Journal of the American Heart Association

ORIGINAL RESEARCH

Survey and Evaluation of Hypertension Machine Learning Research

Clea du Toit, MSc*; Tran Quoc Bao Tran, MSc*; Neha Deo, BS; Sachin Aryal, MS; Stefanie Lip, MBChB; Robert Sykes, MBChB; Ishan Manandhar, MS; Aristeidis Sionakidis, MSc; Leah Stevenson, MS; Harsha Pattnaik, MBBS; Safaa Alsanosi, PhD; Maria Kassi, MSc; Ngoc Le, MSc; Maggie Rostron, MBChB; Sarah Nichol, MBChB; Alisha Aman, MSc; Faisal Nawaz, MBBS; Dhruven Mehta, MD; Ramakumar Tummala, PhD; Linsay McCallum, PhD; Sandeep Reddy, MBBS, PhD; Shyam Visweswaran, MD, PhD; Rahul Kashyap, MD, MBA; Bina Joe, PhD; Sandosh Padmanabhan, MD, PhD

BACKGROUND: Machine learning (ML) is pervasive in all fields of research, from automating tasks to complex decision-making. However, applications in different specialities are variable and generally limited. Like other conditions, the number of studies employing ML in hypertension research is growing rapidly. In this study, we aimed to survey hypertension research using ML, evaluate the reporting quality, and identify barriers to ML's potential to transform hypertension care.

METHODS AND RESULTS: The Harmonious Understanding of Machine Learning Analytics Network survey questionnaire was applied to 63 hypertension-related ML research articles published between January 2019 and September 2021. The most common research topics were blood pressure prediction (38%), hypertension (22%), cardiovascular outcomes (6%), blood pressure variability (5%), treatment response (5%), and real-time blood pressure estimation (5%). The reporting quality of the articles was variable. Only 46% of articles described the study population or derivation cohort. Most articles (81%) reported at least 1 performance measure, but only 40% presented any measures of calibration. Compliance with ethics, patient privacy, and data security regulations were mentioned in 30 (48%) of the articles. Only 14% used geographically or temporally distinct validation data sets. Algorithmic bias was not addressed in any of the articles, with only 6 of them acknowledging risk of bias.

CONCLUSIONS: Recent ML research on hypertension is limited to exploratory research and has significant shortcomings in reporting quality, model validation, and algorithmic bias. Our analysis identifies areas for improvement that will help pave the way for the realization of the potential of ML in hypertension and facilitate its adoption.

Key Words: artificial intelligence ■ hypertension ■ machine learning ■ reporting quality

Recent advances in computational power and the availability of larger and more comprehensive medical data sets have led to an increase in machine learning (ML) in clinical research, which could transform health care. Despite the rapid increase in research and evidence that ML models outperform clinicians in areas such as arrhythmia detection and clinical image processing, the actual impact on health care has been limited.[1-3] Hypertension is the single most important modifiable risk factor worldwide, causing nearly 10 million deaths annually in both high- and low-income countries. The management of hypertension, from screening to diagnosis to treatment, presents a number of obstacles that call for transformational solutions in which ML may play a

Correspondence to: Sandosh Padmanabhan, MD, PhD, School of Cardiovascular and Metabolic Health, University of Glasgow, Glasgow G12 8TA, UK. Email: sandosh.padmanabhan@glasgow.ac.uk and Bina Joe, PhD, Center for Hypertension and Precision Medicine, Department of Physiology and Pharmacology, University of Toledo College of Medicine and Life Sciences, Toledo, OH 43614. Email: bina.joe@utoledo.edu

*C. du Toit and T. Q. B. Tran contributed equally.

This article was sent to Yen-Hung Lin, MD, PhD, Associate Editor, for review by expert referees, editorial decision, and final disposition.

Supplemental Material is available at https://www.ahajournals.org/doi/suppl/10.1161/JAHA.122.027896

For Sources of Funding and Disclosures, see page 7.

JAHA is available at: www.ahajournals.org/journal/jaha

CLINICAL PERSPECTIVE

What Is New?

- The number of hypertension research studies employing machine learning (ML) is increasing quickly, but, as in other clinical domains, there has been almost no translation into clinical practice, and there are concerns about the robustness and generalizability of ML models applied to diverse populations, as well as the quality and reporting of ML methods and results in the clinical research.
- Frameworks are now being developed for conduct and reporting of clinical ML research.
- Because of the variety of input data and the customizability of ML methods, disease- and domain-specific recommendations are to be required for ML.

What Are the Clinical Implications?

- Our analysis of recent hypertension ML publications identifies areas for improvement in reporting, which should inform and support hypertension researchers who are using or planning to use ML.
- This study will help clinicians evaluate commercial ML tools for clinical use effectively and thus minimize patient harm and improve clinical service.

Nonstandard Abbreviations and Acronyms

ML	machine learning

role.[4,5] In fact, the number of research studies employing ML is increasing quickly, but, as in other clinical domains, there has been almost no translation into clinical practice. A solid understanding of the clinical domain, data science, implementation, and regulatory requirements are required to develop ML solutions.[6] Concerns about the robustness and generalizability of models applied to diverse populations, as well as the quality and accessibility of reporting ML methods and results, are growing as ML models in medicine are developed. The evaluations of bias, transparency, and reporting of ML research in a number of medical fields are unstandardized and amenable to improvement. Algorithmic bias (the representation of diversity in input data versus the target algorithm deployment population) is of particular concern for ML in medicine.[7,8] Previously, statistical clinical risk prediction models faced similar challenges, which were addressed by the creation of standardized analysis and reporting frameworks.[3,9,10,11] Similar frameworks are now being developed for clinical ML tools.[12,13] These novel frameworks must consider clinical utility and impact on both the patient and physician, as well as the rapidly evolving range of ML approaches and the data used to develop the models. Because of the variety of input data and the customizability of ML methods, disease- and domain-specific recommendations are likely to be required for ML. While broad research and reporting guidelines are appropriate for more traditional prediction models, disease- and domain-specific recommendations are likely to be necessary for ML.

In this study, we aimed to survey the spectrum of hypertension research employing ML, evaluate the quality of their reporting, and gain insight into the obstacles impeding the realization of ML's potential to transform hypertension care. Understanding where ML has been applied and its limitations will inform the design and reporting of future ML studies that can transform hypertension care.

METHODS

Our goal was to assess the topics covered in hypertension ML research and the current standard of communication of clinical ML research in hypertension using a custom survey developed by incorporating recommendations from existing checklists. The data that support the findings of this study are available from the corresponding author upon reasonable request. Institutional review board approval for this study was not required as this is a survey of published studies.

Identification and Selection of Articles

A search was conducted across 3 widely used databases (Embase, PubMed, and Google Scholar) using 2 groups of medical subject headings search terms: those pertaining to hypertension (eg, "blood pressure," "hypertension," "ambulatory blood pressure monitoring") and those pertaining to ML (eg, "machine learning," "supervised machine learning," "deep learning"). Non–medical subject headings search terms (eg, "random forest" and "Boltzmann machine") were also included in the ML group. The inclusion criteria for search results were peer-reviewed original research, publication date between January 2019 and September 2021, full text availability (either for free or via institutional access), and original English text. The articles were reviewed manually by separate teams at the Universities of Glasgow and Toledo. Selected articles were pooled, and those not meeting eligibility criteria were removed.

Development of the Harmonious Understanding of Machine Learning Analytics Network Survey Questionnaire

A PubMed search identified ML reporting and evaluation frameworks published between January 2015 and

February 2020. A group of ML specialists and hypertension researchers reviewed frameworks ranging from narrow domain-specific to broader high-level checklists.[3,9,12,13,14] Based on this review, a list of survey items was generated and developed into the Harmonious Understanding of Machine Learning Analytics Network survey through an iterative Delphi process. The final survey contains 60 questions with binary, multiple choice, or free-text responses (Table S1). Free-text sections were included to provide additional comments or elaborate when responses like "Other" were selected in multiple-choice questions.

