a review protocol tracking number were analyzed. A final summary score was given to each included record, rated as "critically low," "low," "moderate," or "high" [10].
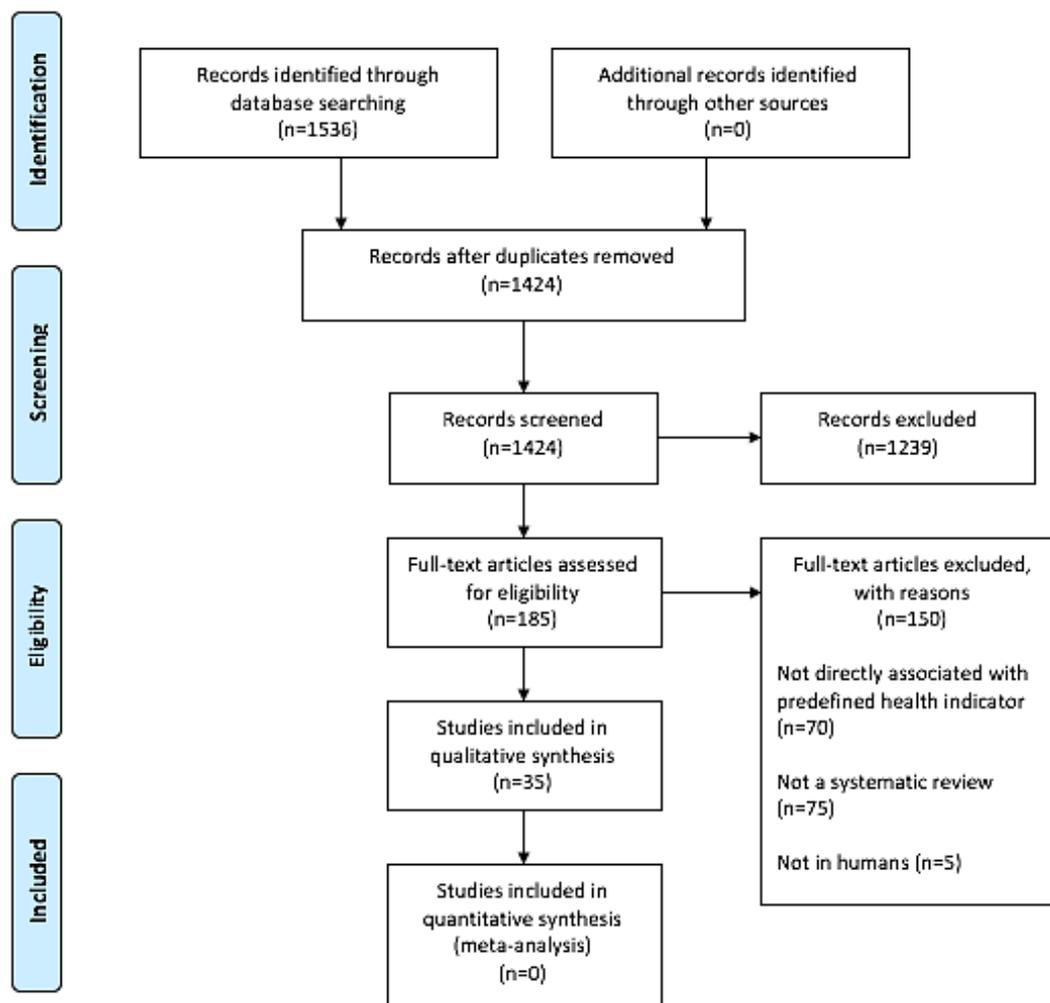
Data Synthesis

Results are reported in summary tables and through a narrative synthesis, grouping studies assessing the same disease or condition, and identifying challenges and opportunities. We also schematically represent the evidence and gaps from these reviews as an overall synthesis.

Results

Reviews Retrieved

The search retrieved 1536 publications, 112 of which were duplicates. Most of the studies were excluded after title and abstract analysis (n=1237), leaving 185 selected for full-text screening, and 35 [11-45] were ultimately included in the final analysis after applying the eligibility criteria according to the QUOROM guidelines [8] (Figure 1). Reference list screening did not retrieve any additional review. One study under "awaiting classification" could not be retrieved.

Figure 1. Flow chart of the different phases of article retrieval.

Quality of Evidence in Individual Systematic Reviews

Multimedia Appendix 2 shows the detailed results of the quality assessment of the 35 systematic reviews. Overall, most of the reviews (n=25) were rated with "critically low" confidence in the results using the AMSTAR 2 criteria, with 7 rated "low" and 3 rated as "moderate." None of the reviews achieved a "high" rating. Common methodological drawbacks included omission of prospective protocol submission or publication, inappropriate search strategy, lack of independent and dual literature screening and data extraction, absence of explanation for heterogeneity among the studies, unclear or no reasons for study exclusion, and lack of risk of bias assessment.

No standard critical appraisal tools were mentioned. Among the 12 reviews that performed any quality assessment, the Quality Assessment of Diagnostic Accuracy Studies 2 tool was used in four reviews demonstrating an overall low risk of bias [14,16,27,28], whereas other tools assessed the risk of bias in studies not specifically aiming at diagnostic accuracy features. El Idrissi et al [18] used their own quality assessment tool and Luo et al [34] used an adapted version of the Critical Appraisal Skills Programme. Appraisal of the quality of evidence aligned with the Grading of Recommendations Assessment, Development and Evaluation method was reported in only one review [17]. Many reviews did not evaluate bias.

Characteristics of Included Reviews

Summary features and main findings of the 35 systematic reviews are presented in Multimedia Appendix 3 and Multimedia Appendix 4, respectively. The included reviews were published in 34 different journals from 2007 to 2020. Most were published in English in a first-quartile journal with an impact factor ranging from 0.977 to 17.679. They covered over 2501 primary studies, involving at least 5,000,000 individuals. Only three reviews included meta-analyses, and one included a randomized clinical trial; the others were based on cohort studies.

