PLOS ONE

RESEARCH ARTICLE

A systematic review of clinical health conditions predicted by machine learning diagnostic and prognostic models trained or validated using real-world primary health care data

Hebatullah Abdulazeem[1]*, **Sera Whitelaw**[2], **Gunther Schauberger**[1], **Stefanie J. Klug**[1]

1 Chair of Epidemiology, Department of Sport and Health Sciences, Technical University of Munich (TUM), Munich, Germany, **2** Faculty of Medicine and Health Sciences, McGill University, Montreal, Quebec, Canada

* hebatullah.abdulazeem@tum.de

Abstract

With the advances in technology and data science, machine learning (ML) is being rapidly adopted by the health care sector. However, there is a lack of literature addressing the health conditions targeted by the ML prediction models within primary health care (PHC) to date. To fill this gap in knowledge, we conducted a systematic review following the PRISMA guidelines to identify health conditions targeted by ML in PHC. We searched the Cochrane Library, Web of Science, PubMed, Elsevier, BioRxiv, Association of Computing Machinery (ACM), and IEEE Xplore databases for studies published from January 1990 to January 2022. We included primary studies addressing ML diagnostic or prognostic predictive models that were supplied completely or partially by real-world PHC data. Studies selection, data extraction, and risk of bias assessment using the prediction model study risk of bias assessment tool were performed by two investigators. Health conditions were categorized according to international classification of diseases (ICD-10). Extracted data were analyzed quantitatively. We identified 106 studies investigating 42 health conditions. These studies included 207 ML prediction models supplied by the PHC data of 24.2 million participants from 19 countries. We found that 92.4% of the studies were retrospective and 77.3% of the studies reported diagnostic predictive ML models. A majority (76.4%) of all the studies were for models' development without conducting external validation. Risk of bias assessment revealed that 90.8% of the studies were of high or unclear risk of bias. The most frequently reported health conditions were diabetes mellitus (19.8%) and Alzheimer's disease (11.3%). Our study provides a summary on the presently available ML prediction models within PHC. We draw the attention of digital health policy makers, ML models developer, and health care professionals for more future interdisciplinary research collaboration in this regard.

OPEN ACCESS

Citation: Abdulazeem H, Whitelaw S, Schauberger G, Klug SJ (2023) A systematic review of clinical health conditions predicted by machine learning diagnostic and prognostic models trained or validated using real-world primary health care data. PLoS ONE 18(9): e0274276. https://doi.org/10.1371/journal.pone.0274276

Editor: Ágnes Vathy-Fogarassy, University of Pannonia: Pannon Egyetem, HUNGARY

Received: August 23, 2022

Accepted: August 29, 2023

Published: September 8, 2023

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pone.0274276

Data Availability Statement: All relevant data are within the paper and its Supporting information files.

Funding: The authors received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Primary health care (PHC) is considered the gatekeeper, where health education and promotion are provided, non-life-threatening health conditions are diagnosed and treated, and chronic diseases are managed [1]. This form of health maintenance, which aims to provide constant access to high-quality care and comprehensive services, is defined and called for by the World Health Organization (WHO) global vision for PHC [2]. Clinicians' skills and experience and the further continuing professional development are fundamental to achieve these PHC aims [3]. Additional health care improvement can be achieved by capitalizing on digital health and AI technologies.

With the high number of patients visiting PHC and the emergence of electronic health records, substantial amounts of data are generated on daily basis. A wide spectrum of data analytics exist to utilize such data; however, meaningful interpretation of large complicated data may not be adequately handled by traditional data analytics [4]. Tools that could more accurately predict diseases incidence and progression and offer advice on adequate treatment could improve the decision-making process. Machine Learning (ML), a subtype of Artificial Intelligence (AI), provides methods to productively mine this large amount of data such as predictive models that potentially forecast and predict diseases occurrence and progression [5]. The variety of ML prediction models' characteristics provide broader opportunities to support the healthcare practice.

Integrating PHC with updated technologies allows for the coordination of numerous disciplines and views. Integrating PHC with such technologies allows for improvements in health care, which may include patient care outcomes and productivity and efficiency within health care facilities [5, 6]. ML models have been developed in health research–most significantly in the last decade—to predict the incidence of diabetes, cancers, and recently COVID-19 pandemic related illness from health records [7]. A systematic overview of 35 studies published in 2021 investigated the existing literature of AI/ML, but exclusively in relation to WHO indicators [8]. Other literature and scoping reviews examined AI/ML in relation to certain health conditions, such as HIV [9], hypertension [10], and diabetes [11]. Other systematic reviews targeted specific health conditions across multiple health sectors, such as pregnancy care [12], melanoma [13], stroke [14], and diabetes [15]. However, reviews investigating PHC specifically have been fewer [16, 17]. It has been reported that research on ML for PHC stands at an early stage of maturity [17]. Similar to ours, a recently published protocol of a systematic review addressing the performance of ML prediction models in multiple different medical fields was published [18]. However, this protocol does not focus specifically on primary care and its search is limited to the years 2018 and 2019. Hence, the current literature is not enough to identify what diseases are targeted by ML prediction models within real-world PHC. Furthermore, literature investigating the validity and the potential impact of such models are not abundant. To direct the focus toward this gap, we conducted this systematic review to encompass the health conditions predicted through using ML models within PHC settings.

Materials and methods

We conducted a systematic review in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [19] and the CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS) [20]. The protocol for our review was registered on PROSPERO CRD42021264582 [21].

Search strategy and selection criteria

A comprehensive and systematic search was performed covering multidisciplinary databases: 1. Cochrane Library, 2. Elsevier (including ScienceDirect, Scopus, and Embase), 3. PubMed, 4.

Web of Science (including nine databases), 5. BioRxiv and MedRxiv, 6. Association for Computer Machinery (ACM) Digital Library, and 7. Institute of Electrical and Electronics Engineers (IEEE) Xplore Digital Library.

To find potentially relevant studies, we searched literature with the last updated search on January 4, 2022, back to January 1, 1990. The utilized search terms included "machine learning", "artificial intelligence", "deep learning", and "primary health care". Boolean operators and symbols were adapted to each literature database. Hand searches of citations of relevant reviews and a cross-reference check of the retrieved articles were also performed. Conference abstracts and gray literature searches were conducted using the available features of some databases. The full search strategy for all the electronic databases is presented in S1 File. A reference management software (EndNote X9) was used to import references and to remove duplicates.

The inclusion criteria were as follows: primary research articles (peer-reviewed, preprint, or abstract) without language restriction, studies reporting AI, DL or ML prediction models for any health condition within PHC settings, and using real-world PHC data, either exclusive or linked to other health care data. We directed our focus toward these supervised ML models (random forest, support vector machine (SVM), boosting models, decision tree, naïve bias, least absolute shrinkage and selection operator (LASSO), and k-nearest neighbors) and the neural networks.

Literature screening, data collection and statistical analysis

Title and abstract screening for all records were conducted independently by two researchers through the Rayyan platform [22]. Discrepancies were resolved by discussion. All studies that met the eligibility criteria were included in the systematic review. The process of data extraction was performed by two authors. Items and definitions of extracted data is presented in Table 1.

Health conditions extracted were categorized according to the International Classification of Diseases (ICD)-10 version 2019 [23]. This coding system was selected because it is applied

Table 1. Items and definitions for data extraction.

Item Extracted	Definition
Meta-data	First author and year of publication
Study Type	According to CHARMS guidelines [20], the types of a prediction modelling studies are: 1. model development without external validation, 2. model development with external validation, or 3. external validation of a predeveloped model with or without model update. The included studies are presented in the Results section in three tables categorized according to these three types.
Study design	Design of the included studies.
Models purpose	1. Incident diagnostic (occurrence probability of a disorder), 2. Prevalent diagnostic (identifying overlooked cases), or 3. Prognostic (occurrence probability of a future event).
Country	Countries, from which health data were collected to train, test, or validate the models.
Source of data	This represents the source of the health data used to train, test, or validate the model. 1. PHC (Data exclusively originated from a PHC settings) 2. Linked data (PHC data linked to other data sources, such as secondary or tertiary health care)
Sample size	Number of the population, whom health data were used to train, test, or validate the models.
Time span of data	Time period, in which the health care data used for modelling were originally available in the health care system.
Health condition	Health condition addressed in the included studies.

https://doi.org/10.1371/journal.pone.0274276.t001

by at least 120 countries across the globe [24]. Considering the countries that apply different coding systems, we used the explicit names of the health condition mentioned in the included studies included to match them to the closest ICD-10 codes.

Descriptive statistics of the extracted data was calculated. The overall number of populations was calculated with considering the potential overlap between the included datasets. This overlap assessment was contemplated based on similarity of source of data, time span of data within each included study, the targeted health condition and the inclusion and exclusion criteria of the participants. The quantitative results were calculated using Microsoft Excel.

Risk of bias and applicability assessment

The 'Prediction model study Risk Of Bias Assessment Tool' (PROBAST) was used to assess the risk of bias and concerns about the applicability of the included studies [25]. The four domains of this tool, which are participants, predictors, outcome, and analysis were addressed. The overall judgement for the risk of bias evaluation and concern of applicability of the prediction models in PROBAST is 'low,' 'high,' or 'unclear.' In cases when all domains were graded 'low' risk of bias, assessment of 'models developed without external validation' was downgraded to 'high' risk of bias even if all the four domains were of low risk of bias, unless the model's development was based on an exceptionally large sample size and included some form of internal validation. External validation was considered if the model was at least validated using a dataset from a later time point in the same data source (temporal external validation) or using a different dataset from inside or outside the source country (geographical or broad external validation, respectively) [20]. Results of risk of bias and concern of applicability assessments were presented in a color-coded graph.

Results

Our search strategy yielded 23,045 publications. After duplicate removal, 19,280 publication abstracts were screened, and 167 publications were eligible for full text screening. A total of 106 publications met our inclusion criteria (Fig 1). A list of the excluded studies with the justification of exclusion is presented (S1 Table). The results of the data extracted in this review are presented in the following subsections: geographical and chronological characteristics of the included studies, studies' type and design, and the ML models addressed, and (frequency of) health conditions investigated.

Geographical and chronological characteristics

The earliest included study was published in 2002 [26], with the most publications occurring over the past four years. Most (77.3%, n = 82/106) of the publications were published between 2018–2021 (Fig 2). The United States of America (US) and the United Kingdom (UK) were reported in 57.1% of the included publications. While the 106 included publications reported countries 126 times, the US was reported 41 times and the UK 31 times. Usage of exclusive real-world PHC data for modelling was reported in 77.7% (n = 115 of 148 counts of data sources) across the studies. The remaining 22.3% of the PHC data sources were linked to different data sources, such as health insurance claims, cancer registries, secondary or tertiary health care, or administrative data. In the US, data were obtained mainly from PHC centers. In contrast, the most common source of the UK data were the Clinical Practice Research Datalink (CPRD), which is the largest patients' data registry in the UK [27]. The overall time span of health data across the studies ranged from 1982 [28] to 2020 [29]. The individual time span of the included studies varied between 2 months to 28 years. Sample sizes across the included studies ranged from 75 [30] to around 4 million [31] participants. The total number of the

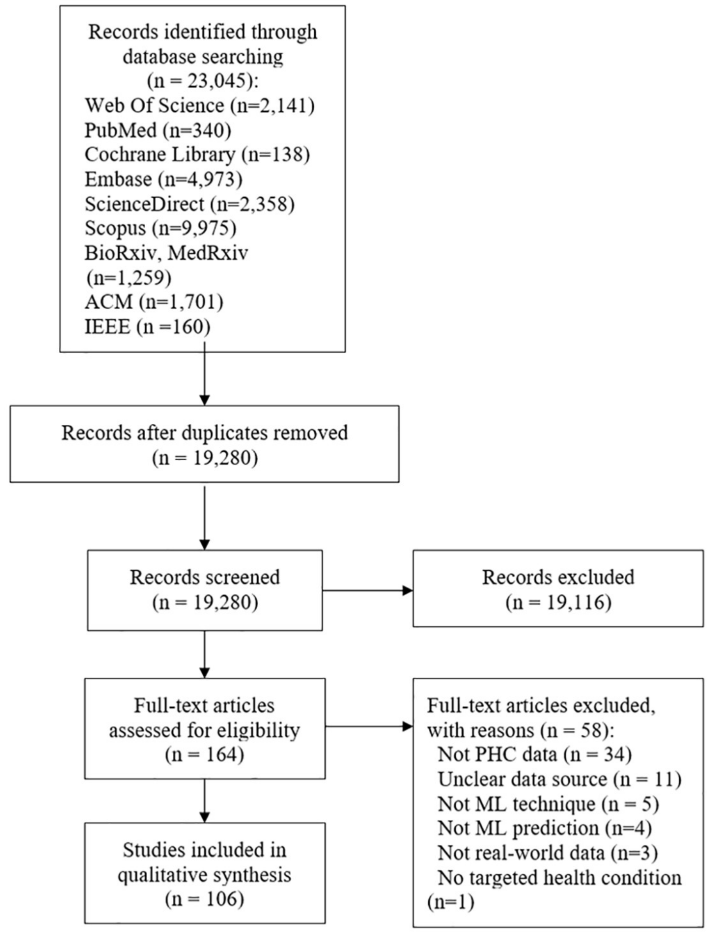

Fig 1. Prisma flow diagram.

https://doi.org/10.1371/journal.pone.0274276.g001

populations within all the included studies was of 23.2 million. After correcting the potential overlaps, the total number of unique populations was reduced to be 22.7 million.