Survey Procedures

The Harmonious Understanding of Machine Learning Analytics Network survey was implemented in REDCap,[15] which is a secure web application for building and managing online surveys. Two researchers (C.D.T. and T.Q.B.T.) read all the papers and completed the survey. In addition, 18 reviewers reflecting the typical readership of cardiovascular research journals also completed the survey. Reviewers were required to have experience with health care data but not with ML. Each article was reviewed by 2 randomly allocated reviewers who independently applied the Harmonious Understanding of Machine Learning Analytics Network survey to the article. Discordance was resolved with the opinion of a third reviewer with ML experience (C.D.T. or T.Q.B.T.). Responses were analyzed for each survey item. Adherence (ie, the proportion of articles that satisfied the questionnaire requirements) was calculated for each individual survey item. Qualitative results were grouped into 9 domains (clinical relevance; defining and addressing the knowledge gap [rationale]; prespecified study design; data suitability; ground truth [basis of supervised machine learning labeling]; performance metrics; replication and validation; ethical, legal, and social implications; and reporting quality). Data from REDCap were analyzed and visualized using the R programming language version 4.1.1 (R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

The search strategy identified 63 articles that applied ML in hypertension research. A Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram outlining the selection process is presented in Figure 1. A list of the articles with main ML methods and objectives is presented in Table S2. The research objectives and data are summarized in Figure 2. The most frequent research aims were blood pressure (BP) prediction (38%), hypertension (22%), cardiovascular outcomes (6%), BP variability (5%), treatment response (5%), and real-time BP estimation (5%).

The main data type used in each study was also identified. The most frequently used input data were routine clinical and demographic data retrieved from medical health records (44%). Thirty percent of studies chiefly used data from noninvasive methods, such as auscultatory or oscillometric BP measurements, photoplethysmography, or electrocardiography. Two of the 5 studies that reported using data from wearable devices did not specify how the measurements were made (eg, Apple watch uses photoplethysmography, but this was not specified).

Survey Responses

All survey questions and responses are presented in Table S1.

Traditional Components of Scientific Papers and Clinical Relevance

Fifty-three of 63 articles (84%) described the relevance of their project in terms of clinical impact (potential savings in cost, lives, or time), and 89% described the rationale for the project and the knowledge gap being addressed. A notable exception to standard reporting requirements was the absence of a description of input data or cohort demographics in many articles (presented in 46% of articles).

Prespecified Analysis Plan; Data; Validation; and Ethical, Legal, and Social Implications

In 59 of the 63 studies, the data sets used were deemed appropriate for the investigation, but in only 30 of the studies were the data obtained from the intended stage of the care pathway if the results were to be implemented. Most studies (44; 70%) also presented a prespecified statistical analysis plan, and 63 studies explained data preprocessing and curation steps.

Internal validation methods, such as cross-fold validation or use of independent training and testing data sets, were described in 73% of studies. External validation with geographically or temporally distinct data sets was carried out in 9 (14%) studies.

Compliance with ethical, patient privacy, and data security regulations were mentioned in 30 (48%) articles. Of the 39 prospective or interventional trials that were deemed by reviewers to require informed consent, acquiring patient consent was mentioned in 17 (44%) studies. Algorithmic bias was not rigorously addressed in any of the reviewed studies, with only 6 articles acknowledging a risk of bias.

Ground Truth and Performance Metrics

Almost all of the studies (58; 92%) applied supervised learning techniques requiring the establishment of

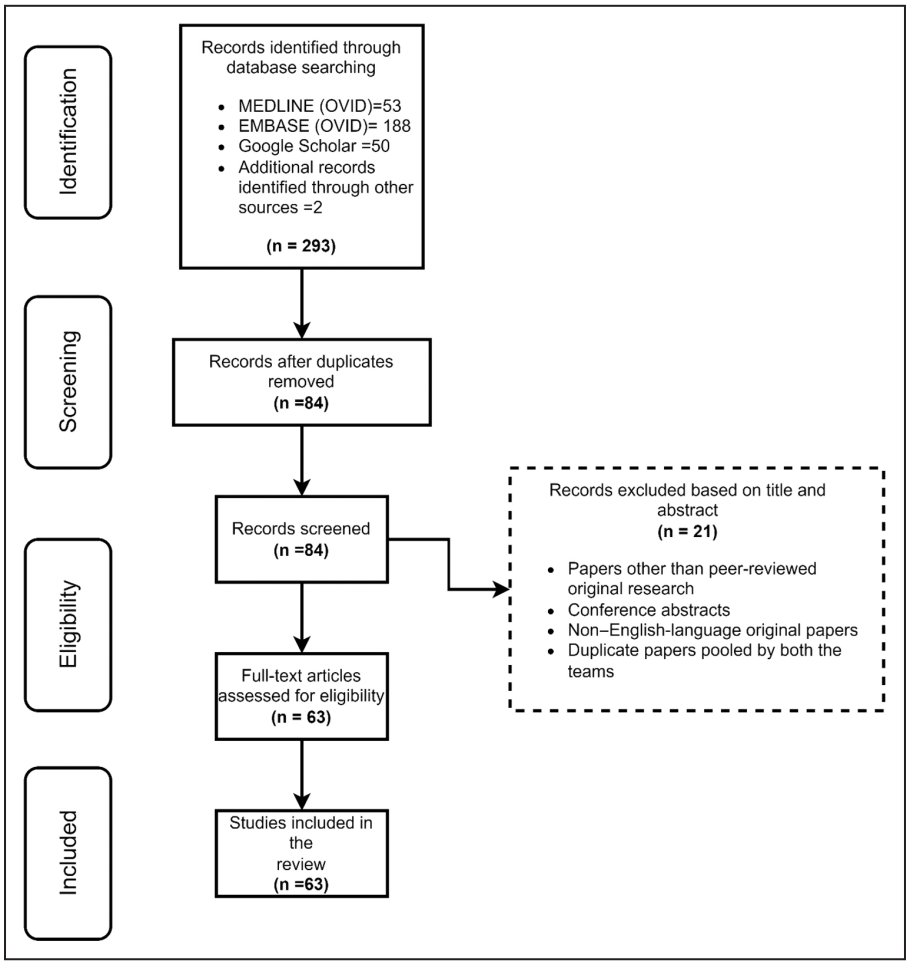

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram of the article screening and identification process.

ground truth for analysis. Ground truth labels in 47 of the 58 studies (81%) were sufficiently explained and backed by guidelines or references.

Most articles (51; 81%) reported at least 1 model performance measure (eg, accuracy, sensitivity, or area under the receiver operating characteristic curve). In contrast, a minority (40%) presented any calibration measures (eg, calibration plot, Hosmer–Lemeshow test, or Brier scores), and 37 of 63 studies described measures to address overfitting.

General Readership Survey

Figure 3 shows the percentage of concordance between non-ML expert reviewers representing real-world readership of the research articles. The highest concordance was seen for items with which the readership is expected to be familiar (namely, general publication quality questions). Lower concordance was observed for questions that covered technical clinical or ML aspects; for example, only 50% of reviewers agreed with their counterpart when assessing items related to overfitting.

DISCUSSION

Our survey of hypertension-related publications over a 33-month period showed that ML use is limited to exploratory research and has significant shortcomings in reporting quality, model validation, and algorithmic bias. Our analysis identifies areas for improvement that will facilitate the full realization of the potential of ML in hypertension and facilitate its adoption.