Data Sources and Purposes of Included Studies

Many reviews included data collected from electronic medical records, hospital information systems, or any databank that used individual patient data to create predictive models or evaluate collective patterns [12,13,16-21,24-27,30,33-35,37,38,40,42-45]. Additionally, four reviews included primary studies based on imaging datasets and databanks, assessing different parameters of accuracy [15,29,31,36]. Other reviews focused on genetic databases [28,35], data from assisted reproductive technologies [30], or publicly available data [11,14,22,32]. Four studies lacked precision about the origin of the datasets used in their analysis or did not specifically use patient data in the investigation [23,37,39,41].

The purposes of the reviews varied broadly. Generally, they (1) outlined AI applications in different medical specialties; (2) analyzed features for the detection, prediction, or diagnosis of multiple diseases or conditions; or (3) pinpointed challenges and opportunities.

WHO Indicators and Core Priorities

Most of the studies assessed the effects of big data analytics on noncommunicable diseases [12-15,17,21,22,24,27,31,32,34,36, 38,40-44]. Furthermore, three reviews covered mental health, associated with the indicator "suicide mortality rate" [19,25,45]; three studies were related to the indicator "probability of dying from any of cardiovascular, cancer, diabetes, or chronic renal disease" [16,18,20,28,29]; and two studies were related to the indicator "proportion of bloodstream infections due to antimicrobial-resistant organisms" [26,33]. One study described technology use in disaster management and preparedness, covering the "number of persons affected by disasters" indicator [11], and one study was associated with the indicator "maternal mortality ratio" [30]. Overlap made precise classification into WHO health indicators challenging, and four studies could not be categorized because they mainly described challenges or opportunities in big data analytics [23,39] or because they were related to the COVID-19 pandemic [35,37].

Diseases or Conditions Assessed

Diabetes Mellitus

AI tools associated with big data analytics in the care of patients with diabetes mellitus (DM) were assessed in six reviews that included 345 primary studies [15,20,32,38,40]. Three studies reviewed AI in screening and diagnosing type 1 or type 2 DM, providing varied ranges of accuracy, sensitivity, and specificity [20,32,40]. Variables included systolic blood pressure, body mass index, triglyceride levels, and others. Two reviews covered DM control and the clinical management of DM patients [32,40]. One noted that techniques for diabetes self-management varied among the tools evaluated and reported mean values for its robust metrics [18]. The other evaluated the use of data-driven tools for predicting blood glucose dynamics and the impact of ML and data mining [20], describing the input parameters used among data-driven analysis models. However, the authors of these reviews concluded that achieving a methodologically precise predictive model is challenging and must consider multiple parameters.

Various studies assessed the ability of big data analytics to predict individual DM complications such as hypoglycemia, nephropathy, and others [15,32,38]. Supervised ML methods, decision trees, deep neural networks, random forests (RF) learning, and support vector machine (SVM) reportedly had the best outcomes for assessing complications. One review assessed deep learning–based algorithms in screening patients for diabetic retinopathy. Of 11 studies, 8 reported sensitivity and specificity of 80.3% to 100% and 84%% to 99%, respectively; two reported accuracies of 78.7% and 81%; and one reported an area under the receiver operating curve (AUC) of 0.955 [15].

Mental Health

Five reviews reported on AI, data mining, and ML in psychiatry/psychology [12,14,19,25,45], most commonly assessing these techniques in the diagnosis of mental disorders. Two reviews assessed the use of ML algorithms for predicting suicidal behaviors. High levels of risk classification accuracy (typically higher than 90%) were reported in two reviews, either for adult primary care patients or teenagers [19,25]. Although the review authors stated the potential of ML techniques in daily clinical practice, limitations were highlighted, including no external validation and reporting inconsistencies.

The use of ML algorithms for early detection of psychiatric conditions was also reported [12,45]. ML was used to develop prediagnosis algorithms for constructing risk models to signal a patient's predisposition or risk for a psychiatric/psychological health issue, for predicting a diagnosis of newly identified patients, and to differentiate mental conditions with overlapping symptomatology. For studies using structural neuroimaging to classify bipolar diseases and other diagnoses, the accuracy ranged from 52.13% to 100%, whereas studies using serum biomarkers reported an accuracy ranging from 72.5% to 77.5%.

Only one review used social media to generate analyzable data on the prevention, recognition, and support for severe mental illnesses [14]. The study included broad descriptions of ML techniques and data types for detection, diagnosis, prognosis, treatment, support, and resulting public health implications. The authors highlighted the potential for monitoring well-being, and providing an ecologically and cost-efficient evaluation of community mental health through social media and electronic records.

COVID-19

Two reviews reported the application of big data analytics and ML to better understand the current novel coronavirus pandemic [35,37]. One assessed data mining and ML techniques in diagnosing COVID-19 cases. Although the study did not define the best methodology to evaluate and detect potential cases, the authors noted an elevated frequency of decision tree models, naïve Bayes classifiers, and SVM algorithms used during previous pandemics.

Another review focused on SARS-CoV-2 immunization, and proposed that AI could expedite vaccine discovery through studying the virus's capabilities, virulence, and genome using genetic databanks. That study merged discussions of deep learning–based drug screening for predicting the interaction between protein and ligands, and using imaging results linked to AI tools for detecting SARS-CoV-2 infections.

Oncology

Four studies described the utility of ML, computerized clinical decision systems, and deep learning in oncology [24,28,29,31]. Using computerized clinical decision support systems (DSS) significantly improves process outcomes in oncology [24]. A compelling example shows that initial decisions were modified in 31% of cases after consultation of clinical DSS, which consistently resulted in improved patient management. Furthermore, implementing clinical DSS led to an average cost

reduction of US $17,000 for lung cancer patients. A remarkable workload decrease reportedly occurs when these systems are implemented in oncology facilities, leading to improved patient management and adherence to guidelines [24].