Studies type and design, and ML models

The main type of the included studies was prediction models development without external validations (76.5%, n = 81 of 106). Of the remaining 25 studies, 13 studies (12.2%) developed and externally validated the models, and 12 studies (10.3%) externally validated previously existing models. Temporal validation [30, 32–36], geographical validation [37, 38], and using different population sample validations [39–44] were reported but none of these studies reported updating the assessed model.

All of the included studies were observational in design. Apart from 8 prospective studies, 92.4% (n = 98 of 106) of the studies were retrospective in design. Of the retrospective studies, 63 were retrospective cohorts. The other reported study designs were case control (n = 29), nested case control (n = 3), and cross sectional (n = 3). The purpose of the models reported was diagnostic in 77.3% (n = 82 of 106) of the studies, either incident (n = 62 of 82) or prevalent (n = 20 of 82). The remaining 23.5% (n = 25 of 106), including one study with two purposes of the models [45]) predicted prognosis of health conditions, such as remission, improvement, complications, hospitalization, or mortality. Despite all studies included used

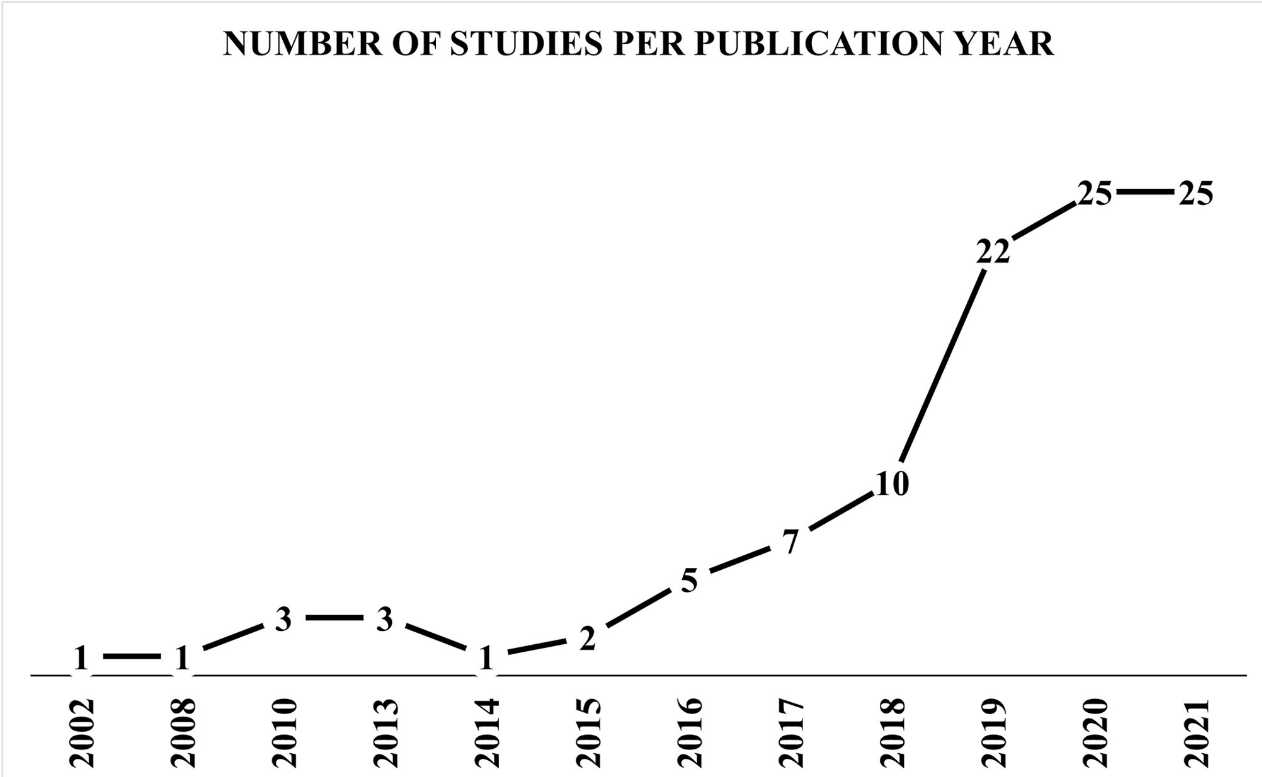

Fig 2. Number of studies for publication years.

https://doi.org/10.1371/journal.pone.0274276.g002

real-world patients' data to develop and/or validate the ML models, four studies reported applying the models develop in real-world primary health care settings [46–48].

Within the 106 included publications, 207 models were developed and/or validated. The most frequently used type of ML was supervised learning 83.1% (n = 172 of 207 models across the included studies). These supervised ML models were identified as follows: random forest (n = 58), SVM (n = 30), boosting models such as extreme, light, and adaptive boosting (n = 28), decision tree (n = 25), and others such as naïve bias, k-nearest neighbors, and LASSO (n = 31). Deep learning techniques, such as neural networks, were reported 35 times (16.9%, of 207models), either exclusively or in comparison to other supervised ML models. Supplementary table (S2 Table) presents advantages and disadvantages of these models in addition to further descriptive results of our included studies. The most frequently reported evaluation approach of models' performance was the area under the receiver operating characteristic curve (AUROC), which was reported as "good" to "moderate" models performance in 62 studies. One study reported the performance measures using decision analysis curve [49]. Other evaluation approaches were reported across the included studies, such as calculating sensitivity, specificity, predictive values, and accuracy.

The data used to develop the models were called predictors, features, or variables across the included studies. These data were mostly textual. Demographic characteristics and clinical picture of the health conditions were the most frequently found data. Medications, comorbidities, and blood tests performed within primary care unit were reported. Data, such as blood test results and imaging results performed within secondary and tertiary health care were additionally reported in some of the individual studies. Referral documentation and clinical notes taken by health care personnel were also reported. Five studies used the natural language processing (NLP) technique to handle free text clinical notes [40, 45, 50–52].

Tables 2–4 present an overview of the included studies characteristics based on the type of the study. They are grouped according to the ICD-10 classification and ordered alphabetically within each classification. A quantitative panel summary of all the included studies is also provided (S1 Panel).

Health conditions

Out of the 22 classifications of the ICD-10, 11 classifications were addressed in the included studies. Frequently reported classifications were the endocrine, nutritional, and metabolic diseases classification (ICD-10: Class E00-E90) (n = 27 studies of 106, 25.5%), circulatory system diseases (ICD-10: Class I00-I99) (n = 23, 21.7%), and the mental and behavioral disorders classification (ICD-10: Class F00-F99) (n = 21, 19.9%). Diseases of the respiratory system classifications (ICD-10: Class J00-J99) and neoplasms (ICD-10: Class C00-C97) were addressed in (n = 10, 9.4% and n = 8, 7.5% respectively). 16% (n = 17) of the included studies investigated other health conditions (ICD 10: Classes G00-G99, K00-K93, M00-M99, N00-N99, O00-O99, and X60-X84).

Endocrine, nutritional, and metabolic diseases (E00-E90). In 27 studies addressing this classification [31, 34, 39, 46, 49, 50, 52, 72–86, 119, 127–130], populations involved were from 12 countries, mainly the US (41.9%). The studies were published since 2008 with the highest number of studies in 2019 (38.7%). 81% of the included studies reported the development and/or training of the proposed models using exclusive primary health care data of a total number of 4.2 million participants. Data were extracted from different sources covering a time span of six months to 23 years. Four health conditions were identified, namely diabetes mellitus (E10, E11) with/without complications (n = 21), familial hypercholesterolemia (E78) (n = 3), childhood obesity (E66) (n = 2), and primary aldosteronism (E26) (n = 1). Incident diagnostic prediction was the most commonly reported outcome (42%). Prevalent diagnostic and prognostic prediction were 32% and 26% respectively. Diabetic retinopathy was the most common complication (n = 5 of 21 related diabetes mellitus studies) reported. Diabetic foot was investigated in one study [50]. Two studies investigated prognostic predictive modelling of the short- and long-term levels of HbA1c after insulin treatment [49, 83].

Mental and behavioral disorder (F00–F99). In 21 studies addressing six health conditions [28, 30, 35, 92–107, 120, 121, 133], the populations were from eight countries, mainly the US and the UK (n = 13). These 21 studies were published since 2013 with the highest number published in 2020 (44.4%). Data were collected from different data sources with time span of data from one year to 28 years. Alzheimer's disease (F00) was addressed in 12 studies for mostly incident or prevalent diagnosis, apart from three studies. Depression (F32) was tackled in three studies, one of which predicted depression prognosis within two years [92]. Psychosis (F29) [35] and anxiety (F41) in cancer survivors seeking care in PHC [97] were addressed in one study each. Lastly, one study used PHC data to predict any mental disorder using different ML models [104].

Circulatory and respiratory health conditions (I00-I99 and J00-J99). In 33 studies, populations involved were from 11 countries, mainly the US and the UK. The included studies were published since 2010 with the highest number in both groups published in 2020 (30.8%). Data were extracted from the different data sources over time span one month to 23 years. Six circulatory health conditions were identified in 23 studies [29, 36, 37, 40, 45, 53–70]. These conditions were hypertension (I10-I15) (n = 5), heart failure (I50) (n = 5), atrial fibrillation (I48) (n = 2), stroke (I64) (n = 2), atherosclerosis (I70) (n = 1), myocardial infarction (I21)

Table 2. Overview of the included studies with the type of ML prediction models development without conducting external validation (n = 81).

Study	Study design	Models purpose	Country	Source of data	Sample size	Time span of data	Health condition
Circulatory System Diseases							
Chen *et al*. 2019 [53]	Retro. nested case control	Incident diagnostic	United States	PHC	34,502	05/2000-05/2013	Heart failure
Choi *et al*. 2017 [54]	Retro. case control	Incident diagnostic	United States	PHC	32,787	05/2000-05/2013	Heart failure
Du *et al*. 2020 [55]	Retro. cohort	Prognostic	China	Linked data	42,676	2010–2018	Hypertension
Farran *et al*. 2013 [56]	Retro. cohort	Incident diagnostic	Kuwait	Linked data	270,172	12 years	Any cardiovascular disease
Hill *et al*. 2019 [57]	Retro. cohort	Incident diagnostic	United Kingdom	PHC	2,994,837	01/2006-12/2016	Atrial fibrillation
Karapetyan *et al*. 2021 [29]	Retro. cohort	Prognostic	Germany	PHC	46,071	02-2020-09/2020	Any cardiovascular disease
Lafreniere *et al*. 2017 [58]	Retro. cohort	Incident diagnostic	Canada	PHC	379,027	Not reported	Hypertension
Li *et al*. 2020 [59]	Retro. cohort	Incident diagnostic	United Kingdom	Linked data	3,661,932	01/1998-12/2018	Any cardiovascular disease
Lip *et al*. 2021 [60]	Retro. cohort	Prevalent diagnostic	Australia	Linked data	926	Not reported	Hypertension
Lorenzoni *et al*. 2019 [61]	Pros. cohort	Prognostic	Italy	Linked data	380	2011–2015	Heart failure
Ng *et al*. 2016 [62]	Retro. nested case control	Incident diagnostic	United States	PHC	152,095	2003–2010	Heart failure
Nikolaou *et al*. 2021 [63]	Retro. cohort	Prognostic	United Kingdom	PHC	6,883	2015–2019	Any cardiovascular disease
Sarraju *et al*. 2021 [64]	Retro. cohort	Incident diagnostic	United States	PHC	32,192	01/2009-12/2018	Any cardiovascular disease
Selskyy *et al*. 2018 [65]	Retro. case control	Prognostic	Ukraine	PHC	63	2011–2012	Hypertension
Shah *et al*. 2019 [45]	Retro. cohort	Prognostic/Incident diagnostic	United Kingdom	Linked data	2,000	Not reported	Myocardial infarction
Solanki *et al*. 2020 [66]	Retro. cohort	Prevalent diagnostic	United States	PHC	495	2007–2017	Hypertension
Solares *et al*. 2019 [67]	Retro. cohort	Incident diagnostic	United Kingdom	PHC	80,964	entry– 01/2014	Any cardiovascular disease
Ward *et al*. 2020 [68]	Retro. cohort	Incident diagnostic	United States	PHC	262,923	01/2009-12/2018	Atherosclerosis
Weng *et al*. 2017 [69]	Retro. cohort	Incident diagnostic	United Kingdom	PHC	378,256	01/2005-01/2015	Any cardiovascular disease
Wu *et al*. 2010 [70]	Retro. case control	Incident diagnostic	United States	PHC	44,895	01/2003-12/2006	Heart failure
Zhao *et al*. 2020 [40]	Retro. cohort	Incident diagnostic	United States	Linked data	4,914	Not reported	Stroke
Digestive System Diseases							
Sáenz Bajo *et al*. 2002 [26]	Retro. cohort	Prevalent diagnostic	Spain	PHC	81	01/1999-06/1999	Gastroesophageal reflux
Waljee *et al*. 2018 [71]	Retro. cohort	Prognostic	United States	Linked data	20,368	2002–2009	Inflammatory bowel disorders
Endocrine, Metabolic, and Nutritional Diseases							
Akyea *et al*. 2020 [31]	Retro. cohort	Incident diagnostic	United Kingdom	PHC	4,027,775	01/1999-06/2019	Familial hypercholesterolemia
Á-Guisasola *et al*. 2010 [72]	Pros. cohort	Incident diagnostic	Spain	PHC	2,662	Not reported	Diabetes mellitus
Crutzen *et al*. 2021 [73]	Retro. cohort	Incident diagnostic	The Netherlands	PHC	138,767	01/2007-01/2014	Diabetes mellitus
Ding *et al*. 2019 [74]	Retro. case control	Prevalent diagnostic	United States	PHC	97,584	1997–2017	Primary Aldosteronism
Dugan *et al*. 2015 [75]	Retro. cohort	Prognostic	United States	PHC	7,519	Over 9 years	Obesity
Farran *et al*. 2019 [76]	Retro. cohort	Prognostic	Kuwait	PHC	1,837	Over 9 years	Diabetes mellitus