The most common research topics were BP prediction, hypertension, and cardiovascular risk, all of which are unquestionably important; however, most of the studies were exploratory and have low translational potential due to the need for multiple validations in independent data sets and long follow-up for definitive outcomes. Successful applications of ML include the automation of tasks and the management of chronic diseases such

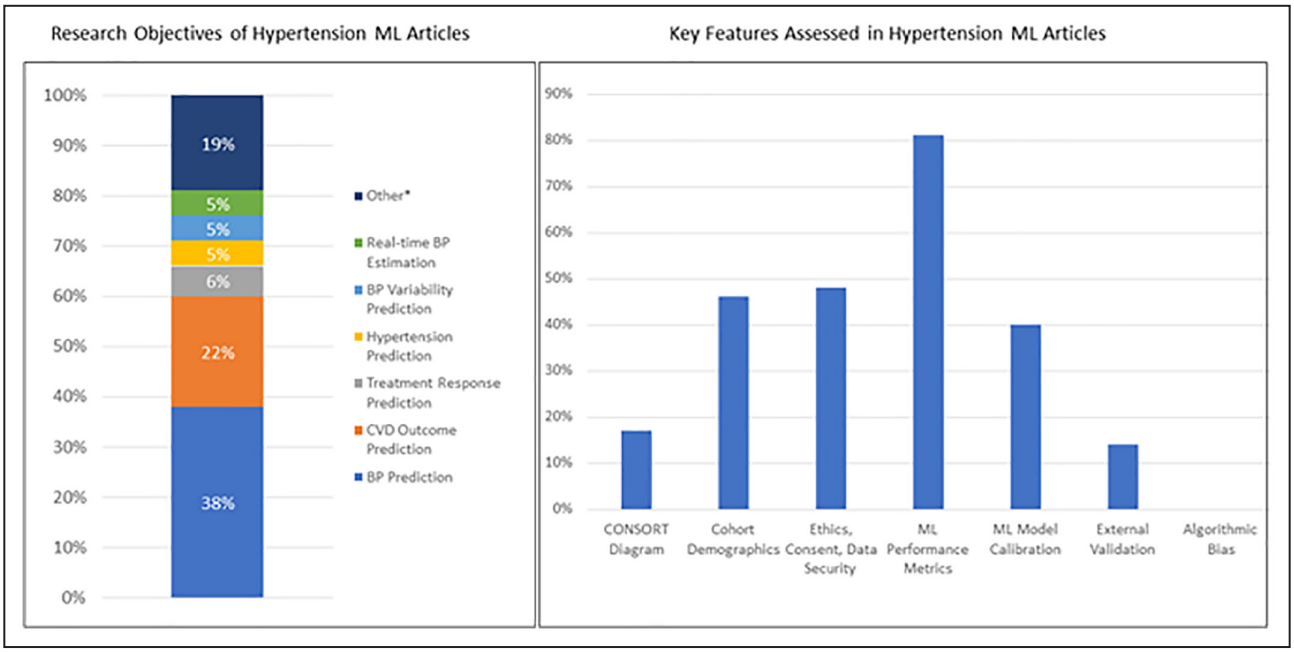

Figure 2. Research objectives of hypertension ML articles (left) and key findings of applying the Harmonious Understanding of Machine Learning Analytics Network survey to hypertension ML articles (right).
*"Other" category includes objectives such as medication adherence, hypertension classification, risk stratification, and investigating the association of BP with the microbiome. BP indicates blood pressure; CONSORT, Consolidated Standards of Reporting Trials; CVD, cardiovascular disease; and ML, machine learning.

as hypertension. These may be the "low-hanging fruit" of implementable ML for the clinical management of hypertension, and studies examining adherence, managing follow-up, monitoring home BP, risk factor management, treatment titration, and education may yield simple solutions that could revolutionize hypertension care.

ML research imposes additional requirements on its design, execution, and reporting that are essential for establishing confidence in novel applications and accelerating their clinical implementation for the benefit of patients. The reporting must be of high quality to demonstrate scientific rigor and should be understandable to a reader who may not be an expert in ML. The engagement of domain experts is crucial, as they are the source of clinical challenges that ML specialists must address.

Using the most suitable data for the research question is crucial to algorithm development. In both

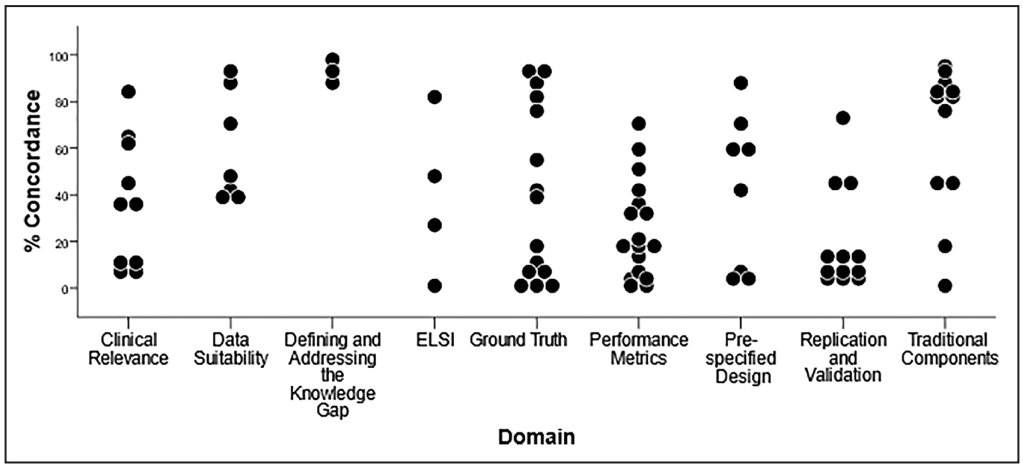

Figure 3. Dot plot showing percentage concordance between non–machine learning expert reviewers, representing real-world readership of the research articles.
Each dot represents 1 question, and its position on the *y* axis represents the concordance between pairs of reviewers. ELSI indicates ethical, legal, and social implications.

prospective and retrospective medical research, best practices, epidemiological research, and other earlier works typically guide the selection of study population and outcome. Frameworks such as Population, Intervention, Comparator, and Outcomes provide guidance on formulating the research question and implementing best practices in clinical research.[10]

Most reviewed studies (89%) employed data sets deemed suitable for the clinical question being investigated. In nearly half of the articles, data selection criteria and study populations were not described. Likewise, 44% of studies lacked adherence to transparency and ethics. It is possible that studies followed regulations but did not explicitly document it.

Presenting data appropriately is essential to convince readers that all efforts were taken to minimize bias.[11] It clarifies the populations to which the study's findings are applicable, which could aid in the future implementation of new interventions or algorithms. Algorithmic bias was the survey item that appeared the least frequently in the articles. Algorithmic bias refers to the extent to which diversity (eg, racial, socioeconomic, sex, and age) is present in the data set used for model development versus the deployment population.[8] Biases in the model's training data may be propagated through its development and eventual deployment, thereby fostering greater inequality. Systematic bias and fairness testing is the first step in informed model selection, which reduces ML-caused inequities.

The most common ML technique in the articles reviewed was supervised learning. As supervised learning depends on models learning from labeled examples, the quality of the ground truth (on which the labels are based) is crucial. Without meticulously selected and labeled data, models cannot be effectively constructed or evaluated. Existing guidelines supported the majority of studies' ground truth labeling, lending credibility to the performance of the resulting models. Studies reported a variety of model performance metrics, but the selection of metrics should be appropriate for the model and the clinical setting in which it will be used.

For prediction models, calibration and discrimination are the minimum requirements for reporting,[2] and only a minority of articles reported calibration. The area under the precision-recall curve should be reported alongside area under the receiver operating characteristic curve metrics for imbalanced data, for which area under the receiver operating characteristic curve metrics were typically reported. Additionally, accuracy and harmonic mean of precision and recall score should be reported, the latter especially when the data set is unbalanced.[2,14]

Most articles viewed overfitting as a threat to the validity of their models. Studies must consider the risk of overfitting as well as countermeasures (eg, oversampling or undersampling). Downsampling is inefficient because reducing the sample size may increase the likelihood of overfitting.[14] Root mean squared error or mean absolute error is recommended for continuous variables. In addition to sample size, number of predictors, and hyperparameter tuning, other factors that influence differences in performance and must therefore be described in detail are sample size, number of predictors, and variance in performance. In varying degrees, these requirements were met in the studies surveyed.