One study evaluated ML techniques in a genomic study of head and neck cancers, and found a wide range of accuracy rates (56.7% to 99.4%) based on the use of genomic data in prognostic prediction. Lastly, two studies reported accuracy levels ranging from 68% to 99.6% when using deep learning algorithms in the automatic detection of pulmonary nodules in computerized tomography images.

Cardiovascular and Neurological Conditions

Six studies described the effect of big data analytics in cardiology [13,16,21,42] and neurology [43,44]. One review assessed the use of ML techniques for predicting cardiac arrest [42]. Different variables were used as predictors among individual studies, including electrocardiographic parameters, heart rate variability, echocardiography, and others. Supervised ML techniques were most frequently applied to predict cardiac arrest events, with clear evidence of regression techniques and SVM algorithms. The authors reported a mean AUC of 0.76 for risk score development and efficiency evaluation [42].

Similarly, two studies assessed the use of intelligent systems in diagnosing acute coronary syndrome and heart failure [13,21], demonstrating high accuracy levels using several methods such as SVM, feature selection, and neural networks. These studies also described useful clinical features for creating prediction and diagnostic models, such as patient clinical data, electrocardiogram characteristics, and cardiac biomarkers.

Scores to identify patients at higher risk to develop QT-interval prolongation have been developed, and predictive analytics incorporated into clinical decision support tools have been tested for their ability to alert physicians of individuals who are at risk of or have QT-interval prolongation [16].

Regarding stroke, two systematic reviews evaluated using ML models for predicting outcomes and diagnosing cerebral ischemic events [43,44]. Generally, ML models were most frequently associated with mortality prediction, functional outcomes, neurological deterioration, and quality of life. The diagnosis of ischemic stroke was associated with similar or better comparative accuracy for detecting large vessel occlusion compared with humans, depending on the AI algorithm employed [44]. RF algorithms had 68% sensitivity and over 80% specificity compared with humans. Analyses of convolutional neural network (CNN) algorithms were limited, but systems using CNNs reported performance metrics on average 8% to 10% greater than those of ML employing RF, with up to 85% mean sensitivity for automatic large vessel occlusion detection. However, AI algorithm performance metrics used different standards, precluding objective comparison. Core and perfusion studies from RAPID-computed tomography and magnetic resonance imaging had the highest metrics for AI accuracy, above 80%, with some datasets showing 100% sensitivity to predict favorable perfusion mismatch. The authors noted several errors of AI use in diagnosing stroke [44].

Miscellaneous Conditions

Several studies reported significant improvement in disease diagnosis and event prediction using big data analytics tools, including remarkable enhancement of sepsis prediction using ML techniques [26]. Another review provided moderate evidence that ML models can reach high performance standards in detecting health care–associated infections [33].

One review focused on the diagnostic accuracy of AI systems in analyzing radiographic images for pulmonary tuberculosis, mostly referring to development instead of clinical evaluation [27]. In studies assessing accuracy, the sensitivity ranged from 56% to 97%, specificity ranged from 36% to 95%, and the AUC ranged from 78% to 99%.

One review also assessed multiple sclerosis diagnosis. Among detection methodologies, rule-based and natural language processing methods were deemed to have superior diagnostic performance based of elevated accuracy and positive predictive value [41]. This study indicates that these methods have potential impacts for early recognition of the disease, increasing quality of life, and allowing prompt pharmacological and nonpharmacological intervention.

Asthma exacerbation events and predictive models for early detection were evaluated in one review, which reached a pooled diagnostic ability of 77% (95% CI 73%-80%) [17]. Among the included studies, most models for predicting asthma development had less than 80% accuracy. None of the 42 studies modeled the reincidence of exacerbation events, and overall accuracy performance was considered inadequate. However, the authors encouraged creating models using large datasets to increase prediction accuracy levels. Logistic regression and Cox proportional hazard regression appeared to be the most commonly used methodologies. Gastric tissue disease and the usability of deep learning techniques were evaluated in one study [36]. CNN was the most common model used for gastric problem classification or detection. Additionally, residual neural network and fully convolutional network were considered to be appropriate models for disease generation, classification, and segmentation.

Two reviews analyzed the use of big data analytics and AI in public health [22,30]. One listed the impact of continuous pharmacological exposure of pregnant women, emphasizing that AI could improve popular understanding of drug effects on pregnancy, mainly through: (i) reliable clinical information disclosure, (ii) adequate scientific research design, and (iii) implementation of DSS [30]. Another review assessing the use of big data in disaster preparedness evidenced that most existing methods are qualitative, covering the response phase of the disaster chain of events [11]. The utilized tools included data originating from geographic information systems, social media interfaces, and disaster prediction modeling studies.

Challenges and Opportunities

Two systematic reviews provided narrative evaluations of the challenges of big data analytics in health care [23,39]. Evidence from these two systematic reviews, and those from the other reviews, are summarized in Textbox 2.

Textbox 2. Current challenges to use big data tools for peoples' health, and future perspectives and opportunities.

Current Challenges

1. Data structure: issues with fragmented data and incompatible or heterogeneous data formats

2. Data security: problems with privacy, lack of transparency, integrity, and inherent data structure

3. Data standardization: concerns with limited interoperability, data obtention, mining, and sharing, along with language barriers

4. Inaccuracy: issues with inconsistencies, lack of precision, and data timeliness

5. Limited awareness of big data analytics capabilities among health managers and health care professionals

6. Lack of evidence of big data analytics on the impact on clinical outcomes for peoples' health

7. Lack of skills and training among professionals to collect, process, or extract data

8. Managerial issues: ownership and government dilemma, along with data management, organizational, and financial issues

9. Regulatory, political, and legal concerns

10. Expenses with data storage and transfer

Future Perspectives and Opportunities

1. To improve the decision-making process with real-time analytics

2. To improve patient-centric health care and to enhance personalized medicine

3. To support early detection of diseases and prognostic assessment by predicting epidemics and pandemics, improving disease monitoring, implementing and tracking health behaviors, predicting patients' vulnerabilities

4. To improve data quality, structure, and accessibility by enabling the improvement of rapid acquisition of large volumes and types of data, in a transparent way, and the improvement of data error detection

5. To enable potential health care cost reduction

6. To improve quality of care by improving efficient health outcomes, reducing the waste of resources, increasing productivity and performance, promoting risk reduction, and optimizing process management

7. To provide better forms to manage population health either through early detection of diseases or establishing ways to support health policy makers.