(*Continued*)

Table 2. (Continued)

Study	Study design	Models purpose	Country	Source of data	Sample size	Time span of data	Health condition
Hammond *et al*. 2019 [77]	Retro. cohort	Prognostic	United States	PHC	3,449	01/2008-08/2016	Obesity
Kopitar *et al*. 2020 [78]	Retro. case control	Incident diagnostic	Slovenia	PHC	27,050	12/2014-09/2017	Diabetes mellitus
Lethebe *et al*. 2019 [79]	Retro. cohort	Prevalent diagnostic	Canada	PHC	1,309	2008–2016	Diabetes mellitus
Looker *et al*. 2015 [80]	Retro. case control	Prognostic	United Kingdom	PHC	309	12/1998-05/2009	Diabetic nephropathy
Metsker *et al*. 2020 [81]	Retro. cohort	Incident diagnostic	Russia	NR	54,252	07/2009-08/2017	Diabetic polyneuropathy
Metzker *et al*. 2020 [82]	Retro. cohort	Incident diagnostic	Russia	NR	58,462	Not reported	Diabetic polyneuropathy
Nagaraj *et al*. 2019 [83]	Retro. cohort	Prognostic	The Netherlands	PHC	11,887	01/2007-12/2013	Diabetes mellitus
Pakhomov *et al*. 2008 [50]	Retro. cohort	Prevalent diagnostic	United States	PHC	145	07/2004-09/2004	Diabetic foot
Rumora *et al*. 2021 [84]	Cross sectional	Incident diagnostic	Denmark	PHC	97	10/2015-06/2016	Diabetic polyneuropathy
Tseng *et al*. 2021 [52]	Cross sectional	Incident diagnostic	United States	PHC	NR	07/2016-12/2018	Diabetes mellitus
Wang *et al*. 2021 [85]	Retro. cohort	Incident diagnostic	China	PHC	1,139	2017–2019	Gestational diabetes
Williamson *et al*. 2020 [86]	Pros. cohort	Incident diagnostic	United States	Linked data	866	Not reported	Familial hypercholesterolemia
External Cause of Mortality							
DelPozo-Banos *et al*. 2018 [87]	Retro. case control	Incident diagnostic	United Kingdom	Linked data	54,684	2001–2015	Suicidality
Penfold *et al*. 2021 [88]	Retro. cohort	Incident diagnostic	United States	Linked data	256,823	Not reported	Suicidality
van Mens *et al*. 2020 [89]	Retro. case control	Incident diagnostic	The Netherlands	PHC	207,882	2017	Suicidality
Genitourinary System Diseases							
Shih *et al*. 2020 [90]	Retro. cohort	Incident diagnostic	Taiwan	Linked data	19,270	01/2015-12/2019	Chronic kidney disease
Zhao *et al*. 2019 [91]	Retro. cohort	Incident diagnostic	United States	PHC	61,740	2009–2017	Chronic kidney disease
Mental and Behavioral Diseases							
Dinga *et al*. 2018 [92]	Pros. cohort	Prognostic	The Netherlands	Linked data	804	Not reported	Depression
Ford *et al*. 2019 [93]	Retro. case control	Incident diagnostic	United Kingdom	PHC	93,120	2000–2012	Alzheimer's disease
Ford *et al*. 2020 [94]	Retro. case control	Incident diagnostic	United Kingdom	PHC	95,202	2000–2012	Alzheimer's disease
Ford *et al*. 2021 [95]	Retro. case control	Prevalent diagnostic	United Kingdom	PHC	93,426	2000–2012	Alzheimer's disease
Fouladvand *et al*. 2019 [96]	Retro. cohort	Prognostic	United States	PHC	3,265	Not reported	Alzheimer's disease
Haun *et al*. 2021 [97]	Cross sectional	Incident diagnostic	Germany	PHC	496	Not reported	Anxiety
Jammeh *et al*. 2018 [98]	Retro. case control	Incident diagnostic	United Kingdom	PHC	3,063	06/2010-06/2012	Alzheimer's disease
Jin *et al*. 2019 [99]	Retro. cohort	Incident diagnostic	United States	PHC	923	2010–2013	Depression
Kaczmarek *et al*. 2019 [100]	Retro. case control	Prevalent diagnostic	Canada	PHC	890	Not reported	Post-traumatic stress disorder
Ljubic *et al*. 2020 [101]	Retro. cohort	Incident diagnostic	United States	PHC	2,324	Not reported	Alzheimer's disease
Mallo *et al*. 2020 [102]	Retro. case control	Prognostic	Spain	PHC	128	2008	Alzheimer's disease
Mar *et al*. 2020 [103]	Retro. case control	Prevalent diagnostic	Spain	Linked data	4,003	Not reported	Alzheimer's disease

(*Continued*)

Table 2. (Continued)

Study	Study design	Models purpose	Country	Source of data	Sample size	Time span of data	Health condition
Półchłopek *et al.* 2020 [104]	Retro. cohort	Incident diagnostic	The Netherlands	PHC	92,621	2007-12/2016	Any mental disorder
Shen *et al.* 2020 [105]	Retro. cohort	Incident diagnostic	China	PHC	2,299	2008–2018	Alzheimer's disease
Suárez-Araujo *et al.* 2021 [106]	Retro. case control	Prevalent diagnostic	United States	PHC	330	Not reported	Alzheimer's disease
Tsang *et al.* 2021 [28]	Retro. cohort	Prognostic	United Kingdom	PHC	59,298	1982–2015	Alzheimer's disease
Zafari *et al.* 2021 [107]	Retro. cohort	Incident diagnostic	Canada	PHC	154,118	01/1995-12/2017	Post-traumatic stress disorder
Musculoskeletal and Connective Tissue Diseases							
Emir *et al.* 2014 [108]	Retro. cohort	Incident diagnostic	United States	PHC	587,961	2011–2012	Fibromyalgia
Jarvik *et al.* 2018 [109]	Pros. cohort	Prognostic	United States	PHC	3,971	03/2011-03/2013	Back pain
Kennedy *et al.* 2021 [110]	Retro. case control	Incident diagnostic	United Kingdom	Linked data	23,528	Over 6 years	Ankylosing spondylitis
Neoplasms							
Kop *et al.* 2016 [111]	Retro. cohort	Incident diagnostic	The Netherlands	PHC	260,000	Not reported	Colorectal cancer
Malhotra *et al.* 2021 [112]	Retro. case control	Incident diagnostic	United Kingdom	PHC	5,695	01/2005-06/2009	Pancreatic cancer
Ristanoski *et al.* 2021 [113]	Retro. case control	Incident diagnostic	Australia	PHC	683	2016–2017	Lung cancer
Nervous System Diseases							
Cox *et al.* 2016 [114]	Retro. case control	Prevalent diagnostic	United Kingdom	PHC	3,960	01/2007-12/2011	Post stroke spasticity
Hrabok *et al.* 2021 [115]	Retro. cohort	Prognostic	United Kingdom	PHC	10,499	01/2000-05/2012	Epilepsy
Kwasny *et al.* 2021 [116]	Retro. case control	Incident diagnostic	Germany	PHC	3,274	01/2010-12/2017	Progressive supranuclear palsy
Respiratory System Diseases							
Afzal *et al.* 2013 [117]	Retro. cohort	Prevalent diagnostic	The Netherlands	PHC	5,032	01/2000-01/2012	Asthma
Doyle *et al.* 2020 [118]	Retro. case control	Incident diagnostic	United Kingdom	PHC	112,784	09/2003-09/2017	Non-tuberculous mycobacterial lung
Kaplan *et al.* 2020 [32]	Retro. cohort	Prevalent diagnostic	United States	Linked data	411,563	Not reported	Asthma/obstructive pulmonary disease
Lisspers *et al.* 2021 [41]	Retro. cohort	Prognostic	Sweden	Linked data	29,396	01/2000-12/2013	Asthma
Marin-Gomez *et al.* 2021 [42]	Retro. cohort	Incident diagnostic	Spain	PHC	7,314	03/04/2020	COVID-19
Ställberg *et al.* 2021 [33]	Retro. cohort	Prognostic	Sweden	Linked data	7,823	01/2000-12/2013	Chronic obstructive pulmonary disease
Stephens *et al.* 2020 [51]	Retro. case control	Incident diagnostic	United States	PHC	7,278	2009–2019	Influenza
Trtica-Majnaric *et al.* 2010 [43]	Retro. cohort	Prognostic	Croatia	PHC	90	2003–2004	Influenza
Zafari *et al.* 2022 [44]	Retro. cohort	Incident diagnostic	Canada	PHC	4,134	Not reported	Chronic obstructive pulmonary disease

https://doi.org/10.1371/journal.pone.0274276.t002

Table 3. Overview of the included studies with the type of ML prediction models development with conduction of external validation (n = 13).

Study	Study design	Models purpose	Country	Source of data	Sample size	Time span of data	Health condition
Endocrine, Metabolic, and Nutritional Diseases							
Hertroijs *et al*. 2018 [49][a]	Retro. cohort	Prognostic	The Netherlands	PHC	10,528	01/2006-12/2014	Diabetes mellitus
			The Netherlands	PHC	3,337	01/2009-12/2013	
Myers *et al*. 2019 [39] [a]	Retro. case control	Incident diagnostic	United States	PHC	33,086	09/2013-08/2016	Familial hypercholesterolemia
			United States	Linked data	7,805		
			United States	Linked data	35,090		
			United States	Linked data	8,094		
Perveen *et al*. 2019 [34] [a]	Retro. cohort	Prognostic	Canada	PHC	911	08/2003-06/2015	Diabetes mellitus
			Canada	PHC	1,970		
Weisman *et al*. 2020 [119] [a]	Retro. cohort	Prevalent diagnostic	Canada	PHC	5,402	2010–2017	Diabetes mellitus
			Canada	Linked data	29,371		
Mental and Behavioral Diseases							
Amit *et al*. 2021 [120] [a]	Retro. cohort	Prevalent diagnostic	United Kingdom	PHC	24,612	2000–2010	Post-partum depression
			United Kingdom	PHC	9,193	2010–2017	
			United Kingdom	PHC	34,525	2000–2017	
Levy *et al*. 2018 [30] [a]	Retro. cohort	Incident diagnostic	United States	PHC	49	Over 9 months	Alzheimer's disease
			United States	Linked data	26	Not reported	
Perlis 2013 [121] [a]	Retro. cohort	Prognostic	United States	PHC	2,094	1999–2006	Depression
			United States	PHC	461		
Raket *et al*. 2020 [35] [a]	Retro. case control	Incident diagnostic	United States	PHC	145,720	1990–2018	Psychosis
			United States	PHC	4,770		
Musculoskeletal and Connective Tissue Diseases							
Fernandez-Gutierrez *et al*. 2021 [122] [a]	Retro. cohort	Incident diagnostic	United Kingdom	Linked data	19,314	2002–2012	Rheumatoid arthritis & Ankylosing spondylitis
			United Kingdom	Linked data	1,868		
Jorge *et al*. 2019 [123] [a]	Retro. cohort	Incident diagnostic	United States	Linked data	400	Not reported	Systematic lupus erythematous
			United States	Linked data	173	Not reported	
Zhou *et al*. 2017 [124] [a]	Retro. cohort	Incident diagnostic	United Kingdom	Linked data	Not reported	10/2013-07/2014	Rheumatoid arthritis
			United Kingdom	Linked data	475,580	03/2009-10/2012	
Neoplasms							
Kinar *et al*. 2016 [125] [a]	Retro. cohort	Incident diagnostic	Israel	PHC	606,403	01/2003-07/2011	Colorectal cancer
			United Kingdom	PHC	30,674	01/2003-05/2012	
Pregnancy, Childbirth, Puerperium							
Sufriyana *et al*. 2020 [126] [a]	Retro. nested case control	Incident diagnostic	Indonesia	Linked data	20,975	2015–2016	Preeclampsia
			Indonesia	Linked data	1,322	Not reported	
			Indonesia	Linked data	904	Not reported	

[a]Each row per study represents a different dataset that was used to develop and/or validate the prediction models.

https://doi.org/10.1371/journal.pone.0274276.t003

Table 4. Overview of the included studies with the type of reporting external validation of previously developed ML prediction models (n = 12).