External validation (in geographically or temporally distinct training and validation data sets) is essential before clinical implementation to demonstrate accuracy and generalizability in settings and populations beyond the original derivation population. Typically, external validation studies are anticipated to diminish the predictive accuracy of models. Only 5 studies reported validating the ML model against an external data set in our review. This may be due to a lack of appropriate external data sets or lack of awareness of the importance of external validation. Another explanation may be the belief that splitting the data set into training and testing sets satisfies the need for validation. Here, we stress the importance of having a totally separate test data set or sometimes several separate test sets, with hyperparameter fine-tuning carried out using a validation data set. One needs to be careful with hyperparameter optimization because changing hyperparameters changes the performance of the whole model and may overfit to the peculiarities of the validation set; cross validation may help to some extent, but an independent test set is the ideal solution.

The clinical usefulness, trustworthiness (to both patients and physicians), and explainability of an algorithm all contribute to its clinical adoption. As a result, providing a detailed description of how the proposed ML model aligns on these dimensions would be beneficial for eventual implementation. If applicable to the stage of the study, plans for deployment and commercialization, including regulatory requirements, may need to be considered. Patients and the general public should be involved in research, and there should be a clear strategy in place to evaluate the acceptability of the proposed model and outcomes to the patients providing the data, the clinicians applying the models, and the patients to whom the model will be applied.

The current study has some limitations. First, it is a scoping review, and while every effort was made to capture the full spectrum of publications in the cross section of ML and hypertension research, individual articles may have been overlooked. Second, the Harmonious Understanding of Machine Learning Analytics Network survey omitted some critical ML-related questions, such as data availability, code sharing, transparency, explainability, and interpretability of ML models.

Finally, with the increasing use of ML methods in hypertension research, our analysis of recent

hypertension ML publications identifies areas for improvement in reporting, which should inform and support hypertension researchers who are using or planning to use ML. This will ensure that ML research in hypertension satisfies the global consensus that ML solutions must be fair and nondiscriminatory, while also having a positive impact in all areas of social and economic life.

ARTICLE INFORMATION

Received December 24, 2022; accepted March 27, 2023.

Affiliations

School of Cardiovascular and Metabolic Health, University of Glasgow, Glasgow, United Kingdom (C.d.T., T.Q.T., S.L., R.S., S.A., M.K., N.L., M.R., S.N., A.A., L.M., S.P.); Mayo Clinic Alix School of Medicine, Rochester, MN (N.D.); Center for Hypertension and Precision Medicine, Department of Physiology and Pharmacology, University of Toledo College of Medicine and Life Sciences, Toledo, OH (S.A., I.M., L.S., R.T., B.J.); Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, United Kingdom (A.S.); Lady Hardinge Medical College, New Delhi, India (H.P.); Department of Pharmacology and Toxicology, Faculty of Medicine, Umm Al Qura University, Makkah, Saudi Arabia (S.A.); College of Medicine, Mohammed Bin Rashid University of Medicine and Health Sciences, Dubai, UAE (F.N.); Department of Internal Medicine, TriStar Centennial Medical Center, HCA Healthcare, Nashville, TN (D.M.); School of Medicine, Deakin University, Geelong, Australia (S.R.); Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA (S.V.); and Department of Anesthesiology and Critical Care Medicine, Mayo Clinic, Rochester, MN (R.K.).

Sources of Funding

Dr Padmanabhan is supported by the British Heart Foundation Centre of Excellence Award (RE/18/6/34217) and the United Kingdom Research and Innovation Strength in Places Fund (SIPF00007/1). T.Q.B. Tran is supported by a British Heart Foundation MBPhD Studentship (FS/MBPhD/22/28005). R01HL143082 funding from the National Heart, Lung, and Blood Institute of the National Institutes of Health to B. Joe is gratefully acknowledged.

Disclosures

None.

Supplemental Material

Tables S1–S2
References 17–78

REFERENCES

1. Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, Ng AY. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med*. 2019;25:65–69. doi: 10.1038/s41591-018-0268-3
2. Liu X, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A, Mahendiran T, Moraes G, Shamdas M, Kern C, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit Health*. 2019;1:e271–e297. doi: 10.1016/S2589-7500(19)30123-2
3. Collins GS, Moons KGM. Reporting of artificial intelligence prediction models. *Lancet*. 2019;393:1577–1579. doi: 10.1016/S0140-6736(19)30037-6
4. Dzau VJ, Balatbat CA. Future of hypertension. *Hypertension*. 2019;74:450–457. doi: 10.1161/HYPERTENSIONAHA.119.13437
5. Padmanabhan S, Tran TQB, Dominiczak AF. Artificial intelligence in hypertension: seeing through a glass darkly. *Circ Res*. 2021;128:1100–1118. doi: 10.1161/CIRCRESAHA.121.318106
6. Verma AA, Murray J, Greiner R, Cohen JP, Shojania KG, Ghassemi M, Straus SE, Pou-Prom C, Mamdani M. Implementing machine learning in medicine. *CMAJ*. 2021;193:E1351–E1357. doi: 10.1503/cmaj.202434
7. Brnabic A, Hess LM. Systematic literature review of machine learning methods used in the analysis of real-world data for patient-provider decision making. *BMC Med Inform Decis Mak*. 2021;21:54. doi: 10.1186/s12911-021-01403-2
8. Altman DG, Simera I, Hoey J, Moher D, Schulz K. EQUATOR: reporting guidelines for health research. *Lancet*. 2008;371:1149–1150. doi: 10.1016/S0140-6736(08)60505-X
9. Cruz Rivera S, Liu X, Chan AW, Denniston AK, Calvert MJ; SPIRIT-AI and CONSORT-AI Working Group, SPIRIT-AI and CONSORT-AI Steering Group, SPIRIT-AI and CONSORT-AI Consensus Group. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Nat Med*. 2020;26:1351–1363. doi: 10.1038/s41591-020-1037-7
10. Sounderajah V, Ashrafian H, Aggarwal R, De Fauw J, Denniston AK, Greaves F, Karthikesalingam A, King D, Liu X, Markar SR, et al. Developing specific reporting guidelines for diagnostic accuracy studies assessing AI interventions: the STARD-AI steering group. *Nat Med*. 2020;26:807–808. doi: 10.1038/s41591-020-0941-1
11. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42:377–381. doi: 10.1016/j.jbi.2008.08.010
12. Eriksen MB, Frandsen TF. The impact of patient, intervention, comparison, outcome (PICO) as a search strategy tool on literature search quality: a systematic review. *J Med Libr Assoc*. 2018;106:420–431. doi: 10.5195/jmla.2018.345
13. Guyatt GH, Oxman AD, Kunz R, Atkins D, Brozek J, Vist G, Alderson P, Glasziou P, Falck-Ytter Y, Schunemann HJ. GRADE guidelines: 2. Framing the question and deciding on important outcomes. *J Clin Epidemiol*. 2011;64:395–400. doi: 10.1016/j.jclinepi.2010.09.012
14. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366:447–453. doi: 10.1126/science.aax2342
15. van Smeden M, Moons KG, de Groot JA, Collins GS, Altman DG, Eijkemans MJ, Reitsma JB. Sample size for binary logistic prediction models: beyond events per variable criteria. *Stat Methods Med Res*. 2019;28:2455–2474. doi: 10.1177/0962280218784726Aziz
16. Aziz F, Malek S, Ali AM, Wong MS, Mosleh M, Milow P. Determining hypertensive patients' beliefs towards medication and associations with medication adherence using machine learning methods. *PeerJ*. 2020;8:e8286. doi: 10.7717/peerj.8286
17. Argha A, Celler BG. Blood pressure estimation using time domain features of auscultatory waveforms and deep learning. *Annu Int Conf IEEE Eng Med Biol Soc*. 2019;2019:1821–1824. doi: 10.1109/embc.2019.8857464
18. Argha A, Celler BG, Lovell NH. A novel automated blood pressure estimation algorithm using sequences of Korotkoff sounds. *IEEE J Biomed Health Inform*. 2021;25:1257–1264. doi: 10.1109/jbhi.2020.3012567
19. Pan F, He P, Chen F, Zhang J, Wang H, Zheng D. A novel deep learning based automatic auscultatory method to measure blood pressure. *Int J Med Inform*. 2019;128:71–78. doi: 10.1016/j.ijmedinf.2019.04.023
20. Pan F, He P, Chen F, Pu X, Zhao Q, Zheng D. Deep learning-based automatic blood pressure measurement: evaluation of the effect of deep breathing, talking and arm movement. *Ann Med*. 2019;51:397–403. doi: 10.1080/07853890.2019.1694170
21. Persell SD, Peprah YA, Lipiszko D, Lee JY, Li JJ, Ciolino JD, Karmali KN, Sato H. Effect of home blood pressure monitoring via a smartphone hypertension coaching application or tracking application on adults with uncontrolled hypertension: a randomized clinical trial. *JAMA Netw Open*. 2020;3:e200255. doi: 10.1001/jamanetworkopen.2020.0255
22. Miao F, Wen B, Hu Z, Fortino G, Wang XP, Liu ZD, Tang M, Li Y. Continuous blood pressure measurement from one-channel electrocardiogram signal using deep-learning techniques. *Artif Intell Med*. 2020;108:101919. doi: 10.1016/j.artmed.2020.101919
23. Soh DCK, Ng EYK, Jahmunah V, Oh SL, San TR, Acharya UR. A computational intelligence tool for the detection of hypertension using empirical mode decomposition. *Comput Biol Med*. 2020;118:103630. doi: 10.1016/j.compbiomed.2020.103630
24. Li YH, Harfiya LN, Purwandari K, Lin YD. Real-time cuffless continuous blood pressure estimation using deep learning model. *Sensors (Basel)*. 2020;20:20. doi: 10.3390/s20195606
25. Yan C, Li Z, Zhao W, Hu J, Jia D, Wang H, You T. Novel deep convolutional neural network for cuff-less blood pressure measurement