8. To enhance fraud detection

9. To enhance health-threat detection plans by governmental entities

10. To support the creation of new research hypotheses

Discussion

This overview is the first to assess the effects of big data analytics on the prioritized WHO indicators, which offers utility for noncommunicable diseases and the ongoing COVID-19 pandemic. Although the research question focused on the impact of big data analytics on people's health, studies assessing the impact on clinical outcomes are still scarce. Most of the reviews assessed performance values using big data tools and ML techniques, and demonstrated their applications in medical practice. Most of the reviews were associated with the GPW13 indicator "probability of dying from any cardiovascular disease, cancer, diabetes, chronic respiratory disease." This indicator outranks others because of the incidence, prevalence, premature mortality, and economic impact of these diseases [46]. Similarly, many reviews were related to "people requiring interventions against noncommunicable diseases." The included reviews in this study addressed many necessary health-related tasks; however, the quality of evidence was found to be low to moderate, and studies assessing the impact on clinical outcomes are notably scarce.

The low to moderate quality of evidence suggests that big data analytics has moderate to high accuracy for the (1) diagnosis and prediction of complications of DM, (2) diagnosis of mental diseases, (3) prediction of suicidal behaviors, and (4) diagnosis of chronic diseases. Most studies presented performance values, although no study assessed whether big data analytics or ML could improve the early detection of specific diseases.

Clinical research and clinical trials significantly contribute to understanding the patterns and characteristics of diseases, as well as for improving detection of acute or chronic pathologies and to guide the development of novel medical interventions [47]. However, experimental/theoretical investigations, mathematical approaches, and computer-based studies hinge on handling sample size limitations and performing data imputation [48,49]. Computer-driven analysis can easily handle missing data, examine variable mechanisms in complex systems, and employ essential tools for exploratory evaluations using voluminous input data. Big data analytics can execute an operation on/process data within microseconds after generation of the dataset, allowing for real-time follow up [50,51]. These studies and prospective applications could generate innovative knowledge and promote actionable insights; however, adapting, validating, and translating scientific data into practical medical protocols or evaluation studies is necessary.

XSL•FO

RenderX

Many systematic reviews reported simple or inappropriate evaluation measures for the task at hand. The most common metric used to evaluate the performance of a classification predictive model is accuracy, which is calculated as the proportion of correct predictions in the test set divided by the total number of predictions that were made on the test set. This metric is easy to use and to interpret, as a single number summarizes the model capability. However, accuracy values and error rate, which is simply the complement of accuracy, are not adequate for skewed or imbalanced classification tasks (ie, when the distribution of observations in the training dataset across the classes is not equal), because of the bias toward the majority class. When the distribution is slightly skewed, accuracy can still be a useful metric; however, when the distribution is severely skewed, accuracy becomes an unreliable measure of model performance.

For instance, in a binary classification task with a distribution of (95%, 5%) for the classes (eg, healthy vs sick), a "dumb classifier" that simply chooses the class "healthy" for all instances will have 95% of accuracy in this task, although the most important issue in this task would be correctly classifying the "sick" class. Precision (also called the positive predictive value), which captures the fraction of correctly classified instances among the instances predicted for a given class (eg, "sick"); recall or sensitivity, which captures the fraction of instances of a class (eg, "sick") that were correctly classified; and F-measure, the harmonic mean of precision and recall calculated per class of interest, are more robust metrics for several practical situations. The proper choice of an evaluation metric should be carefully determined, as these indices ought to be used by regulatory bodies for screening tests and not for diagnostic reasoning [52]. The most important issue is to choose the appropriate (most robust) performance metric given the particularities of each case.

Another pitfall identified among the included reviews was the lack of reporting the precise experimental protocols used for testing ML algorithms and the specific type of replication performed.

There is no formal tool for assessing quality and risk of bias in big data studies. This is an area that is ripe for development. In Textbox 3, we summarize our recommendations for systematic reviews on the application of big data and ML for people's health based on our experience, the findings of this systematic review, and inspired by Cunha et al [53].

High variability in the results was evident across different ML techniques and approaches among the 35 reviews, even for those assessing the same disease or condition. Indeed, designing big data analysis and ML experiments involves elevated model complexity and commonly requires testing of several modeling algorithms [54]. The diversity of big data tools and ML algorithms requires proper standardization of protocols and comparative approaches. Additionally, the process of tuning the hyperparameters of the algorithms is not uniformly reported. Important characteristics essential for replicability and external validation were not frequently available. Lastly, most of the studies provide little guidance to explain the results. Without knowing *how* and *why* the models achieve their results, applicability and trust of the models in real-world scenarios are severely compromised. Therefore, we urge the testing and assessment of supervised, unsupervised, and semisupervised methodologies, with explanation and interpretation to justify the results. Moreover, we encourage hyperparameter optimization to achieve adjusted improvement of models, enhance model generalizations for untrained data, and avoid overfitting to increase predictive accuracy.

Only two published systematic reviews evaluated the impact of big data analytics on the COVID-19 pandemic. Primary studies on COVID-19 are lacking, which indicates an opportunity to apply big data and ML to this and future epidemics/pandemics [35,37]. As of November 30, 2020, many published protocols were retrieved through a standard search on PROSPERO. The titles of these review protocols showed an intention to evaluate ML tools in diagnosis and prediction, the impact of telemedicine using ML techniques, and the use of AI-based disease surveillance [55].