Study	Study design	Models purpose	Country	Source of data	Sample size	Time span of data	Health condition
Circulatory System Diseases							
Kostev *et al.* 2021 [37]	Retro. cohort	Incident diagnostic	Germany	PHC	11,466	01/2010-12/2018	Stroke
Sekelj *et al.* 2020 [36]	Retro. cohort	Incident diagnostic	United Kingdom	PHC	604,135	01/2001-12/2016	Atrial fibrillation
Endocrine, Metabolic, and Nutritional Diseases							
Abramoff *et al.* 2019 [127]	Pros. cohort	Prevalent diagnostic	United States	PHC	819	01/2017-07/2017	Diabetic retinopathy
Bhaskaranand *et al.* 2019 [128]	Retro. cohort	Prevalent diagnostic	United States	PHC	1,017,001	01/2014-09/2015	Diabetic retinopathy
González-Gonzalo *et al.* 2019 [129] [a]	Retro. case control	Prevalent diagnostic	Spain	PHC	288	08/2011-10/2016	Diabetic retinopathy
			Sweden	PHC			
			Denmark	PHC			
			United States	Linked data	4,613	Over 2014	
			United Kingdom	Linked data			
Kanagasingam *et al.* 2018 [130]	Pros. cohort	Incident diagnostic	Australia	PHC	193	12.2016-05/2017	Diabetic retinopathy
Verbraak *et al.* 2019 [46]	Retro. cohort	Prevalent diagnostic	The Netherlands	PHC	1,425	2015	Diabetic retinopathy
Neoplasms							
Birks *et al.* 2017 [38]	Retro. case control	Incident diagnostic	United Kingdom	PHC	2,550,119	01/2000-04/2015	Colorectal cancer
Hoogendoorn *et al.* 2016 [131]	Retro. case control	Prevalent diagnostic	The Netherlands	PHC	90,000	07/2006-12/2011	Colorectal cancer
Hornbrook *et al.* 2017 [47]	Retro. case control	Incident diagnostic	United States	Linked data	17,095	1998–2013	Colorectal cancer
Kinar *et al.* 2017 [48]	Pros. cohort	Incident diagnostic	Israel	Linked data	112,584	07/2007-12/2007	Colorectal cancer
Respiratory System Diseases							
Morales *et al.* 2018 [132]	Retro. cohort	Prognostic	United Kingdom	PHC	2,044,733	01/2000-04/2014	Chronic obstructive pulmonary disease

[a]Each row per study represents a different dataset that was used to develop and/or validate the prediction models.

https://doi.org/10.1371/journal.pone.0274276.t004

(n = 1), and any cardiovascular event or disease (n = 7). Five respiratory health conditions were investigated in 10 studies [32, 33, 41–44, 51, 117, 118, 132, 134, 135]. Four studies predicted mortality and hospitalization risks on top of chronic obstructive pulmonary disease (COPD) (J40). Two studies investigated prevalent diagnosis of Asthma (J45) and its exacerbation risk. Influenza was predicated in two studies [117, 124] for incident cases and prognosis. COVID-19 (U07) incident cases were predicted within routine PHC visits in one study [42].

Other health conditions. Eight studies colorectal cancer (CRC) (C18) (n = 6), lung cancer (C34) (n = 1), and pancreatic cancer (C25) (n = 1). Four studies addressed the same incidence prediction model known as ColonFlag (previously MeScore) to identify CRC cases [38, 47, 48, 125]. Each study predicted incident cases within different time windows before diagnosis; from three months to two years. Three health conditions affecting the nervous system were addressed [114–116], which were post stroke spasticity, epilepsy specifically mortality four years before and after its diagnosis (G40) [115], and a rare neurodegenerative disease progressive supra-nuclear palsy (G23) [116]. A few studies investigated musculoskeletal and connective tissue disorders as well as gastrointestinal and kidney diseases [122–124, 108–110]. The musculoskeletal and connective tissue condition were back pain (M54) prognosis within PHC settings [109], ankylosing spondylitis (M45) [110]. The gastrointestinal and kidney diseases were examined in four studies, namely inflammatory bowel diseases (K50-K52), including Crohn's disease and ulcerative colitis [26, 71], peptic ulcers (K27)/gastroesophageal reflux (K21), and chronic kidney disease (N18) [90, 133]. Three studies tackled suicidality (X60-X84) [87–89]. Lastly, one study addressed preeclampsia (O14) [126].

Quality assessment

Quality was assessed using the PROBAST tool and 90.5% (n = 96 of 106) of the included studies were of high and unclear risk of bias (Fig 3). Analysis domain was the main source of bias, because of underreporting. It was found that only a few studies (n = 11) were reported in accordance with transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) guidelines [136]. Nevertheless, studies of low risk of bias were downgraded to be of high risk of bias due to the of lack of external validation of the proposed models (n = 20). The second concern assessed using this tool was the concern of applicability, which was estimated as low to moderate concern (66%). The dependence of the predictive models on not-routine PHC data as a concern of models' applicability within PHC settings was raised in 34% of the studies.

Most of the included studies (n = 98 of 106, 92.5%) were published as peer-reviewed articles in biomedical (e.g., PLOS ONE, n = 8) and technical journals (e.g., IEEE, n = 3). Eight studies were preprint and abstracts. National research institutes and universities were the most frequently reported funding support. Most of the studies reported that the funders were not involved in the published work.

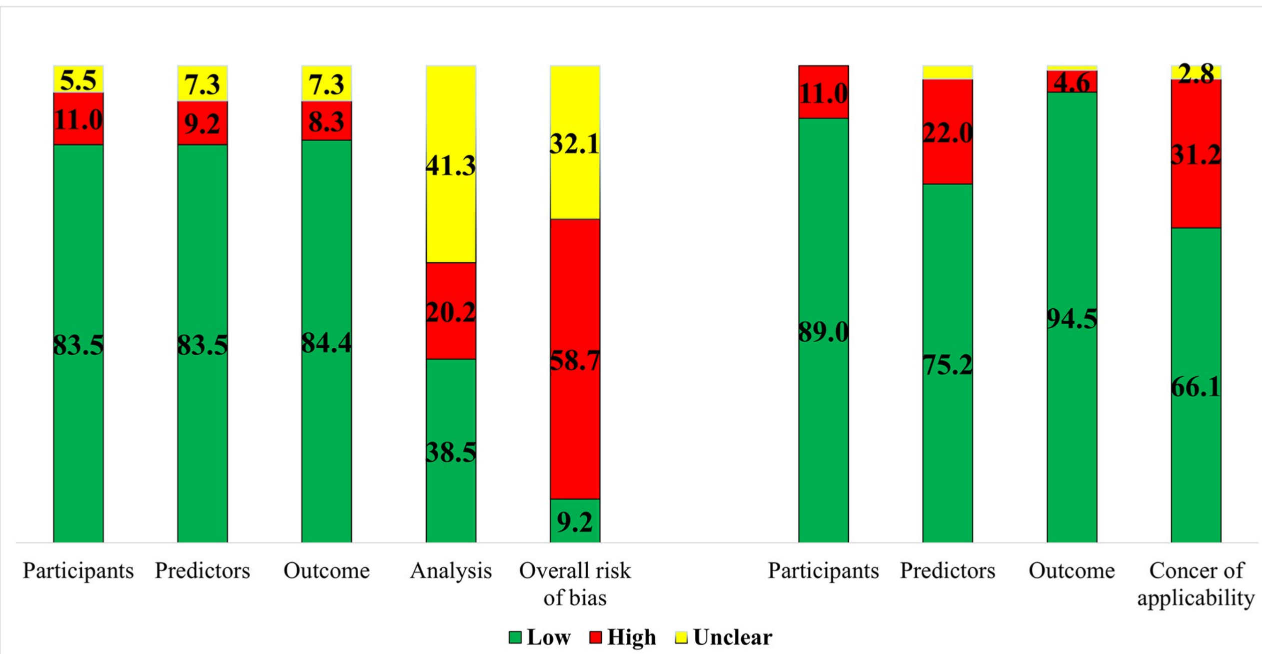

Fig 3. Percentage presentation of the results of (PROBAST) tool. The tool has two components. Component 1. Risk of bias (4 domains: Participants, predictors, outcome, and analysis). Component 2. Concern of applicability (3 domains: Participants, predictors, and outcome).

https://doi.org/10.1371/journal.pone.0274276.g003

Discussion

ML prediction models could have an immense potential to augment health practice and clinical decision making in various health sectors. Our systematic review provides an outline of the health conditions investigated with ML prediction models using PHC data.

Summary of findings

In 106 observational studies, we identified 42 health conditions targeted by 207 ML prediction models, of which 42.5% were random forest and SVM. The included models used PHC data documented over the past 40 years for a total of 22.7 million patients. Half of the studies were conducted in the US and the UK. While the majority of the included studies (77.3%) focused on diagnosis prediction, a significant portion also addressed predictive aspects related to complications, hospitalization, and mortality. The most frequently targeted health conditions included Alzheimer's disease, diabetes mellitus, heart failure, colorectal cancer, and chronic obstructive pulmonary diseases, while other conditions such as asthma, childhood obesity, and dyspepsia received comparatively less attention. A considerable portion of the models (76.4% of the included studies) were trained and internally validated without evaluating their generalizability.

Results in perspective

Detection and management of health conditions, particularly those that are preventable and controllable like diabetes mellitus, stand for the fundamental role of PHC [3]. Advances in such technologies might enhance health care and quality of life. Noticeably, they have gained more attention in many countries [11]. Our findings of common and rare health conditions targeted by ML prediction models in PHC indicates increase of research interest. However, clinical implication of such models is still limited to the theoretical good performance. Furthermore, the unequal distribution of publications across countries could be related to the low publication rate or lack of proper health data documentation systems in lower income countries, which impose further limitation to validate and implement such models.

The coding system used in health records does not universally follow the same criteria for all diseases, posing challenges for the consistency of models' performance [137]. Moreover, the lack of globally standardized definitions and terminology of diseases and the wide variability of the services provided across different health systems further limit the effectiveness of the models [137]. For example, uncoded free-text clinical notes as well as using 'race' and 'ethnicity' or 'suicide' and 'suicide attempts' to be documented as a single input can affect the predictive power of the models [138]. Other drawbacks reported include underrepresentation of healthy persons, retrospective temporal dimension of predictors, and the absence of confirmatory diagnostic services in PHC pose significant limitations [139, 140].

Technical biases can significantly influence the clinical utility of technologies. Models trained on historical data without adaptation to policy changes may reinforce outdated practices, leading to erroneous results [141]. Additionally, validating models using different populations data can create a mismatch between the data or environment on which the models was trained; this mismatch may impact the accuracy of the models' prediction [141]. Therefore, documenting characteristics of the health systems may highlight the discrepancies between the data used to train and validate the models. This may improve the validation and implementation processes of the models. Models that are known for their high prediction accuracy, such as random forest and SVM might support better health outcomes when developed using high quality health data [139]. Additionally, the variety of the ML prediction models characteristics provide opportunities to improve healthcare practice. Using large data documented as electronic health records, random forest models and ensemble models such as boosting models have the ability to handle large datasets with numerous predictors variables [140]. Artificial neural network can also perform complex images processing that can boost the primary health care services [140]. Furthermore, SVM and decision tree models can provide nonlinear solutions, thus will support our understanding of complex and dynamic diseases for earlier health conditions prediction [142].

Nature of diseases append further challenges. The most challenging diseases for ML prediction are multifaceted long-term health conditions, such as DM, that are influenced by combination of genetic, environmental, and lifestyle factors. The complex health conditions further tangle the models, making it harder to identify accurate predictive patterns. Furthermore, the subjective nature of symptoms, especially symptoms related to mental health disorders, pose additive challenges toward ML models accuracy. Rare diseases, if documented, often suffer from limited data availability, leading to difficulty to train ML models effectively [143].