using ECG and PPG signals. *Annu Int Conf IEEE Eng Med Biol Soc*. 2019;2019:1917–1920. doi: 10.1109/embc.2019.8857108
26. Zhang B, Ren H, Huang G, Cheng Y, Hu C. Predicting blood pressure from physiological index data using the SVR algorithm. *BMC Bioinformatics*. 2019;20:109. doi: 10.1186/s12859-019-2667-y
27. Sannino G, De Falco I, De Pietro G. Non-invasive risk stratification of hypertension: a systematic comparison of machine learning algorithms. *J Sens Actuator Netw*. 2020;9:34. doi: 10.3390/jsan9030034
28. Li C, Sun D, Liu J, Li M, Zhang B, Liu Y, Wang Z, Wen S, Zhou J. A prediction model of essential hypertension based on genetic and environmental risk factors in northern Han Chinese. *Int J Med Sci*. 2019;16:793–799. doi: 10.7150/ijms.33967
29. Widen E, Raben TG, Lello L, Hsu SDH. Machine learning prediction of biomarkers from SNPs and of disease risk from biomarkers in the UK biobank. *Genes (Basel)*. 2021;12:12. doi: 10.3390/genes12070991
30. Kissas G, Yang Y, Hwuang E, Witschey WR, Detre JA, Perdikaris P. Machine learning in cardiovascular flows modeling: predicting arterial blood pressure from non-invasive 4D flow MRI data using physics-informed neural networks. *Comput Methods Appl Mech Eng*. 2020;358:112623. doi: 10.1016/j.cma.2019.112623
31. Lacson RC, Baker B, Suresh H, Andriole K, Szolovits P, Lacson E Jr. Use of machine-learning algorithms to determine features of systolic blood pressure variability that predict poor outcomes in hypertensive patients. *Clin Kidney J*. 2019;12:206–212. doi: 10.1093/ckj/sfy049
32. Barbieri C, Cattinelli I, Neri L, Mari F, Ramos R, Brancaccio D, Canaud B, Stuard S. Development of an artificial intelligence model to guide the management of blood pressure, fluid volume, and dialysis dose in end-stage kidney disease patients: proof of concept and first clinical assessment. *Kidney Dis (Basel)*. 2019;5:28–33. doi: 10.1159/000493479
33. Cho IJ, Sung JM, Kim HC, Lee SE, Chae MH, Kavousi M, Rueda-Ochoa OL, Ikram MA, Franco OH, Min JK, et al. Development and external validation of a deep learning algorithm for prognostication of cardiovascular outcomes. *Korean Circ J*. 2020;50:72–84. doi: 10.4070/kcj.2019.0105
34. Du Z, Yang Y, Zheng J, Li Q, Lin D, Li Y, Fan J, Cheng W, Chen XH, Cai Y. 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 Inform*. 2020;8:e17257. doi: 10.2196/17257
35. Wu CC, Hsu WD, Islam MM, Poly TN, Yang HC, Nguyen PA, Wang YC, Li YJ. An artificial intelligence approach to early predict non-ST-elevation myocardial infarction patients with chest pain. *Comput Methods Programs Biomed*. 2019;173:109–117. doi: 10.1016/j.cmpb.2019.01.013
36. Wu X, Yuan X, Wang W, Liu K, Qin Y, Sun X, Ma W, Zou Y, Zhang H, Zhou X, et al. Value of a machine learning approach for predicting clinical outcomes in young patients with hypertension. *Hypertension*. 2020;75:1271–1278. doi: 10.1161/hypertensionaha.119.13404
37. Bertsimas D, Borenstein A, Mingardi L, Nohadani O, Orfanoudaki A, Stellato B, Wiberg H, Sarin P, Varelmann DJ, Estrada V, et al. Personalized prescription of ACEI/ARBs for hypertensive COVID-19 patients. *Health Care Manag Sci*. 2021;24:339–355. doi: 10.1007/s10729-021-09545-5
38. Zheng J, Yu Z. A novel machine learning-based systolic blood pressure predicting model. *J Nanomaterials*. 2021;2021:9934998. doi: 10.1155/2021/9934998
39. AlKaabi LA, Ahmed LS, Al Attiyah MF, Abdel-Rahman ME. Predicting hypertension using machine learning: findings from Qatar biobank study. *PLoS ONE*. 2020;15:e0240370. doi: 10.1371/journal.pone.0240370
40. Chang W, Liu Y, Xiao Y, Yuan X, Xu X, Zhang S, Zhou S. A machine-learning-based prediction method for hypertension outcomes based on medical data. *Diagnostics (Basel)*. 2019;9:178. doi: 10.3390/diagnostics9040178
41. Elshawi R, Al-Mallah MH, Sakr S. On the interpretability of machine learning-based model for predicting hypertension. *BMC Med Inform Decis Mak*. 2019;19:146. doi: 10.1186/s12911-019-0874-0
42. Fang M, Chen Y, Xue R, Wang H, Chakraborty N, Su T, Dai Y. A hybrid machine learning approach for hypertension risk prediction. *Neural Comput Appl*. 2021. doi: 10.1007/s00521-021-06060-0
43. Islam MM, Rahman MJ, Chandra Roy D, Tawabunnahar M, Jahan R, Ahmed N, Maniruzzaman M. Machine learning algorithm for characterizing risks of hypertension, at an early stage in Bangladesh. *Diabetes Metab Syndr*. 2021;15:877–884. doi: 10.1016/j.dsx.2021.03.035
44. Kanegae H, Suzuki K, Fukatani K, Ito T, Harada N, Kario K. Highly precise risk prediction model for new-onset hypertension using artificial intelligence techniques. *J Clin Hypertens (Greenwich)*. 2020;22:445–450. doi: 10.1111/jch.13759
45. López-Martínez F, Núñez-Valdez ER, Crespo RG, García-Díaz V. An artificial neural network approach for predicting hypertension using NHANES data. *Sci Rep*. 2020;10:10620. doi: 10.1038/s41598-020-67640-z
46. Marin I, Goga N. Hypertension detection based on machine learning. *Proceedings of the 6th Conference on the Engineering of Computer Based Systems*. Association for Computing Machinery; 2019:Article 23.
47. Nour M, Polat K. Automatic classification of hypertension types based on personal features by machine learning algorithms. *Math Probl Eng*. 2020;2020:2742781. doi: 10.1155/2020/2742781
48. Xu F, Zhu J, Sun N, Wang L, Xie C, Tang Q, Mao X, Fu X, Brickell A, Hao Y, et al. Development and validation of prediction models for hypertension risks in rural Chinese populations. *J Glob Health*. 2019;9:020601. doi: 10.7189/jogh.09.020601
49. Diao X, Huo Y, Yan Z, Wang H, Yuan J, Wang Y, Cai J, Zhao W. An application of machine learning to etiological diagnosis of secondary hypertension: retrospective study using electronic medical records. *JMIR Med Inform*. 2021;9:e19739. doi: 10.2196/19739
50. Boutilier JJ, Chan TCY, Ranjan M, Deo S. Risk stratification for early detection of diabetes and hypertension in resource-limited settings: machine learning analysis. *J Med Internet Res*. 2021;23:e20123. doi: 10.2196/20123
51. Chunyu L, Ran L, Junteng Z, Miye W, Jing X, Lan S, Yixuan Z, Rui Z, Yizhou F, Chen W, et al. Characterizing the critical features when personalizing antihypertensive drugs using spectrum analysis and machine learning methods. *Artif Intell Med*. 2020;104:101841. doi: 10.1016/j.artmed.2020.101841
52. Angelaki E, Marketou ME, Barmparis GD, Patrianakos A, Vardas PE, Parthenakis F, Tsironis GP. Detection of abnormal left ventricular geometry in patients without cardiovascular disease through machine learning: an ECG-based approach. *J Clin Hypertens (Greenwich)*. 2021;23:935–945. doi: 10.1111/jch.14200
53. Gupta K, Balyan K, Lamba B, Puri M, Sengupta D, Kumar M. Ultrasound placental image texture analysis using artificial intelligence to predict hypertension in pregnancy. *J Matern Fetal Neonatal Med*. 2021;35:5587–5594. doi: 10.1080/14767058.2021.1887847
54. Koshimizu H, Kojima R, Kario K, Okuno Y. Prediction of blood pressure variability using deep neural networks. *Int J Med Inform*. 2020;136:104067. doi: 10.1016/j.ijmedinf.2019.104067
55. Esmaelpoor J, Moradi MH, Kadkhodamohammadi A. A multistage deep neural network model for blood pressure estimation using photoplethysmogram signals. *Comput Biol Med*. 2020;120:103719. doi: 10.1016/j.compbiomed.2020.103719
56. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK, Spirit AI; Group C-AW. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med*. 2020;26:1364–1374. doi: 10.1038/s41591-020-1034-x
57. Verhaar BJH, Collard D, Prodan A, Levels JHM, Zwinderman AH, Bäckhed F, Vogt L, Peters MJL, Muller M, Nieuwdorp M, et al. Associations between gut microbiota, faecal short-chain fatty acids, and blood pressure across ethnic groups: the HELIUS study. *Eur Heart J*. 2020;41:4259–4267. doi: 10.1093/eurheartj/ehaa704
58. Alghamdi AS, Polat K, Alghoson A, Alshdadi AA, Abd El-Latif AA. A novel blood pressure estimation method based on the classification of oscillometric waveforms using machine-learning methods. *Appl Acoust*. 2020;164:107279. doi: 10.1016/j.apacoust.2020.107279
59. Argha A, Celler BG. Blood pressure estimation from time-domain features of oscillometric waveforms using long short-term memory recurrent neural networks. *IEEE Trans Instrum Meas*. 2020;69:3614–3622. doi: 10.1109/TIM.2019.2941037
60. Argha A, Wu J, Su SW, Celler BG. Blood pressure estimation from beat-by-beat time-domain features of oscillometric waveforms using deep-neural-network classification models. *IEEE Access*. 2019;7:113427–113439. doi: 10.1109/ACCESS.2019.2933498
61. Celler BG, Le PN, Argha A, Ambikairajah E. GMM-HMM-based blood pressure estimation using time-domain features. *IEEE Trans Instrum Meas*. 2020;69:3631–3641. doi: 10.1109/TIM.2019.2937074
62. Magbool A, Bahloul MA, Ballal T, Al-Naffouri TY, Laleg-Kirati T-M. Aortic blood pressure estimation: a hybrid machine-learning and cross-relation approach. *Biomed Signal Process Control*. 2021;68:102762. doi: 10.1016/j.bspc.2021.102762