Although DSS are an important application of big data analytics and may benefit patient care [56-58], only two reviews assessed such systems [16,24]. One focused on predictive analytics for identifying patients at risk of drug-induced QTc interval prolongation, discussing the efficacy of a DSS that has shown evidence of reduced prescriptions for QT interval–prolonging drugs. Similarly, one study exploring the impact of DSS on quality care in oncology showed that implementing these systems might positively impact physician-prescribing behaviors, health care costs, and clinician workload.

This overview of systematic reviews updates the available evidence from multiple primary studies intersecting computer science, engineering, medicine, and public health. We used a comprehensive search strategy (performed by an information specialist) with a predefined published protocol, precise inclusion criteria, rigorous data extraction, and quality assessment of retrieved records. We avoided reporting bias through the dual and blinded examination of systematic reviews and by having one review author standardizing the extracted data.

Textbox 3. Recommendations for systematic reviews on the application of big data and machine learning for people's health.

- *Choose an appropriate evaluation measure for the task and data characteristics, and justify your choice*

Different evaluation measures such as accuracy, area under the receiver operating characteristic curve, precision, recall, and F-measure capture different aspects of the task and are influenced by data characteristics such as skewness (ie, imbalance), sampling bias, etc. Choose your measures wisely and justify your choice based on the aforementioned aspects of the task and the data.

- *Ensure the employment of appropriate experimental protocols/design to guarantee generalization of the results*

Authors should use experimental protocols based on cross-validation or multiple training/validation/test splits of the employed datasets with more than one repetition of the experimental procedure. The objective of this criterion is to analyze whether the study assesses the capacity of generalization of each method compared in the experiments. The use of a single default split of the input dataset with only one training/test split does not fit this requirement. Repetitions are essential to demonstrate the generalization of the investigated methods for multiple training and test sets, and to avoid any suspicion of a "lucky" (single) partition that favors the authors' method.

- *Properly tune, and explicitly report the tuning process and values of the hyperparameters of all compared methods*

The effectiveness of big data solutions and machine-learning methods is highly affected by the choice of the parameters of these methods (ie, parameter tuning). The wrong or improper choice of parameters may make a highly effective method exhibit very poor behavior in a given task. Ideally, the parameters should be chosen for each specific task and dataset using a partition of the training set (ie, validation), which is different from the dataset used to train and to test the model. This procedure is known as cross-validation on the training set or nested cross-validation.

Even if the tuning of all methods is properly executed, this should be explicitly reported in the paper, with the exact values (or range of values) used for each parameter and the best choices used. When the tuning information is missing or absent, it is impossible to determine whether the methods have been implemented appropriately and if they have achieved their maximum potential in a given task. It is also impossible to assess whether the comparison is fair, as some methods may have been used at their maximum capacity and others not.

- *Pay attention to the appropriate statistical tests*

Authors should employ statistical significance tests to contrast the compared strategies in their experimental evaluation. Statistical tests are essential to assess whether the performance of the analyzed methods in the sample (ie, the considered datasets) is likely to reflect, with certain confidence, their actual performance in the whole population. As such, they are key to support any claim of superiority of a particular method over others. Without such tests, the relative performance observed in the sample cannot, by any means, be extrapolated to the population. The choice of the tests should also reflect the characteristics of the data (ie, determining whether the data follow a normal distribution).

- *Make the data and code freely available with proper documentation*

One of the issues that hampers reproducibility of studies, and therefore scientific progress, is the lack of original implementation (with proper documentation) of the methods and techniques, and the unavailability of the original data used to test the methods. Therefore, it is important to make all data, models, code, documentation, and other digital artifacts used in the research available for others to reuse. The artifacts made available must be sufficient to ensure that published results can be accurately reproduced.

- *Report other dimensions of the problem such as model costs (time) and potential for explainability*

Effectiveness of the solutions, as captured by accuracy-oriented measures, is not the only dimension that should be evaluated. Indeed, if the effectiveness of the studied models is similar and sufficient for a given health-related application, other dimensions such as time efficiency (or the costs) to train and deploy (test) the models are essential to evaluate the practical applicability of such solutions. Another dimension that may influence the decision for the practical use of a big data or a machine-learning method in a real practical situation is the ability to understand why the model has produced certain outputs (ie, explainability). Solutions such as those based on neural networks may be highly effective when presented with huge amounts of data, but their training and deployment costs as well as their opaqueness may not make them the best choice for a given health-related application.

However, limitations exist. The inferior quality scores based on the AMSTAR 2 tool might reflect incomplete reporting and lack of adherence to substandardized review methods. There is neither an established bias risk tool specifically for big data or ML studies nor any systematic way of presenting the findings of such studies. Furthermore, most studies provided a narrative description of results, requiring summarization. Nevertheless, all of the reviews were inspected by most authors, and the most relevant data were condensed in the text or in descriptive tables.

Big data analytics provide public health and health care with powerful instruments to gather and analyze large volumes of heterogeneous data. Although research in this field has been growing exponentially in the last decade, the overall quality of evidence is found to be low to moderate. High variability of results was observed across different ML techniques and approaches, even for the same disease or condition. The diversity of big data tools and ML algorithms require proper standardization of protocols and comparative approaches, and the process of tuning the hyperparameters of the algorithms is not uniformly reported. Important characteristics essential for replicability and external validation were not frequently available.

Additionally, the included reviews in this systematic review addressed different health-related tasks; however, studies assessing the impact on clinical outcomes remain scarce. Thus, evidence of applicability in daily medical practice is still needed. Further studies should focus on how big data analytics impact clinical outcomes and on creating proper methodological guidelines for reporting big data/ML studies, as well as using robust performance metrics to assess accuracy.