Health care professionals are fundamental to the process of implementing and integrating ML prediction models in their healthcare practice. Despite that, our review did not report outcomes related to healthcare professionals. Significant variability of opinions on the utilization of ML in PHC among primary health care providers hinder its acceptance. Furthermore, the black-box nature of ML prediction models precludes the clinical interpretability of models' outcomes. Additional workload and training are needed to implement such technology in the routine practice. Trust, data protection, and ethical and clinical responsibility legislation are further intractable issues that represent major obstacles toward ML prediction models implementation [5].

A considerable lack of usage of studies reporting guidelines across the included studies lead to deficient description of the populations' demographics and underreporting of the models' related statistical analysis, which lead to high risk of bias of majority of studies. These shortcomings negatively affect the reproducibility of the models [144]. Navarro and colleagues investigated this underreporting, and they claimed that the available reporting guidelines of modelling studies might be less apposite for ML models studies [145].

Implication of results and recommendation for future contributions

This review provided a comprehensive outline of ML prediction models in PHC and raises important considerations for future research and implementation of this technology in PHC settings. Interdisciplinary collaboration among health care workers, developers of ML models, and creators of digital documentation systems is required. This is especially important given the increasing popularity of digitally connected health systems [5]. It is recommended to augment the participation of health professionals through the development process of the PHC predictive models to critically evaluate, assess, adopt, and challenge the validation of the models within practices. This collaboration may assist ML engineers to recognize unintended negative implications of their algorithms, such as accidentally fitting of confounders, and unintended discriminatory bias, among others, for better health outcomes [146]. Health care systems need to provide comprehensive population health data repositories as an enabler for medical analyses [137]. Well-designed and -documented repositories which provide representative health data for the healthy and diseased populations are needed [137, 139]. These high-quality data repositories might provide future modelling studies with data that match the studies' clinical research questions for more accurate prediction. Further ML prediction studies are needed to target more health conditions using PHC data. Despite the additional burden, it is beneficial also to continuously assess the potential significance of models, such as improved health outcomes, reduced medical errors, increased professional effectiveness and productivity, and enhanced patients' quality of life [147]. It is recommended to follow reporting guidelines for producing valid and reproducible ML modelling studies. Developing robust frameworks to enable the adoption and integration of ML models in the routine practice is also essential for effective transition from conventional health care systems to digital health [148, 149]. Sophisticated technical infrastructure and strong academic and governmental support are essential for promoting and supporting long-term and broad-reaching PHC ML-based services [138, 150]. However, balanced arguments [151, 152] regarding the potential benefits and limitations of ML models support better health care without overestimating or hampering the use of such technology. It is also suggested to integrate the basic understanding of ML concepts and techniques in education programs for health science and medical students.

Strengths and limitations of the review

Our review was conducted following a predesigned comprehensive protocol [21]. We identified the health conditions targeted within PHC settings and identified the gaps that need to be addressed. The main limitation of our review is the low quality of evidence of the primary evidence. It is also possible due to the wide array of descriptors that exist to describe ML, our search strategy could have missed some studies if they exclusively used terms outside of our search string [153]. Limiting the scope of our review to clinical health conditions might have excluded other conditions, such as domestic violence and drug abuse [3]. Guiding our work using ICD-10 might have led to the exclusion of some health conditions, such as frailty studies [154]. Lastly, we did not present the statistical analysis of the models' attributes or conduct a meta-analysis, because of the broad heterogeneity across studies. In the future, we plan to update our review–considering the noticeable rise of ML studies within PHC, while also modifying our methodology to reduce the identified limitations. It is also planned to use the specific ML guidelines TRIPOD-AI and PROBAST-AI when published to strengthen quality and reporting of our findings [155].

In conclusion, ML prediction models within PHC are gaining traction. Further studies examining the use of ML in real PHC settings are needed, especially those with prospective designs and more representative samples. Collaborating amongst multidisciplinary teams to tackle ML in PHC will increase the confidence in models and their implementations in clinical practice.

Supporting information

S1 Table. List of excluded studies with reasons (n = 58).
(PDF)

S2 Table. Characteristics of the included ML predictive models.
(PDF)

S1 File. Search strategy.
(PDF)

S2 File. Prisma checklist.
(DOC)

S1 Panel. Quantitative summary of the included studies' characteristics (n = 106).
(PDF)

Acknowledgments

Marcos André Gonçalves, PhD and Bruna Zanotto, MSc, provided their feedback on the project's primary draft. Luana Fiengo Tanaka, PhD, helped retrieved inaccessible studies.

Author Contributions

Conceptualization: Hebatullah Abdulazeem.

Data curation: Hebatullah Abdulazeem, Sera Whitelaw.

Formal analysis: Hebatullah Abdulazeem.

Methodology: Hebatullah Abdulazeem.

Project administration: Hebatullah Abdulazeem.

Supervision: Gunther Schauberger, Stefanie J. Klug.

Visualization: Hebatullah Abdulazeem.

Writing – original draft: Hebatullah Abdulazeem.

Writing – review & editing: Hebatullah Abdulazeem, Sera Whitelaw, Gunther Schauberger, Stefanie J. Klug.

References

1. Aoki M. Editorial: Science and roles of general medicine. Japanese J Natl Med Serv. 2001; 55: 111–114. https://doi.org/10.11261/iryo1946.55.111
2. Troncoso EL. The Greatest Challenge to Using AI/ML for Primary Health Care: Mindset or Datasets? Front Artif Intell. 2020; 3: 53. https://doi.org/10.3389/frai.2020.00053 PMID: 33733170
3. Hashim MJ. A definition of family medicine and general practice. J Coll Physicians Surg Pakistan. 2018; 28: 76–77. https://doi.org/10.29271/jcpsp.2018.01.76 PMID: 29290201
4. Cao L. Data science: A comprehensive overview. ACM Comput Surv. 2018; 50: 1–42. https://doi.org/10.1145/3076253
5. Liyanage H, Liaw ST, Jonnagaddala J, Schreiber R, Kuziemsky C, Terry AL, et al. Artificial Intelligence in Primary Health Care: Perceptions, Issues, and Challenges. Yearb Med Inform. 2019; 28: 41–46. https://doi.org/10.1055/s-0039-1677901 PMID: 31022751
6. Debray TPA, Damen JAAG, Snell KIE, Ensor J, Hooft L, Reitsma JB, et al. A guide to systematic review and meta-analysis of prediction model performance. BMJ. 2017; 356: i6460. https://doi.org/10.1136/bmj.i6460 PMID: 28057641
7. Sarker IH. Machine Learning: Algorithms, Real-World Applications and Research Directions. SN Comput Sci. 2021; 2: 160. https://doi.org/10.1007/s42979-021-00592-x PMID: 33778771
8. Do Nascimento IJB, Marcolino MS, Abdulazeem HM, Weerasekara I, Azzopardi-Muscat N, Goncalves MA, et al. Impact of big data analytics on people's health: Overview of systematic reviews and recommendations for future studies. J Med Internet Res. 2021; 23: e27275. https://doi.org/10.2196/27275 PMID: 33847586
9. Marcus JL, Sewell WC, Balzer LB, Krakower DS. Artificial Intelligence and Machine Learning for HIV Prevention: Emerging Approaches to Ending the Epidemic. Curr HIV/AIDS Rep. 2020; 17: 171–179. https://doi.org/10.1007/s11904-020-00490-6 PMID: 32347446
10. Amaratunga D, Cabrera J, Sargsyan D, Kostis JB, Zinonos S, Kostis WJ. Uses and opportunities for machine learning in hypertension research. Int J Cardiol Hypertens. 2020; 5: 100027. https://doi.org/10.1016/j.ijchy.2020.100027 PMID: 33447756
11. 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. https://doi.org/10.1016/j.csbj.2016.12.005 PMID: 28138367
12. Sufriyana H, Husnayain A, Chen YL, Kuo CY, Singh O, Yeh TY, et al. Comparison of multivariable logistic regression and other machine learning algorithms for prognostic prediction studies in pregnancy care: Systematic review and meta-analysis. JMIR Med Informatics. 2020; 8: e16503. https://doi.org/10.2196/16503 PMID: 33200995

13. Rajpara SM, Botello AP, Townend J, Ormerod AD. Systematic review of dermoscopy and digital dermoscopy/ artificial intelligence for the diagnosis of melanoma. Br J Dermatol. 2009; 161: 591–604. https://doi.org/10.1111/j.1365-2133.2009.09093.x PMID: 19302072

14. 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: e0234722. https://doi.org/10.1371/journal.pone.0234722 PMID: 32530947

15. Contreras I, Vehi J. Artificial intelligence for diabetes management and decision support: Literature review. J Med Internet Res. 2018; 20: e10775. https://doi.org/10.2196/10775 PMID: 29848472

16. Rahimi SA, Légaré F, Sharma G, Archambault P, Zomahoun HTV, Chandavong S, et al. Application of artificial intelligence in community-based primary health care: Systematic scoping review and critical appraisal. Journal of Medical Internet Research J Med Internet Res; Sep 1, 2021. https://doi.org/10.2196/29839 PMID: 34477556

17. Kueper JK, Terry AL, Zwarenstein M, Lizotte DJ. Artificial intelligence and primary care research: A scoping review. Ann Fam Med. 2020; 18: 250–258. https://doi.org/10.1370/afm.2518 PMID: 32393561

18. Andaur Navarro CL, Damen JAAG, Takada T, Nijman SWJ, Dhiman P, Ma J, et al. Protocol for a systematic review on the methodological and reporting quality of prediction model studies using machine learning techniques. BMJ Open. 2020; 10: e038832. https://doi.org/10.1136/bmjopen-2020-038832 PMID: 33177137

19. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. The BMJ. British Medical Journal Publishing Group; 2021. https://doi.org/10.1136/bmj.n71 PMID: 33782057

20. Moons KGM, de Groot JAH, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist. PLoS Med. 2014; 11: e1001744. https://doi.org/10.1371/journal.pmed.1001744 PMID: 25314315

21. Abdulazeem H, Whitelaw S, Schauberger G, Klug S. Development and Performance of Prediction Machine Learning Models supplied by Real-World Primary Health Care Data: A Systematic Review and Meta-analysis. In: PROSPERO 2021 CRD42021264582 [Internet]. 2021. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021264582

22. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev. 2016; 5: 210. https://doi.org/10.1186/s13643-016-0384-4 PMID: 27919275

23. World Health Organization. ICD-10 Version:2019. In: International Classification of Diseases [Internet]. 2019 [cited 1 Sep 2021]. https://icd.who.int/browse10/2019/en#/XIV

24. International Classification of Diseases (ICD). [cited 6 Apr 2023]. https://www.who.int/standards/classifications/classification-of-diseases

25. Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A tool to assess risk of bias and applicability of prediction model studies: Explanation and elaboration. Ann Intern Med. 2019; 170: W1–W33. https://doi.org/10.7326/M18-1377 PMID: 30596876

26. Sáenz Bajo N, Barrios Rueda E, Conde Gómez M, Domínguez Macías I, López Carabaño A, Méndez Díez C. Use of neural networks in medicine: concerning dyspeptic pathology. Aten Primaria. 2002; 30: 99–102. https://doi.org/10.1016/s0212-6567(02)78978-6 PMID: 12106560

27. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). Int J Epidemiol. 2015; 44: 827–836. https://doi.org/10.1093/ije/dyv098 PMID: 26050254

28. Tsang G, Zhou SM, Xie X. Modeling Large Sparse Data for Feature Selection: Hospital Admission Predictions of the Dementia Patients Using Primary Care Electronic Health Records. IEEE J Transl Eng Heal Med. 2021; 9. https://doi.org/10.1109/JTEHM.2020.3040236 PMID: 33354439

29. Karapetyan S, Schneider A, Linde K, Donnachie E, Hapfelmeier A. SARS-CoV-2 infection and cardiovascular or pulmonary complications in ambulatory care: A risk assessment based on routine data. PLoS One. 2021; 16: e0258914. https://doi.org/10.1371/journal.pone.0258914 PMID: 34673818

30. Boaz L, Samuel G, Elena T, Nurit H, Brianna W, Rand W, et al. Machine Learning Detection of Cognitive Impairment in Primary Care. Alzheimers Dis Dement. 2017; 1: S111. https://doi.org/10.36959/734/372

31. Akyea RK, Qureshi N, Kai J, Weng SF. Performance and clinical utility of supervised machine-learning approaches in detecting familial hypercholesterolaemia in primary care. NPJ Digit Med. 2020; 3: 142. https://doi.org/10.1038/s41746-020-00349-5 PMID: 33145438

32. Kaplan A, Cao H, Fitzgerald JM, Yang E, Iannotti N, Kocks JWH, et al. Asthma/COPD Differentiation Classification (AC/DC): Machine Learning to Aid Physicians in Diagnosing Asthma, COPD and Asthma-COPD Overlap (ACO). D22 COMORBIDITIES IN PEOPLE WITH COPD. American Thoracic Society; 2020. p. A6285.