63. Singh AK, Sharma V, Kesarwani S. Exploration of monitoring and detection of blood pressure via machine learning. *Ann Romanian Soc Cell Biol*. 2021;25:1743–1748.
64. Pulido M, Melin P, Prado-Arechiga G. Blood pressure classification using the method of the modular neural networks. *Int J Hypertens*. 2019;2019:7320365. doi: 10.1155/2019/7320365
65. Chowdhury MH, Shuzan MNI, Chowdhury MEH, Mahbub ZB, Uddin MM, Khandakar A, Reaz MBI. Estimating blood pressure from the Photoplethysmogram signal and demographic features using machine learning techniques. *Sensors (Basel)*. 2020;20:20. doi: 10.3390/s20113127
66. Fujita D, Suzuki A, Ryu K. PPG-based systolic blood pressure estimation method using PLS and level-crossing feature. *Appl Sci*. 2019;9:304. doi: 10.3390/app9020304
67. Maher N, Elsheikh GA, Anis WR, Emara T. Enhancement of blood pressure estimation method via machine learning. *Alex Eng J*. 2021;60:5779–5796. doi: 10.1016/j.aej.2021.04.035
68. Mejía-Mejía E, May JM, Elgendi M, Kyriacou PA. Classification of blood pressure in critically ill patients using photoplethysmography and machine learning. *Comput Methods Programs Biomed*. 2021;208:106222. doi: 10.1016/j.cmpb.2021.106222
69. Chen S, Ji Z, Wu H, Xu Y. A non-invasive continuous blood pressure estimation approach based on machine learning. *Sensors (Basel)*. 2019;19:2585. doi: 10.3390/s19112585
70. Huttunen JMJ, Kärkkäinen L, Lindholm H. Pulse transit time estimation of aortic pulse wave velocity and blood pressure using machine learning and simulated training data. *PLoS Comput Biol*. 2019;15:e1007259. doi: 10.1371/journal.pcbi.1007259
71. Duan T, Rajpurkar P, Laird D, Ng AY, Basu S. Clinical value of predicting individual treatment effects for intensive blood pressure therapy. *Circ Cardiovasc Qual Outcomes*. 2019;12:e005010. doi: 10.1161/circoutcomes.118.005010
72. Tsoi KKF, Chan NB, Yiu KKL, Poon SKS, Lin B, Ho K. Machine learning clustering for blood pressure variability applied to systolic blood pressure intervention trial (SPRINT) and the Hong Kong community cohort. *Hypertension*. 2020;76:569–576. doi: 10.1161/hypertensionaha.119.14213
73. Ankışhan H. Blood pressure prediction from speech recordings. *Biomed Signal Process Control*. 2020;58:101842. doi: 10.1016/j.bspc.2019.101842
74. Chiang PH, Dey S. Offline and online learning techniques for personalized blood pressure prediction and health behavior recommendations. *IEEE Access*. 2019;7:130854–130864. doi: 10.1109/ACCESS.2019.2939218
75. El Attaoui A, Largo S, Jilbab A, Bourouhou A. Wireless medical sensor network for blood pressure monitoring based on machine learning for real-time data classification. *J Ambient Intell Humaniz Comput*. 2021;12:8777–8792. doi: 10.1007/s12652-020-02660-1
76. Huang KH, Tan F, Wang TD, Yang YJ. A highly sensitive pressure-sensing array for blood pressure estimation assisted by machine-learning techniques. *Sensors (Basel)*. 2019;19:19. doi: 10.3390/s19040848
77. Guthrie NL, Carpenter J, Edwards KL, Appelbaum KJ, Dey S, Eisenberg DM, Katz DL, Berman MA. Emergence of digital biomarkers to predict and modify treatment efficacy: machine learning study. *BMJ Open*. 2019;9:e030710. doi: 10.1136/bmjopen-2019-030710
78. Zhang L, Hurley NC, Ibrahim B, Spatz E, Krumholz HM, Jafari R, Mortazavi BJ. Developing personalized models of blood pressure estimation from wearable sensors data using minimally-trained domain adversarial neural networks. *Proc Mach Learn Res*. 2020;126:97–120.