XSL•FO

RenderX

Authors' Contributions

IJBdN, MM, MG, NAM, and DNO designed the study. HA, IW, and IJBdN performed first- and second-stage screening, and extracted the presented data. MM solved any disagreements. HA, IW, and IBdN carried out the quality assessment. IJBdN, MM, MG, and DNO drafted the manuscript and its final version. DNO and NAM are staff members of the WHO. The authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy, or views of the WHO.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search strategy used in the research.
[DOCX File , 26 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Quality assessment judgment using the AMSTAR 2 tool.
[DOCX File , 28 KB-Multimedia Appendix 2]

Multimedia Appendix 3

Main characteristics of included studies.
[DOCX File , 38 KB-Multimedia Appendix 3]

Multimedia Appendix 4

Results and limitations of included systematic reviews.
[DOCX File , 53 KB-Multimedia Appendix 4]

References

1. Ristevski B, Chen M. Big data analytics in medicine and healthcare. J Integr Bioinform 2018 May 10;15(3):20170030 [FREE Full text] [doi: 10.1515/jib-2017-0030] [Medline: 29746254]
2. Panahiazar M, Taslimitehrani V, Jadhav A, Pathak J. Empowering personalized medicine with big data and semantic web technology: promises, challenges, and use cases. Proc IEEE Int Conf Big Data 2014 Oct;2014:790-795 [FREE Full text] [doi: 10.1109/BigData.2014.7004307] [Medline: 25705726]
3. 13th General Programme of Work (GPW13) WHO Impact Framework. World Health Organization. 2019. URL: https://www.who.int/about/what-we-do/GPW13_WHO_Impact_Framework_Indicator_Metadata.pdf [accessed 2020-12-05]
4. The European Programme of Work, 2020–2025 – "United Action for Better Health". World Health Organization, regional office for Europe. 2020. URL: https://www.euro.who.int/en/health-topics/health-policy/european-programme-of-work/about-the-european-programme-of-work/european-programme-of-work-20202025-united-action-for-better-health-in-europe [accessed 2020-12-15]
5. Whitlock EP, Lin JS, Chou R, Shekelle P, Robinson KA. Using existing systematic reviews in complex systematic reviews. Ann Intern Med 2008 May 20;148(10):776-782. [doi: 10.7326/0003-4819-148-10-200805200-00010] [Medline: 18490690]
6. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
7. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011 Oct 18;343:d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
8. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. Lancet 1999 Nov 27;354(9193):1896-1900. [doi: 10.1016/s0140-6736(99)04149-5] [Medline: 10584742]

9.   Baştanlar Y, Ozuysal M. Introduction to machine learning. Methods Mol Biol 2014;1107:105-128. [doi: 10.1007/978-1-62703-748-8_7] [Medline: 24272434]

10. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ 2017 Sep 21;358:j4008 [FREE Full text] [doi: 10.1136/bmj.j4008] [Medline: 28935701]

11. Freeman JD, Blacker B, Hatt G, Tan S, Ratcliff J, Woolf TB, et al. Use of big data and information and communications technology in disasters: an integrative review. Disaster Med Public Health Prep 2019 Apr;13(2):353-367. [doi: 10.1017/dmp.2018.73] [Medline: 30047353]

12. Librenza-Garcia D, Kotzian BJ, Yang J, Mwangi B, Cao B, Pereira Lima LN, et al. The impact of machine learning techniques in the study of bipolar disorder: A systematic review. Neurosci Biobehav Rev 2017 Sep;80:538-554. [doi: 10.1016/j.neubiorev.2017.07.004] [Medline: 28728937]

13. Sprockel J, Tejeda M, Yate J, Diaztagle J, González E. Intelligent systems tools in the diagnosis of acute coronary syndromes: A systemic review. Arch Cardiol Mex 2018;88(3):178-189. [doi: 10.1016/j.acmx.2017.03.002] [Medline: 28359602]

14. Shatte ABR, Hutchinson DM, Teague SJ. Machine learning in mental health: a scoping review of methods and applications. Psychol Med 2019 Jul;49(9):1426-1448. [doi: 10.1017/S0033291719000151] [Medline: 30744717]

15. Nielsen KB, Lautrup ML, Andersen JKH, Savarimuthu TR, Grauslund J. Deep learning-based algorithms in screening of diabetic retinopathy: a systematic review of diagnostic performance. Ophthalmol Retina 2019 Apr;3(4):294-304. [doi: 10.1016/j.oret.2018.10.014] [Medline: 31014679]

16. Tomaselli Muensterman E, Tisdale JE. Predictive analytics for identification of patients at risk for QT interval prolongation: a systematic review. Pharmacotherapy 2018 Aug;38(8):813-821. [doi: 10.1002/phar.2146] [Medline: 29882591]

17. Bridge J, Blakey JD, Bonnett L. A systematic review of methodology used in the development of prediction models for future asthma exacerbation. BMC Med Res Methodol 2020 Feb 05;20(1):22 [FREE Full text] [doi: 10.1186/s12874-020-0913-7] [Medline: 32024484]

18. El Idrissi T, Idri A, Bakkoury Z. Systematic map and review of predictive techniques in diabetes self-management. Int J Inf Manage 2019 Jun;46:263-277. [doi: 10.1016/j.ijinfomgt.2018.09.011]

19. Bernert RA, Hilberg AM, Melia R, Kim JP, Shah NH, Abnousi F. Artificial intelligence and suicide prevention: a systematic review of machine learning investigations. Int J Environ Res Public Health 2020 Aug 15;17(16):5929 [FREE Full text] [doi: 10.3390/ijerph17165929] [Medline: 32824149]

20. Woldaregay AZ, Årsand E, Walderhaug S, Albers D, Mamykina L, Botsis T, et al. Data-driven modeling and prediction of blood glucose dynamics: Machine learning applications in type 1 diabetes. Artif Intell Med 2019 Jul;98:109-134 [FREE Full text] [doi: 10.1016/j.artmed.2019.07.007] [Medline: 31383477]

21. Tripoliti EE, Papadopoulos TG, Karanasiou GS, Naka KK, Fotiadis DI. Heart failure: diagnosis, severity estimation and prediction of adverse events through machine learning techniques. Comput Struct Biotechnol J 2017;15:26-47 [FREE Full text] [doi: 10.1016/j.csbj.2016.11.001] [Medline: 27942354]