33. Ställberg B, Lisspers K, Larsson K, Janson C, Müller M, Łuczko M, et al. Predicting hospitalization due to copd exacerbations in swedish primary care patients using machine learning–based on the arctic study. Int J COPD. 2021; 16: 677–688. https://doi.org/10.2147/COPD.S293099 PMID: 33758504
34. Perveen S, Shahbaz M, Keshavjee K, Guergachi A. Prognostic Modeling and Prevention of Diabetes Using Machine Learning Technique. Sci Rep. 2019; 9: 13805. https://doi.org/10.1038/s41598-019-49563-6 PMID: 31551457
35. Raket LL, Jaskolowski J, Kinon BJ, Brasen JC, Jönsson L, Wehnert A, et al. Dynamic ElecTronic hEalth reCord deTection (DETECT) of individuals at risk of a first episode of psychosis: a case-control development and validation study. Lancet Digit Heal. 2020; 2: e229–e239. https://doi.org/10.1016/S2589-7500(20)30024-8 PMID: 33328055
36. Sekelj S, Sandler B, Johnston E, Pollock KG, Hill NR, Gordon J, et al. Detecting undiagnosed atrial fibrillation in UK primary care: Validation of a machine learning prediction algorithm in a retrospective cohort study. Eur J Prev Cardiol. 2021; 28: 598–605. https://doi.org/10.1177/2047487320942338 PMID: 34021576
37. Kostev K, Wu T, Wang Y, Chaudhuri K, Tanislav C. Predicting the risk of stroke in patients with late-onset epilepsy: A machine learning approach. Epilepsy Behav. 2021; 122: 108211. https://doi.org/10.1016/j.yebeh.2021.108211 PMID: 34325155
38. Birks J, Bankhead C, Holt TA, Fuller A, Patnick J. Evaluation of a prediction model for colorectal cancer: retrospective analysis of 2.5 million patient records. Cancer Med. 2017; 6: 2453–2460. https://doi.org/10.1002/cam4.1183 PMID: 28941187
39. Myers KD, Knowles JW, Staszak D, Shapiro MD, Howard W, Yadava M, et al. Precision screening for familial hypercholesterolaemia: a machine learning study applied to electronic health encounter data. Lancet Digit Heal. 2019; 1: e393–e402. https://doi.org/10.1016/S2589-7500(19)30150-5 PMID: 33323221
40. Zhao Y, Fu S, Bielinski SJ, Decker P, Chamberlain AM, Roger VL, et al. Abstract P259: Using Natural Language Processing and Machine Learning to Identify Incident Stroke From Electronic Health Records. Circulation. 2020; 141. https://doi.org/10.1161/circ.141.suppl_1.p259
41. Lisspers K, Ställberg B, Larsson K, Janson C, Müller M, Łuczko M, et al. Developing a short-term prediction model for asthma exacerbations from Swedish primary care patients' data using machine learning—Based on the ARCTIC study. Respir Med. 2021; 185: 106483. https://doi.org/10.1016/j.rmed.2021.106483 PMID: 34077873
42. Marin-Gomez FX, Fàbregas-Escurriola M, Seguí FL, Pérez EH, Camps MB, Peña JM, et al. Assessing the likelihood of contracting COVID-19 disease based on a predictive tree model: A retrospective cohort study. PLoS One. 2021; 16: e0247995. https://doi.org/10.1371/journal.pone.0247995 PMID: 33657164
43. Trtica-Majnaric L, Zekic-Susac M, Sarlija N, Vitale B. Prediction of influenza vaccination outcome by neural networks and logistic regression. J Biomed Inform. 2010; 43: 774–781. https://doi.org/10.1016/j.jbi.2010.04.011 PMID: 20451660
44. Zafari H, Langlois S, Zulkernine F, Kosowan L, Singer A. AI in predicting COPD in the Canadian population. BioSystems. 2022; 211: 104585. https://doi.org/10.1016/j.biosystems.2021.104585 PMID: 34864143
45. Shah AD, Bailey E, Williams T, Denaxas S, Dobson R, Hemingway H. Natural language processing for disease phenotyping in UK primary care records for research: A pilot study in myocardial infarction and death. J Biomed Semantics. 2019; 10. https://doi.org/10.1186/s13326-019-0214-4 PMID: 31711543
46. Verbraak FD, Abramoff MD, Bausch GCF, Klaver C, Nijpels G, Schlingemann RO, et al. Diagnostic accuracy of a device for the automated detection of diabetic retinopathy in a primary care setting. Diabetes Care. 2019; 42: 651–656. https://doi.org/10.2337/dc18-0148 PMID: 30765436
47. Hornbrook MC, Goshen R, Choman E, O'Keeffe-Rosetti M, Kinar Y, Liles EG, et al. Early Colorectal Cancer Detected by Machine Learning Model Using Gender, Age, and Complete Blood Count Data. Dig Dis Sci. 2017; 62: 2719–2727. https://doi.org/10.1007/s10620-017-4722-8 PMID: 28836087
48. Kinar Y, Akiva P, Choman E, Kariv R, Shalev V, Levin B, et al. Performance analysis of a machine learning flagging system used to identify a group of individuals at a high risk for colorectal cancer. PLoS One. 2017; 12: e0171759. https://doi.org/10.1371/journal.pone.0171759 PMID: 28182647
49. Hertroijs DFL, Elissen AMJ, Brouwers MCGJ, Schaper NC, Köhler S, Popa MC, et al. A risk score including body mass index, glycated haemoglobin and triglycerides predicts future glycaemic control in people with type 2 diabetes. Diabetes, Obes Metab. 2018; 20: 681–688. https://doi.org/10.1111/dom.13148 PMID: 29095564
50. Pakhomov SVS, Hanson PL, Bjornsen SS, Smith SA. Automatic Classification of Foot Examination Findings Using Clinical Notes and Machine Learning. J Am Med Informatics Assoc. 2008; 15: 198–202. https://doi.org/10.1197/jamia.M2585 PMID: 18096902

51. Stephens KA, Au MA, Yetisgen M, Lutz B, Suchsland MZ, Ebell MH, et al. Leveraging UMLS-driven NLP to enhance identification of influenza predictors derived from electronic medical record data. In: BioRxiv [preprint] [Internet]. 2020 [cited 4 Jan 2022].
52. Tseng E, Schwartz JL, Rouhizadeh M, Maruthur NM. Analysis of Primary Care Provider Electronic Health Record Notes for Discussions of Prediabetes Using Natural Language Processing Methods. J Gen Intern Med. 2021; 35: S11–S12. https://doi.org/10.1007/s11606-020-06400-1 PMID: 33469758
53. Chen R, Stewart WF, Sun J, Ng K, Yan X. Recurrent neural networks for early detection of heart failure from longitudinal electronic health record data: Implications for temporal modeling with respect to time before diagnosis, data density, data quantity, and data type. Circ Cardiovasc Qual Outcomes. 2019; 12: e005114. https://doi.org/10.1161/CIRCOUTCOMES.118.005114 PMID: 31610714
54. Choi E, Schuetz A, Stewart WF, Sun J. Using recurrent neural network models for early detection of heart failure onset. J Am Med Informatics Assoc. 2017; 24: 361–370. https://doi.org/10.1093/jamia/ocw112 PMID: 27521897
55. Du Z, Yang Y, Zheng J, Li Q, Lin D, Li Y, et al. Accurate prediction of coronary heart disease for patients with hypertension from electronic health records with big data and machine-learning methods: Model development and performance evaluation. JMIR Med Informatics. 2020; 8: e17257. https://doi.org/10.2196/17257 PMID: 32628616
56. Farran B, Channanath AM, Behbehani K, Thanaraj TA. Predictive models to assess risk of type 2 diabetes, hypertension and comorbidity: Machine-learning algorithms and validation using national health data from Kuwait-a cohort study. BMJ Open. 2013; 3. https://doi.org/10.1136/bmjopen-2012-002457 PMID: 23676796
57. Hill NR, Ayoubkhani D, McEwan P, Sugrue DM, Farooqui U, Lister S, et al. Predicting atrial fibrillation in primary care using machine learning. PLoS One. 2019; 14: e0224582. https://doi.org/10.1371/journal.pone.0224582 PMID: 31675367
58. LaFreniere D, Zulkernine F, Barber D, Martin K. Using machine learning to predict hypertension from a clinical dataset. 2016 IEEE Symposium Series on Computational Intelligence (SSCI). IEEE; 2016. pp. 1–7.
59. Li Y, Sperrin M, Ashcroft DM, Van Staa TP. Consistency of variety of machine learning and statistical models in predicting clinical risks of individual patients: Longitudinal cohort study using cardiovascular disease as exemplar. BMJ. 2020; 371: m3919. https://doi.org/10.1136/bmj.m3919 PMID: 33148619
60. Lip S, Mccallum L, Reddy S, Chandrasekaran N, Tule S, Bhaskar RK, et al. Machine Learning Based Models for Predicting White-Coat and Masked Patterns of Blood Pressure. J Hypertens. 2021; 39: e69. https://doi.org/10.1097/01.hjh.0000745092.07595.a5
61. Lorenzoni G, Sabato SS, Lanera C, Bottigliengo D, Minto C, Ocagli H, et al. Comparison of machine learning techniques for prediction of hospitalization in heart failure patients. J Clin Med. 2019/08/28. 2019; 8. https://doi.org/10.3390/jcm8091298 PMID: 31450546
62. Ng K, Steinhubl SR, Defilippi C, Dey S, Stewart WF. Early Detection of Heart Failure Using Electronic Health Records: Practical Implications for Time before Diagnosis, Data Diversity, Data Quantity, and Data Density. Circ Cardiovasc Qual Outcomes. 2016; 9: 649–658. https://doi.org/10.1161/CIRCOUTCOMES.116.002797 PMID: 28263940
63. Nikolaou V, Massaro S, Garn W, Fakhimi M, Stergioulas L, Price D. The cardiovascular phenotype of Chronic Obstructive Pulmonary Disease (COPD): Applying machine learning to the prediction of cardiovascular comorbidities. Respir Med. 2021/07/15. 2021; 186: 106528. https://doi.org/10.1016/j.rmed.2021.106528 PMID: 34260974
64. Sarraju A, Ward A, Chung S, Li J, Scheinker D, Rodríguez F. Machine learning approaches improve risk stratification for secondary cardiovascular disease prevention in multiethnic patients. Open Hear. 2021; 8: e001802. https://doi.org/10.1136/openhrt-2021-001802 PMID: 34667093
65. Selskyy P, Vakulenko D, Televiak A, Veresiuk T. On an algorithm for decision-making for the optimization of disease prediction at the primary health care level using neural network clustering. Fam Med Prim Care Rev. 2018; 20: 171–175. https://doi.org/10.5114/fmpcr.2018.76463
66. Solanki P, Ajmal I, Ding X, Cohen J, Cohen D, Herman D. Abstract P185: Using Electronic Health Records To Identify Patients With Apparent Treatment Resistant Hypertension. Hypertension. 2020; 76. https://doi.org/10.1161/hyp.76.suppl_1.p185
67. Ayala Solares JR, Canoy D, Raimondi FED, Zhu Y, Hassaine A, Salimi-Khorshidi G, et al. Long-Term Exposure to Elevated Systolic Blood Pressure in Predicting Incident Cardiovascular Disease: Evidence From Large-Scale Routine Electronic Health Records. J Am Heart Assoc. 2019; 8. https://doi.org/10.1161/JAHA.119.012129 PMID: 31164039
68. Ward A, Sarraju A, Chung S, Li J, Harrington R, Heidenreich P, et al. Machine learning and atherosclerotic cardiovascular disease risk prediction in a multi-ethnic population. NPJ Digit Med. 2020; 3: 125. https://doi.org/10.1038/s41746-020-00331-1 PMID: 33043149

69. Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. Can Machine-learning improve cardiovascular risk prediction using routine clinical data? PLoS One. 2017; 12. https://doi.org/10.1371/journal.pone.0174944 PMID: 28376093

70. Wu J, Roy J, Stewart WF. Prediction modeling using EHR data: Challenges, strategies, and a comparison of machine learning approaches. Med Care. 2010; 48: S106–S113. https://doi.org/10.1097/MLR.0b013e3181de9e17 PMID: 20473190

71. Waljee AK, Lipson R, Wiitala WL, Zhang Y, Liu B, Zhu J, et al. Predicting Hospitalization and Outpatient Corticosteroid Use in Inflammatory Bowel Disease Patients Using Machine Learning. Inflamm Bowel Dis. 2018; 24: 45–53. https://doi.org/10.1093/ibd/izx007 PMID: 29272474