Supplemental Material

Table S1: Results from the application of the HUMANE checklist to articles included in analysis. Each question shows the total number of papers that were scored for each choice. The responses are adjudicated responses, where the two main authors CDT and TQBT resolved any discordant responses to a single response.

Question		Options		Responses	Response (%)
Clinical Relevance					
Q1	**Is the importance of research (e.g., cost/life/time/process savings) explained?**	Yes		53	84%
		No		10	16%
Q2	**Which of the following domain(s) did the article explore for potential impact of the model? (check all that apply)**	Triage	Checked	4	6%
			Unchecked	59	94%
		Early Diagnosis	Checked	23	37%
			Unchecked	40	63%
		Improved Diagnosis	Checked	29	46%
			Unchecked	34	54%
		Allowed personalized/targeted treatment	Checked	6	10%
			Unchecked	57	90%
		Prevent/reduce hospital admissions	Checked	5	8%
			Unchecked	58	92%
		Improve survival	Checked	6	10%
			Unchecked	57	90%
		Other	Checked	22	35%
			Unchecked	41	65%
Q3	**Is the intended role of the model (e.g., triage or diagnosis) clear?**	Yes		41	65%
		No		6	10%
		NA		16	25%
Q4	**Is it clear whether the model be used as an isolated test or in combination with other diagnostic elements?**	Yes		39	62%
		No		11	17%
		NA		13	21%
Defining and Addressing the Knowledge Gap					
Q1	**Have the authors detailed what is already known in the field?**	Yes		62	98%
		No		1	2%
Q2	**Is the knowledge gap defined?**	Yes		56	89%
		No		7	11%
Q3	**Have the authors explained how they aim to address the knowledge gap?**	Yes		59	94%
		No		4	6%
Pre-specified Study Design					
Q1	**Is the experimental protocol designed to prevent overfitting?**	Yes		37	59%
		No		22	35%
		NA		4	6%
Q2		Yes		27	43%

	Are there pre-defined inclusion and exclusion criteria for different model/study datasets?	No		27	43%
		NA		9	14%
Q3	**Does the outcome tested by the ML model align with written methods?**	Yes		56	89%
		No		6	10%
		NA		1	2%
Q4	**Has the study described any other multivariable prediction models?**	Yes		38	60%
		No		20	32%
		NA		5	8%
Q5	**Has the study pre-specified a statistical analysis plan?**	Yes		44	70%
		No		19	30%
Q6	**Has the study applied any of the following methods to address class imbalance?**	Oversampling - adding copies of underrepresented class		4	6%
		Undersampling - removing copies of overrepresented class		3	5%
		Replicate the class distribution in the validation test set		3	5%
		Other			
		None Reported		53	84%
Data Suitability					
Q1	**Is the study methodology and study pre-specified in terms of the study design (e.g., retrospective/prospective, derivation/validation, supervised/unsupervised/deep learning), including characteristics of the data type collected?**	Yes		55	87%
		No		8	13%
Q2	**Is the study timeline specified in terms of initiation of data collection/model development and the end date of the completed (or ongoing) data collection/model validation?**	Yes		26	41%
		No		37	59%
Q3	**Is the dataset obtained from within the intended stage in the care pathway?**	Yes		30	48%
		No		5	8%
		Unclear		28	44%
Q4	**Are the key data pre-processing/pre-curation steps described?**	Yes		45	71%
		No		18	29%
Q5	**Is the dataset appropriate for the healthcare conditions studied?**	Yes		59	94%
		No		4	6%
Q6	**Is there sufficient clarity on dataset for model development (training/test/validation)?**	Clear		24	38%
		Partially Clear		25	40%
		Unclear		14	22%
ELSI					
Q1	**Is it explicitly mentioned that study is compliant with local ethical**	Yes		30	48%
		No		27	43%

	committee/IRB/patient privacy/data security regulations?	NA		6	10%
Q2	**Has documented consent been obtained from the participants involved in the prospective/intervention study?**	Yes		17	27%
		No		22	35%
		NA		24	38%
Q3	**Has the article evaluated algorithmic bias? (e.g., gender, race, ethnicity, socioeconomic status etc.)**	Yes		1	2%
		No		57	90%
		Partial		5	8%
Q4	**Have the authors listed their conflict of interest(s)?**	Yes		51	81%
		No		12	19%
Ground Truth					
Q1	**Is ground truth applicable for supervised learning method in this article?**	Yes		58	92%
		No		5	8%
Q2	**How much do you agree with the accuracy of the ground truth labels (is labelling backed by clinical guidelines or references; are sufficient details provided on the ground truth labelling process)?**	Strongly Agree		22	38%[a]
		Agree		25	43%[a]
		Neutral		10	17%[a]
		Disagree		0	0%[a]
		Strongly Disagree		1	2%[a]
Q3	**Were ground truth labels manually determined by experts?**	Yes		32	55%[a]
		No		26	45%[a]
Q4	**Were ground truth labels automatically generated?**	Yes		7	12%[a]
		No		51	88%[a]
Q5	**Were any ground truth labels missing?**	Yes		0	0%[a]
		No		58	100%[a]
Q6	**How were the ground truth labels added?**	Prospectively		47	81%[a]
		Retrospectively		11	19%[a]
Q7	**Which of the following is applicable for the number of experts involved in the review?**	Single		54	93%[a]
		Multiple Independent		4	7%[a]
		Use of Adjudicator(s)		0	0%[a]
Q8	**Which of the following is applicable regarding the qualification of the expert(s) in the review?**	Sub-specialist with experience		4	7%[a]
		Board-certified specialist		1	2%[a]
		Specialist in the domain without sub-specialty accreditation		0	0%[a]
		Others		53	91%[a]
Q9	**Was there sufficient availability of clinical information to the expert to make the diagnosis?**	Yes		48	83%[a]
		No		0	0%[a]
		Unclear		10	17%[a]
Q10	**Is an inter-observer agreement presented?**	Yes		0	0%[a]
		No		4	7%[a]
		NA		54	93%[a]
Performance Metrics					
Q1		Yes		22	35%

	Was the distribution of outcomes similar in all training, test and validation datasets?	No		6	10%
		NA		35	56%
Q2	**Has the study specified a range of statistical measures used to compare the accuracy/precision/sensitivity/specificity of the proposed model?**	Yes		44	70%
		No		19	30%
Q3	**Has the article presented any difference between the training, testing, and validation data sets in inclusion criteria, model outcome, and predictors?**	Yes		8	13%
		No		39	62%
		NA		16	25%
Q4	**Has the study reported any discrimination measures of performance? (Check all that apply)**	Accuracy	Checked	32	51%
			Unchecked	31	49%
		Sensitivity/Recall	Checked	20	32%
			Unchecked	43	68%
		Specificity	Checked	12	19%
			Unchecked	51	81%
		Precision	Checked	13	21%
			Unchecked	50	79%
		ROC curve	Checked	20	32%
			Unchecked	43	68%
		Precision recall (PR) curve	Checked	3	5%
			Unchecked	60	95%
		Other	Checked	26	41%
			Unchecked	37	59%
		None reported	Checked	12	19%
			Unchecked	51	81%
Q5	**Has the article reported any calibration measures of performance? (Check all that apply)**	Calibration plot	Checked	4	6%
			Unchecked	59	94%
		Hosmer-Lemeshaw test	Checked	1	2%
			Unchecked	62	98%
		Excepted calibration error	Checked	0	0%
			Unchecked	63	100%
		Brier score	Checked	2	3%
			Unchecked	61	97%
		Mean square error (MSE)	Checked	11	17%
			Unchecked	52	83%
		Other	Checked	12	19%
			Unchecked	51	81%
		None reported	Checked	38	60%
			Unchecked	25	40%
Replication and Validation					
Q1	**Is the validation dataset distinct from training and test datasets?**	Temporally		3	5%
		Geographically		2	3%
		Both		5	8%
		None		53	84%