22. Yin Z, Sulieman LM, Malin BA. A systematic literature review of machine learning in online personal health data. J Am Med Inform Assoc 2019 Jun 01;26(6):561-576 [FREE Full text] [doi: 10.1093/jamia/ocz009] [Medline: 30908576]

23. Kruse CS, Goswamy R, Raval Y, Marawi S. Challenges and opportunities of big data in health care: a systematic review. JMIR Med Inform 2016 Nov 21;4(4):e38 [FREE Full text] [doi: 10.2196/medinform.5359] [Medline: 27872036]

24. Klarenbeek SE, Weekenstroo HH, Sedelaar JM, Fütterer JJ, Prokop M, Tummers M. The effect of higher level computerized clinical decision support systems on oncology care: a systematic review. Cancers (Basel) 2020 Apr 22;12(4):1032 [FREE Full text] [doi: 10.3390/cancers12041032] [Medline: 32331449]

25. Burke TA, Ammerman BA, Jacobucci R. The use of machine learning in the study of suicidal and non-suicidal self-injurious thoughts and behaviors: A systematic review. J Affect Disord 2019 Feb 15;245:869-884. [doi: 10.1016/j.jad.2018.11.073] [Medline: 30699872]

26. Fleuren LM, Klausch TLT, Zwager CL, Schoonmade LJ, Guo T, Roggeveen LF, et al. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. Intensive Care Med 2020 Mar;46(3):383-400 [FREE Full text] [doi: 10.1007/s00134-019-05872-y] [Medline: 31965266]

27. Harris M, Qi A, Jeagal L, Torabi N, Menzies D, Korobitsyn A, et al. A systematic review of the diagnostic accuracy of artificial intelligence-based computer programs to analyze chest x-rays for pulmonary tuberculosis. PLoS One 2019;14(9):e0221339 [FREE Full text] [doi: 10.1371/journal.pone.0221339] [Medline: 31479448]

28. Patil S, Habib Awan K, Arakeri G, Jayampath Seneviratne C, Muddur N, Malik S, et al. Machine learning and its potential applications to the genomic study of head and neck cancer-A systematic review. J Oral Pathol Med 2019 Oct;48(9):773-779. [doi: 10.1111/jop.12854] [Medline: 30908732]

29. Pehrson L, Nielsen M, Ammitzbøl Lauridsen C. Automatic pulmonary nodule detection applying deep learning or machine learning algorithms to the LIDC-IDRI database: a systematic review. Diagnostics (Basel) 2019 Mar 07;9(1):29 [FREE Full text] [doi: 10.3390/diagnostics9010029] [Medline: 30866425]

30. Davidson L, Boland MR. Enabling pregnant women and their physicians to make informed medication decisions using artificial intelligence. J Pharmacokinet Pharmacodyn 2020 Aug;47(4):305-318 [FREE Full text] [doi: 10.1007/s10928-020-09685-1] [Medline: 32279157]

31. Li D, Mikela Vilmun B, Frederik Carlsen J, Albrecht-Beste E, Ammitzbøl Lauridsen C, Bachmann Nielsen M, et al. The performance of deep learning algorithms on automatic pulmonary nodule detection and classification tested on different datasets that are not derived from LIDC-IDRI: a systematic review. Diagnostics (Basel) 2019 Nov 29;9(4):207 [FREE Full text] [doi: 10.3390/diagnostics9040207] [Medline: 31795409]

32. Chaki J, Thillai Ganesh S, Cidham S, Ananda Theertan S. Machine learning and artificial intelligence based diabetes mellitus detection and self-management: a systematic review. J King Saud Univ Comput Inf Sci 2020 Jul:In press [FREE Full text] [doi: 10.1016/j.jksuci.2020.06.013]

33. Scardoni A, Balzarini F, Signorelli C, Cabitza F, Odone A. Artificial intelligence-based tools to control healthcare associated infections: A systematic review of the literature. J Infect Public Health 2020 Aug;13(8):1061-1077 [FREE Full text] [doi: 10.1016/j.jiph.2020.06.006] [Medline: 32561275]

34. Luo G, Nkoy FL, Stone BL, Schmick D, Johnson MD. A systematic review of predictive models for asthma development in children. BMC Med Inform Decis Mak 2015 Nov 28;15:99 [FREE Full text] [doi: 10.1186/s12911-015-0224-9] [Medline: 26615519]

35. Kannan S, Subbaram K, Ali S, Kannan H. The role of artificial intelligence and machine learning techniques: race for COVID-19 vaccine. Arch Clin Infect Dis 2020 May 10;15(2):e103232. [doi: 10.5812/archcid.103232]

36. Gonçalves WGE, Dos Santos M, Lobato FMF, Ribeiro-Dos-Santos Â, de Araújo GS. Deep learning in gastric tissue diseases: a systematic review. BMJ Open Gastroenterol 2020 Mar 26;7(1):e000371 [FREE Full text] [doi: 10.1136/bmjgast-2019-000371] [Medline: 32337060]

37. Albahri AS, Hamid RA, Alwan JK, Al-Qays ZT, Zaidan AA, Zaidan BB, et al. Role of biological data mining and machine learning techniques in detecting and diagnosing the novel coronavirus (COVID-19): a systematic review. J Med Syst 2020 May 25;44(7):122 [FREE Full text] [doi: 10.1007/s10916-020-01582-x] [Medline: 32451808]

38. Kavakiotis I, Tsave O, Salifoglou A, Maglaveras N, Vlahavas I, Chouvarda I. Machine learning and data mining methods in diabetes research. Comput Struct Biotechnol J 2017;15:104-116 [FREE Full text] [doi: 10.1016/j.csbj.2016.12.005] [Medline: 28138367]

39. Galetsi P, Katsaliaki K, Kumar S. Values, challenges and future directions of big data analytics in healthcare: A systematic review. Soc Sci Med 2019 Nov;241:112533. [doi: 10.1016/j.socscimed.2019.112533] [Medline: 31585681]