72. Álvarez-Guisasola F, Conget I, Franch J, Mata M, Mediavilla JJ, Sarria A, et al. Adding questions about cardiovascular risk factors improve the ability of the ADA questionnaire to identify unknown diabetic patients in Spain. Diabetologia. 2010; 26: 347–352. https://doi.org/10.1016/S1134-3230(10)65008-9

73. Crutzen S, Belur Nagaraj S, Taxis K, Denig P. Identifying patients at increased risk of hypoglycaemia in primary care: Development of a machine learning-based screening tool. Diabetes Metab Res Rev. 2021; 37: e3426. https://doi.org/10.1002/dmrr.3426 PMID: 33289318

74. Ding X, Ajmal I, Trerotola OSc, Fraker D, Cohen J, Wachtel H, et al. EHR-based modeling specifically identifies patients with primary aldosteronism. In: Circulation [Internet]. 2019 [cited 22 Sep 2021]. https://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed20&AN=630921513

75. Dugan TM, Mukhopadhyay S, Carroll A, Downs S. Machine learning techniques for prediction of early childhood obesity. Appl Clin Inform. 2015; 6: 506–520. https://doi.org/10.4338/ACI-2015-03-RA-0036 PMID: 26448795

76. Farran B, AlWotayan R, Alkandari H, Al-Abdulrazzaq D, Channanath A, Thanaraj TA. Use of Non-invasive Parameters and Machine-Learning Algorithms for Predicting Future Risk of Type 2 Diabetes: A Retrospective Cohort Study of Health Data From Kuwait. Front Endocrinol (Lausanne). 2019; 10. https://doi.org/10.3389/fendo.2019.00624 PMID: 31572303

77. Hammond R, Athanasiadou R, Curado S, Aphinyanaphongs Y, Abrams C, Messito MJ, et al. Predicting childhood obesity using electronic health records and publicly available data. PLoS One. 2019; 14: e0215571. https://doi.org/10.1371/journal.pone.0215571 PMID: 31009509

78. Kopitar L, Kocbek P, Cilar L, Sheikh A, Stiglic G. Early detection of type 2 diabetes mellitus using machine learning-based prediction models. Sci Rep. 2020; 10: 11981. https://doi.org/10.1038/s41598-020-68771-z PMID: 32686721

79. Lethebe BC, Williamson T, Garies S, McBrien K, Leduc C, Butalia S, et al. Developing a case definition for type 1 diabetes mellitus in a primary care electronic medical record database: an exploratory study. C open. 2019; 7: E246–E251. https://doi.org/10.9778/cmajo.20180142 PMID: 31061005

80. Looker HC, Colombo M, Hess S, Brosnan MJ, Farran B, Dalton RN, et al. Biomarkers of rapid chronic kidney disease progression in type 2 diabetes. Kidney Int. 2015; 88: 888–896. https://doi.org/10.1038/ki.2015.199 PMID: 26200946

81. Metsker O, Magoev K, Yanishevskiy S, Yakovlev A, Kopanitsa G, Zvartau N. Identification of diabetes risk factors in chronic cardiovascular patients. Stud Health Technol Inform. 2020; 273: 136–141. https://doi.org/10.3233/SHTI200628 PMID: 33087603

82. Metzker O, Magoev K, Yanishevskiy S, Yakovlev A, Kopanitsa G. Risk factors for chronic diabetes patients. Stud Health Technol Inform. 2020; 270: 1379–1380. https://doi.org/10.3233/SHTI200451 PMID: 32570668

83. Nagaraj SB, Sidorenkov G, van Boven JFM, Denig P. Predicting short- and long-term glycated haemoglobin response after insulin initiation in patients with type 2 diabetes mellitus using machine-learning algorithms. Diabetes, Obes Metab. 2019; 21: 2704–2711. https://doi.org/10.1111/dom.13860 PMID: 31453664

84. Rumora AE, Guo K, Alakwaa FM, Andersen ST, Reynolds EL, Jørgensen ME, et al. Plasma lipid metabolites associate with diabetic polyneuropathy in a cohort with type 2 diabetes. Ann Clin Transl Neurol. 2021; 8: 1292–1307. https://doi.org/10.1002/acn3.51367 PMID: 33955722

85. Wang J, Lv B, Chen X, Pan Y, Chen K, Zhang Y, et al. An early model to predict the risk of gestational diabetes mellitus in the absence of blood examination indexes: application in primary health care centres. BMC Pregnancy Childbirth. 2021; 21: 814. https://doi.org/10.1186/s12884-021-04295-2 PMID: 34879850

86. Williamson L, Wojcik C, Taunton M, McElheran K, Howard W, Staszak D, et al. Finding Undiagnosed Patients With Familial Hypercholesterolemia in Primary Care Usingelectronic Health Records. J Am Coll Cardiol. 2020; 75: 3502. https://doi.org/10.1016/s0735-1097(20)34129-2

87. DelPozo-Banos M, John A, Petkov N, Berridge DM, Southern K, Loyd KL, et al. Using neural networks with routine health records to identify suicide risk: Feasibility study. JMIR Ment Heal. 2018; 5: e10144. https://doi.org/10.2196/10144 PMID: 29934287

88. Penfold RB, Johnson E, Shortreed SM, Ziebell RA, Lynch FL, Clarke GN, et al. Predicting suicide attempts and suicide deaths among adolescents following outpatient visits. J Affect Disord. 2021; 294: 39–47. https://doi.org/10.1016/j.jad.2021.06.057 PMID: 34265670

89. van Mens K, Elzinga E, Nielen M, Lokkerbol J, Poortvliet R, Donker G, et al. Applying machine learning on health record data from general practitioners to predict suicidality. Internet Interv. 2020; 21: 100337. https://doi.org/10.1016/j.invent.2020.100337 PMID: 32944503

90. Shih CC, Lu CJ, Chen G Den, Chang CC. Risk prediction for early chronic kidney disease: Results from an adult health examination program of 19,270 individuals. Int J Environ Res Public Health. 2020; 17: 1–11. https://doi.org/10.3390/ijerph17144973 PMID: 32664271

91. Zhao J, Gu S, McDermaid A. Predicting outcomes of chronic kidney disease from EMR data based on Random Forest Regression. Math Biosci. 2019; 310: 24–30. https://doi.org/10.1016/j.mbs.2019.02.001 PMID: 30768948

92. Dinga R, Marquand AF, Veltman DJ, Beekman ATF, Schoevers RA, van Hemert AM, et al. Predicting the naturalistic course of depression from a wide range of clinical, psychological, and biological data: a machine learning approach. Transl Psychiatry. 2018; 8: 241. https://doi.org/10.1038/s41398-018-0289-1 PMID: 30397196

93. Ford E, Rooney P, Oliver S, Hoile R, Hurley P, Banerjee S, et al. Identifying undetected dementia in UK primary care patients: A retrospective case-control study comparing machine-learning and standard epidemiological approaches. BMC Med Inform Decis Mak. 2019; 19: 248. https://doi.org/10.1186/s12911-019-0991-9 PMID: 31791325

94. Ford E, Starlinger J, Rooney P, Oliver S, Banerjee S, van Marwijk H, et al. Could dementia be detected from UK primary care patients' records by simple automated methods earlier than by the treating physician? A retrospective case-control study. Wellcome Open Res. 2020; 5: 120. https://doi.org/10.12688/wellcomeopenres.15903.1 PMID: 32766457

95. Ford E, Sheppard J, Oliver S, Rooney P, Banerjee S, Cassell JA. Automated detection of patients with dementia whose symptoms have been identified in primary care but have no formal diagnosis: A retrospective case-control study using electronic primary care records. BMJ Open. 2021; 11: e039248. https://doi.org/10.1136/bmjopen-2020-039248 PMID: 33483436

96. Fouladvand S, Mielke MM, Vassilaki M, St. Sauver J, Petersen RC, Sohn S. Deep Learning Prediction of Mild Cognitive Impairment using Electronic Health Records. 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE; 2019. pp. 799–806.

97. Haun MW, Simon L, Sklenarova H, Zimmermann-Schlegel V, Friederich H-CHC, Hartmann M. Predicting anxiety in cancer survivors presenting to primary care–A machine learning approach accounting for physical comorbidity. Cancer Med. 2021; 10: 5001–5016. https://doi.org/10.1002/cam4.4048 PMID: 34076372

98. Jammeh EA, Carroll CB, Pearson Stephen W, Escudero J, Anastasiou A, Zhao P, et al. Machine-learning based identification of undiagnosed dementia in primary care: A feasibility study. BJGP Open. 2018; 2: https://doi.org/10.3399/bjgpopen18X101589 PMID: 30564722

99. Jin H, Wu S. Use of patient-reported data to match depression screening intervals with depression risk profiles in primary care patients with diabetes: Development and validation of prediction models for major depression. JMIR Form Res. 2019; 3: e13610–e13610. https://doi.org/10.2196/13610 PMID: 31573900

100. Kaczmarek E, Salgo A, Zafari H, Kosowan L, Singer A, Zulkernine F. Diagnosing PTSD using electronic medical records from Canadian primary care data. ACM International Conference Proceeding Series. School of Computing, Queen's University, Kingston, Canada; 2019. pp. 23–29.

101. Ljubic B, Roychoudhury S, Cao XH, Pavlovski M, Obradovic S, Nair R, et al. Influence of medical domain knowledge on deep learning for Alzheimer's disease prediction. Comput Methods Programs Biomed. 2020; 197: 105765. https://doi.org/10.1016/j.cmpb.2020.105765 PMID: 33011665

102. Mallo SC, Valladares-Rodriguez S, Facal D, Lojo-Seoane C, Fernández-Iglesias MJ, Pereiro AX. Neuropsychiatric symptoms as predictors of conversion from MCI to dementia: A machine learning approach. Int Psychogeriatrics. 2020; 32: 381–392. https://doi.org/10.1017/S1041610219001030 PMID: 31455461

103. Mar J, Gorostiza A, Ibarrondo O, Cernuda C, Arrospide A, Iruin A, et al. Validation of Random Forest Machine Learning Models to Predict Dementia-Related Neuropsychiatric Symptoms in Real-World Data. J Alzheimer's Dis. 2020; 77: 855–864. https://doi.org/10.3233/JAD-200345 PMID: 32741825

104. Półchłopek O, Koning NR, Büchner FL, Crone MR, Numans ME, Hoogendoorn M. Quantitative and temporal approach to utilising electronic medical records from general practices in mental health

prediction. Comput Biol Med. 2020;125. https://doi.org/10.1016/j.compbiomed.2020.103973 PMID: 32916386

105. Shen X, Wang G, Rick Yiu-Cho Kwan, Choi KS. Using dual neural network architecture to detect the risk of dementia with community health data: Algorithm development and validation study. JMIR Med Informatics. 2020; 8: e19870. https://doi.org/10.2196/19870 PMID: 32865498

106. Suárez-Araujo CP, García Báez P, Cabrera-León Y, Prochazka A, Rodríguez Espinosa N, Fernández Viadero C, et al. A Real-Time Clinical Decision Support System, for Mild Cognitive Impairment Detection, Based on a Hybrid Neural Architecture. Bangyal WH, editor. Comput Math Methods Med. 2021; 2021: 1–9. https://doi.org/10.1155/2021/5545297 PMID: 34257699

107. Zafari H, Kosowan L, Zulkernine F, Signer A. Diagnosing post-traumatic stress disorder using electronic medical record data. Health Informatics J. 2021; 27. https://doi.org/10.1177/14604582211053259 PMID: 34818936

108. Emir B, Mardekian J, Masters ET, Clair A, Kuhn M, Silverman SL. Predictive modeling of a fibromyalgia diagnosis: Increasing the accuracy using real world data. Meeting: 2014 ACR/ARHP Annual Meeting. ACR; 2014.

109. Jarvik JG, Gold LS, Tan K, Friedly JL, Nedeljkovic SS, Comstock BA, et al. Long-term outcomes of a large, prospective observational cohort of older adults with back pain. Spine J. 2018; 18: 1540–1551. https://doi.org/10.1016/j.spinee.2018.01.018 PMID: 29391206

110. Kennedy J, Kennedy N, Cooksey R, Choy E, Siebert S, Rahman M, et al. Predicting a diagnosis of ankylosing spondylitis using primary care health records–a machine learning approach. medRxiv. 2021; 2021.04.22.21255659.