40. Abhari S, Niakan Kalhori SR, Ebrahimi M, Hasannejadasl H, Garavand A. Artificial intelligence applications in type 2 diabetes mellitus care: focus on machine learning methods. Healthc Inform Res 2019 Oct;25(4):248-261 [FREE Full text] [doi: 10.4258/hir.2019.25.4.248] [Medline: 31777668]

41. Arani LA, Hosseini A, Asadi F, Masoud SA, Nazemi E. Intelligent computer systems for multiple sclerosis diagnosis: a systematic review of reasoning techniques and methods. Acta Inform Med 2018 Dec;26(4):258-264 [FREE Full text] [doi: 10.5455/aim.2018.26.258-264] [Medline: 30692710]

42. Layeghian Javan S, Sepehri MM, Aghajani H. Toward analyzing and synthesizing previous research in early prediction of cardiac arrest using machine learning based on a multi-layered integrative framework. J Biomed Inform 2018 Dec;88:70-89 [FREE Full text] [doi: 10.1016/j.jbi.2018.10.008] [Medline: 30389440]

43. Wang W, Kiik M, Peek N, Curcin V, Marshall IJ, Rudd AG, et al. A systematic review of machine learning models for predicting outcomes of stroke with structured data. PLoS One 2020;15(6):e0234722 [FREE Full text] [doi: 10.1371/journal.pone.0234722] [Medline: 32530947]

44. Murray NM, Unberath M, Hager GD, Hui FK. Artificial intelligence to diagnose ischemic stroke and identify large vessel occlusions: a systematic review. J Neurointerv Surg 2020 Feb;12(2):156-164. [doi: 10.1136/neurintsurg-2019-015135] [Medline: 31594798]

45. Alonso SG, de la Torre-Díez I, Hamrioui S, López-Coronado M, Barreno DC, Nozaleda LM, et al. Data mining algorithms and techniques in mental health: a systematic review. J Med Syst 2018 Jul 21;42(9):161. [doi: 10.1007/s10916-018-1018-2] [Medline: 30030644]

46. GBD 2016 Disease Injury Incidence Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 2017 Sep 16;390(10100):1211-1259 [FREE Full text] [doi: 10.1016/S0140-6736(17)32154-2] [Medline: 28919117]

47. Nass SJ, Levit LA, Gostin LO. Rule I of M (US) C on HR and the P of HITHP. In: The value, importance, and oversight of health research. Washington, DC: National Academy of Sciences; 2009:334.

48. Sivarajah U, Kamal MM, Irani Z, Weerakkody V. Critical analysis of Big Data challenges and analytical methods. J Bus Res 2017 Jan;70:263-286. [doi: 10.1016/j.jbusres.2016.08.001]

49. Cassidy R, Singh NS, Schiratti P, Semwanga A, Binyaruka P, Sachingongu N, et al. Mathematical modelling for health systems research: a systematic review of system dynamics and agent-based models. BMC Health Serv Res 2019 Nov 19;19(1):845 [FREE Full text] [doi: 10.1186/s12913-019-4627-7] [Medline: 31739783]

50. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. Lancet Oncol 2019 May;20(5):e262-e273. [doi: 10.1016/S1470-2045(19)30149-4] [Medline: 31044724]

XSL•FO
RenderX

51. Lee S, Mohr NM, Street WN, Nadkarni P. Machine learning in relation to emergency medicine clinical and operational scenarios: an overview. West J Emerg Med 2019 Mar;20(2):219-227 [FREE Full text] [doi: 10.5811/westjem.2019.1.41244] [Medline: 30881539]

52. Llewelyn H. Sensitivity and specificity are not appropriate for diagnostic reasoning. BMJ 2017 Sep 06;358:j4071. [doi: 10.1136/bmj.j4071] [Medline: 28877897]

53. Cunha W, Mangaravite V, Gomes C, Canuto S, Resende E, Nascimento C, et al. On the cost-effectiveness of neural and non-neural approaches and representations for text classification: A comprehensive comparative study. Inf Process Manage 2021 May;58(3):102481. [doi: 10.1016/j.ipm.2020.102481]

54. Deo RC. Machine learning in medicine. Circulation 2015 Nov 17;132(20):1920-1930 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.115.001593] [Medline: 26572668]

55. National Institute for Health Research: PROSPERO - International Prospective Register of Systematic Reviews. University of York Centre for Reviews and Dissemination. URL: https://www.crd.york.ac.uk/prospero/ [accessed 2020-12-04]

56. Rumsfeld JS, Joynt KE, Maddox TM. Big data analytics to improve cardiovascular care: promise and challenges. Nat Rev Cardiol 2016 Jun;13(6):350-359. [doi: 10.1038/nrcardio.2016.42] [Medline: 27009423]

57. Tabary MY, Memariani A, Ebadati E. Chapter 3 - Developing a decision support system for big data analysis and cost allocation in national healthcare. In: Dey N, Ashour AS, Bhat C, Fong SJ, editors. Healthcare data analytics and management. Cambridge, MA: Academic Press; 2019:89-109.

58. Dagliati A, Tibollo V, Sacchi L, Malovini A, Limongelli I, Gabetta M, et al. Big data as a driver for clinical decision support systems: a learning health systems perspective. Front Digit Humanit 2018 May 1;5:8. [doi: 10.3389/fdigh.2018.00008]

Abbreviations

AI: artificial intelligence
AUC: area under the receiver operating characteristic curve
AMSTAR 2: A Measurement Tool to Assess Systematic Reviews 2
CNN: convolutional neural network
DM: diabetes mellitus
DSS: decision support system
EPW: European Programme of Work
GPW13: Thirteenth General Programme of Work
ML: machine learning
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-analyses
QUOROM: Quality of Reporting of Meta-analyses
RF: random forest
SVM: support vector machine
WHO: World Health Organization

XSL•FO

RenderX