111. Kop R, Hoogendoorn M, Teije A ten, Büchner FL, Slottje P, Moons LMG, et al. Predictive modeling of colorectal cancer using a dedicated pre-processing pipeline on routine electronic medical records. Comput Biol Med. 2016; 76: 30–38. https://doi.org/10.1016/j.compbiomed.2016.06.019 PMID: 27392227

112. Malhotra A, Rachet B, Bonaventure A, Pereira SP, Woods LM. Can we screen for pancreatic cancer? Identifying a sub-population of patients at high risk of subsequent diagnosis using machine learning techniques applied to primary care data. PLoS One. 2021; 16: e0251876–e0251876. https://doi.org/10.1371/journal.pone.0251876 PMID: 34077433

113. Ristanoski G, Emery J, Gutierrez JM, McCarthy D, Aickelin U. Primary Care Datasets for Early Lung Cancer Detection: An AI Led Approach. Lecture Notes in Computer Science. AIME; 2021. pp. 83–92. https://doi.org/10.1007/978-3-030-77211-6_9

114. Cox AP, Raluy M, Wang M, Bakheit AMO, Moore AP, Dinet J, et al. Predictive analysis for identifying post stroke spasticity patients in UK primary care data. Pharmacoepidemiol Drug Saf. 2014; 23: 422–423.

115. Hrabok M, Engbers JDT, Wiebe S, Sajobi TT, Subota A, Almohawes A, et al. Primary care electronic medical records can be used to predict risk and identify potentially modifiable factors for early and late death in adult onset epilepsy. Epilepsia. 2021; 62: 51–60. https://doi.org/10.1111/epi.16738 PMID: 33316095

116. Kwasny MJ, Oleske DM, Zamudio J, Diegidio R, Höglinger GU. Clinical Features Observed in General Practice Associated With the Subsequent Diagnosis of Progressive Supranuclear Palsy. Front Neurol. 2021; 12: 637176. https://doi.org/10.3389/fneur.2021.637176 PMID: 33967937

117. Afzal Z, Engelkes M, Verhamme KMC, Janssens HM, Sturkenboom MCJM, Kors JA, et al. Automatic generation of case-detection algorithms to identify children with asthma from large electronic health record databases. Pharmacoepidemiol Drug Saf. 2013; 22: 826–833. https://doi.org/10.1002/pds.3438 PMID: 23592573

118. Doyle OM, van der Laan R, Obradovic M, McMahon P, Daniels F, Pitcher A, et al. Identification of potentially undiagnosed patients with nontuberculous mycobacterial lung disease using machine learning applied to primary care data in the UK. Eur Respir J. 2020/05/21. 2020; 56: 2000045. https://doi.org/10.1183/13993003.00045-2020 PMID: 32430411

119. Weisman A, Tu K, Young J, Kumar M, Austin PC, Jaakkimainen L, et al. Validation of a type 1 diabetes algorithm using electronic medical records and administrative healthcare data to study the population incidence and prevalence of type 1 diabetes in Ontario, Canada. BMJ Open Diabetes Res Care. 2020; 8. https://doi.org/10.1136/bmjdrc-2020-001224 PMID: 32565422

120. Amit G, Girshovitz I, Marcus K, Zhang Y, Pathak J, Bar V, et al. Estimation of postpartum depression risk from electronic health records using machine learning. BMC Pregnancy Childbirth. 2021; 21: 630. https://doi.org/10.1186/s12884-021-04087-8 PMID: 34535116

121. Perlis RH. A clinical risk stratification tool for predicting treatment resistance in major depressive disorder. Biol Psychiatry. 2013; 74: 7–14. https://doi.org/10.1016/j.biopsych.2012.12.007 PMID: 23380715

122. Fernández-Gutiérrez F, Kennedy JI, Cooksey R, Atkinson M, Choy E, Brophy S, et al. Mining Primary Care Electronic Health Records for Automatic Disease Phenotyping: A Transparent Machine Learning Framework. Diagnostics. 2021; 11: 1908. https://doi.org/10.3390/diagnostics11101908 PMID: 34679609

123. Jorge A, Castro VM, Barnado A, Gainer V, Hong C, Cai T, et al. Identifying lupus patients in electronic health records: Development and validation of machine learning algorithms and application of rule-based algorithms. Semin Arthritis Rheum. 2019; 49: 84–90. https://doi.org/10.1016/j.semarthrit.2019.01.002 PMID: 30665626

124. Zhou S-M, Fernandez-Gutierrez F, Kennedy J, Cooksey R, Atkinson M, Denaxas S, et al. Defining Disease Phenotypes in Primary Care Electronic Health Records by a Machine Learning Approach: A Case Study in Identifying Rheumatoid Arthritis. Pappalardo F, editor. PLoS One. 2016; 11: e0154515. https://doi.org/10.1371/journal.pone.0154515 PMID: 27135409

125. Kinar Y, Kalkstein N, Akiva P, Levin B, Half EE, Goldshtein I, et al. Development and validation of a predictive model for detection of colorectal cancer in primary care by analysis of complete blood counts: A binational retrospective study. J Am Med Informatics Assoc. 2016; 23: 879–890. https://doi.org/10.1093/jamia/ocv195 PMID: 26911814

126. Sufriyana H, Wu YW, Su ECY. Artificial intelligence-assisted prediction of preeclampsia: Development and external validation of a nationwide health insurance dataset of the BPJS Kesehatan in Indonesia. EBioMedicine. 2020; 54: 102710. https://doi.org/10.1016/j.ebiom.2020.102710 PMID: 32283530

127. Abràmoff MD, Lavin PT, Birch M, Shah N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. NPJ Digit Med. 2019/07/16. 2018; 1: 39. https://doi.org/10.1038/s41746-018-0040-6 PMID: 31304320

128. Bhaskaranand M, Ramachandra C, Bhat S, Cuadros J, Nittala MG, Sadda SR, 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. 2019; 21: 635–643. https://doi.org/10.1089/dia.2019.0164 PMID: 31335200

129. González-Gonzalo C, Sánchez-Gutiérrez V, Hernández-Martínez P, Contreras I, Lechanteur YT, Domanian A, et al. Evaluation of a deep learning system for the joint automated detection of diabetic retinopathy and age-related macular degeneration. Acta Ophthalmol. 2020; 98: 368–377. https://doi.org/10.1111/aos.14306 PMID: 31773912

130. Kanagasingam Y, Xiao D, Vignarajan J, Preetham A, Tay-Kearney ML, Mehrotra A. Evaluation of Artificial Intelligence-Based Grading of Diabetic Retinopathy in Primary Care. JAMA Netw open. 2018; 1: e182665. https://doi.org/10.1001/jamanetworkopen.2018.2665 PMID: 30646178

131. Hoogendoorn M, Szolovits P, Moons LMG, Numans ME. Utilizing uncoded consultation notes from electronic medical records for predictive modeling of colorectal cancer. Artif Intell Med. 2016; 69: 53–61. https://doi.org/10.1016/j.artmed.2016.03.003 PMID: 27085847

132. Morales DR, Flynn R, Zhang J, Trucco E, Quint JK, Zutis K. External validation of ADO, DOSE, COTE and CODEX at predicting death in primary care patients with COPD using standard and machine learning approaches. Respir Med. 2018; 138: 150–155. https://doi.org/10.1016/j.rmed.2018.04.003 PMID: 29724388

133. Alexander N, Alexander DC, Barkhof F, Denaxas S. Identifying and evaluating clinical subtypes of Alzheimer's disease in care electronic health records using unsupervised machine learning. BMC Med Inform Decis Mak. 2021; 21. https://doi.org/10.1186/s12911-021-01693-6 PMID: 34879829

134. Nikolaou V, Massaro S, Garn W, Fakhimi M, Stergioulas L, Price DB. Fast decliner phenotype of chronic obstructive pulmonary disease (COPD): Applying machine learning for predicting lung function loss. BMJ Open Respir Res. 2021;8. https://doi.org/10.1136/bmjresp-2021-000980 PMID: 34716217

135. Pikoula M, Quint JK, Nissen F, Hemingway H, Smeeth L, Denaxas S. Identifying clinically important COPD sub-types using data-driven approaches in primary care population based electronic health records. BMC Med Inform Decis Mak. 2019; 19: 86. https://doi.org/10.1186/s12911-019-0805-0 PMID: 30999919

136. Collins GS, Reitsma JB, Altman DG, Moons KGMM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. Ann Intern Med. 2015; 162: 55–63. https://doi.org/10.7326/M14-0697 PMID: 25560714

137. Nickel B, Barratt A, Copp T, Moynihan R, McCaffery K. Words do matter: a systematic review on how different terminology for the same condition influences management preferences. BMJ Open. 2017; 7: e014129. https://doi.org/10.1136/bmjopen-2016-014129 PMID: 28698318

138. Ghassemi M, Naumann T, Schulam P, Beam AL, Chen IY, Ranganath R. A Review of Challenges and Opportunities in Machine Learning for Health. AMIA Joint Summits on Translational Science proceedings. AMIA Joint Summits on Translational Science American Medical Informatics Association; 2020 pp. 191–200.

139. Fernández-Delgado Manuel, Cernadas Eva, Barro Senén, Amorim Dinani. Do we need hundreds of classifiers to solve real world classification problems? J Mach Learn Res. 2014. https://doi.org/10.5555/2627435.2697065

140. Juarez-Orozco LE, Martinez-Manzanera O, Nesterov S V., Kajander S, Knuuti J. The machine learning horizon in cardiac hybrid imaging. Eur J Hybrid Imaging. 2018; 2: 1–15. https://doi.org/10.1186/S41824-018-0033-3/FIGURES/5

141. Challen R, Denny J, Pitt M, Gompels L, Edwards T, Tsaneva-Atanasova K. Artificial intelligence, bias and clinical safety. BMJ Qual Saf. 2019; 28: 231–237. https://doi.org/10.1136/bmjqs-2018-008370 PMID: 30636200

142. Higgins JP. Nonlinear systems in medicine. Yale J Biol Med. 2002; 75: 247–260. PMID: 14580107

143. Decherchi S, Pedrini E, Mordenti M, Cavalli A, Sangiorgi L. Opportunities and Challenges for Machine Learning in Rare Diseases. Front Med. 2021; 8: 747612. https://doi.org/10.3389/fmed.2021.747612 PMID: 34676229

144. Bozkurt S, Cahan EM, Seneviratne MG, Sun R, Lossio-Ventura JA, Ioannidis JPA, et al. Reporting of demographic data and representativeness in machine learning models using electronic health records. J Am Med Informatics Assoc. 2020; 27: 1878–1884. https://doi.org/10.1093/jamia/ocaa164 PMID: 32935131

145. Andaur Navarro CL, Damen JAA, Takada T, Nijman SWJ, Dhiman P, Ma J, et al. Completeness of reporting of clinical prediction models developed using supervised machine learning: a systematic review. BMC Med Res Methodol. 2022; 22: 1–13. https://doi.org/10.1186/S12874-021-01469-6/FIGURES/3

146. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med. 2019; 17: 195. https://doi.org/10.1186/s12916-019-1426-2 PMID: 31665002

147. El-Sherbini AH, Hassan Virk HU, Wang Z, Glicksberg BS, Krittanawong C. Machine-Learning-Based Prediction Modelling in Primary Care: State-of-the-Art Review. Ai. 2023; 4: 437–460. https://doi.org/10.3390/ai4020024

148. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: A multidisciplinary view. J Med Internet Res. 2016; 18: e5870. https://doi.org/10.2196/jmir.5870 PMID: 27986644

149. Collins GS, Moons KGMM, GS Collins KM. Reporting of artificial intelligence prediction models. Lancet (London, England). 2019; 393: 1577–1579. https://doi.org/10.1016/S0140-6736(19)30037-6 PMID: 31007185

150. Gentil M-L, Cuggia M, Fiquet L, Hagenbourger C, Le Berre T, Banâtre A, et al. Factors influencing the development of primary care data collection projects from electronic health records: A systematic review of the literature. BMC Med Inform Decis Mak. 2017; 17. https://doi.org/10.1186/s12911-017-0538-x PMID: 28946908

151. Cabitza F, Rasoini R, Gensini GF. Unintended Consequences of Machine Learning in Medicine. JAMA. 2017; 318: 517–518. https://doi.org/10.1001/jama.2017.7797 PMID: 28727867

152. McDonald L, Ramagopalan S V., Cox AP, Oguz M. Unintended consequences of machine learning in medicine? F1000Research. 2017; 6. https://doi.org/10.12688/f1000research.12693.1 PMID: 29250316

153. Bakker L, Aarts J, Uyl-de Groot C, Redekop W, Groot CUD, Redekop W. Economic evaluations of big data analytics for clinical decision-making: A scoping review. J Am Med Informatics Assoc. 2020; 27: 1466–1475. https://doi.org/10.1093/jamia/ocaa102 PMID: 32642750

154. Williamson T, Aponte-Hao S, Mele B, Lethebe BC, Leduc C, Thandi M, et al. Developing and validating a primary care EMR-based frailty definition using machine learning. Int J Popul Data Sci. 2020; 5: 1344. https://doi.org/10.23889/ijpds.v5i1.1344 PMID: 32935059

155. Collins GS, Dhiman P, Andaur Navarro CL, Ma J, Hooft L, Reitsma JB, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. BMJ Open. 2021; 11: e048008. https://doi.org/10.1136/bmjopen-2020-048008 PMID: 34244270