Q2	**Has the study described the predictor model using an internal validation technique?**	Yes		46	73%
		No		11	17%
		NA		6	10%
Q3	**How was the experimental protocol developed to prevent overfitting?**	Independent train and test dataset validation		5	8%
		Crossfold validation		29	46%
		Leave one out validation		3	5%
		Other		0	0%
		Not Applicable (NA)		26	41%
Q4	**Was model validation performed using an out-of-sample external validation dataset?**	Yes		9	14%
		No		54	86%
Q5	**What other steps are reported to support external validity?**	Disease prevalence in the internal validation test dataset representative of the target population in the real world		9	14%
		Presence of subgroups within the training dataset		5	8%
		Authors have not applied any inclusion or exclusion criteria which create a selection bias		28	44%
		Authors have applied a sampling method (i.e. random sampling) to reduce the risk of spectrum bias?		8	13%
		Other		13	21%
Traditional components of scientific papers					
Q1	**Is the title relevant to research in the field of AI/ML in medicine?**	Yes		52	83%
		No		11	17%
Q2	**Does the title align with any of the following terms or related terms: AI, ML, or deep learning?**	Yes		55	87%
		No		8	13%
Q3	**Does the abstract provide a summary of the following: objectives, study design, setting, target population, statistical analysis, results, and conclusion pertinent to ML in healthcare?**	Agree		29	46%
		Partially Agree		26	41%
		Disagree		8	13%
Q4	**Has the article defined the objectives including validation or development of ML?**	Yes		53	84%
		No		10	16%
Q5	**Is there a pre-specified threshold for inclusion of cases where there is non-consensus?**	Yes		1	2%
		No		20	32%
		NA		42	66%

Q6	**Has the study described key demographics/characteristics of the cohorts? (Table 1- age, gender, chronic co-morbidities, patient type etc.)**	Yes		29	46%
		No		34	54%
Q7	**Has the study described either in text or by a flow diagram the impact of applying stated inclusion/exclusion criteria on the final sample size?**	Yes		11	17%
		No		52	83%
Q8	**Has the study provided a succinct summary of their primary result findings?**	Yes		60	95%
		No		3	5%
Q9	**Has the study compared their results with existing literature, by supporting or challenging their findings?**	Yes		52	83%
		No		11	17%
Q10	**Has the article mentioned strengths of their research?**	Yes		53	84%
		No		10	16%
Q11	**Has the article mentioned weaknesses of their research?**	Yes		48	76%
		No		15	24%
Q12	**Have the authors provided a justifiable conclusion based on the results presented with a take-home message and implications of the results?**	Yes		59	94%
		No		4	6%

[a] These percentages are out of 58, the number of 'Yes' responses to Ground Truth Q1.

Table S2: Articles included in analysis. 4D MRI: 4-dimensional magnetic resonance imaging; ANN: Artificial Neural Network; BiLSTM: Bidirectional LSTM; BP: blood pressure; CART: Classification And Regression Trees; CNN: Convolutional Neural Network; DANN: Domain-Adversarial Training of Neural Networks; DBN: Deep Belief Network; DNN: Deep Neural Network; ECG: electrocardiogram; GNN: Graph Neural Network GPR: Gaussian process regression; HTN: hypertension; KNN: k-Nearest Neighbors; LASSO: Least Absolute Shrinkage and Selection Operator; LDA: Linear Discriminant Analysis; LightGBM: Light Gradient Boosting Machine; LSTM: Long Short-Term Memory networks; LSVM: Lagrangian Support Vector Machine; ML: machine learning; MLP: Multilayer perceptron; MNN: Modular Neural Network; NBC: Naive Bayes Classifier; PPG: photoplethysmography; RCT: Randomised Controlled Trial; RF: Random Forest; RFE: Recursive Feature Elimination; RL: Reinforcement Learning; RNN: Recurrent Neural Network; SOM: Self-Organizing Map; SVM: Support Vector Machines; SVR: Support Vector Regression.

Publication	**Data source**	**ML task**	**ML methods and study objectives**	**Ref.**
Aziz et al. 2020	Adherence questionnaire, demographics, medical records	Drug adherence	Use ML (RF ANN, SVR, SOM) to find determinants of antihypertensive medication adherence & predict precise adherence scores.	16
Argha et al. 2019	Auscultatory waveforms	Predict BP	Use DL (LSTM-RNN) to estimate SBP & DBP from auscultatory waveforms.	17
Argha et al. 2021	Auscultatory waveforms	Predict BP	Use DL (BiLSTM-RNN) to estimate SBP & DBP from auscultatory waveforms.	18
Pan et al. 2019	Auscultatory waveforms	Predict BP	Use ML (CNN) to determine BP from Korotkoff sound recordings.	19
Pan et al. 2019	Auscultatory waveforms	Predict BP	Use ML (CNN) to determine impact of movement disturbance on BP measurement.	20
Persell et al. 2020	Medical records (clinical trial)	HTN management	AI based coaching app for HTN management.	21
Miao et al. 2020	ECG	Predict BP	Use ML (CNN with LSTM) to estimate BP from ECG data.	22
Soh et al. 2020	ECG	Predict BP	Use ML (k-NN, decision tree, LDA) to identify masked HTN from ECG data without ABPM.	23
Li et al. 2020	ECG & PPG	Predict BP	Use ML (LSTM) to estimate BP from PPG &E CG signals in real time.	24
Yan et al. 2019	ECG & PPG	Predict BP	Use ML (CNN) to estimate BP from PPG & ECG signals in real time.	25

Zhang et al. 2019	ECG & PPG	Predict BP	Use ML (SVR) to estimate BP PPG & ECG signals & other physiological measurements.	26
Sannino et al. 2020	ECG & PPG	Predict HTN	Comparison of discriminative performance of several ML models (in classifying HTN from PPG & ECG data).	27
Li et al. 2019	Genetic data	Predict HTN	Use ML (SVM) to predict HTN from genetic & environmental risk factors.	28
Widen et al. 2021	Genetic data & medical data	Predict BP	Use ML (LASSO) to predict quantitative traits from genomic data	29
Kissas et al. 2020	Imaging, computational fluid dynamics, 4D MRI	Predict BP	Use physics informed neural networks to predict BP from 4D flow MRI	30
Lacson et al. 2019	Medical records	BP variability	Use ML (random forest) to identify features affecting SBP variability.	31
Barbieri et al. 2019	Medical records	BP, fluid management and dialysis	Use ML (ANN) to guide BP, fluid volume & dialysis dose in ESKD	32
Cho et al. 2020	Medical records	CVD/ outcomes	Use DL (RNN-LSTM) & Cox regression to predict CVD.	33
Du et al. 2020	Medical records	CVD/ outcomes	Use ML (XGBoost, kNN, SVM, decision tree, random forest) & logistic regression to predict CHD risk factors.	34
Wu et al. 2019	Medical records	CVD/ outcomes	Use ML (ANN) to predict NSTEMI.	35
Wu et al. 2020	Medical records	CVD/ outcomes	Use ML (XGBoost) to predict outcomes of young patients with HTN.	36
Bertsimas et al. 2021	Medical records	Personalised treatment	Use ML (ensemble of multiple methods) to personalise ACEI/ARB treatment for hypertensive COVID-19 patients.	37
Zheng et al. 2021	Medical records	Predict BP	Use ML (SVM, decision tree, GPR, ANN, logistic regression) to predict SBP from clinical features.	38
AlKaabi et al. 2020	Medical records	Predict HTN	Use supervised ML models (decision tree, random forest, logistic regression) to predict hypertension from 987 biobank records.	39
Chang et al. 2019	Medical records	Predict HTN	Use ML (SVM, decision tree, random forest, XGBoost) to predict HTN from clinical data.	40
Elshawi et al. 2019	Medical records	Predict HTN	Use ML (random forest) to predict hypertension risk from fitness data & evaluate interpretability.	41

Fang et al. 2021	Medical records	Predict HTN	Use ML (k-NN, LightGBM, SVM, random forest) to predict 5-year HTN risk from medical records.	42
Islam et al. 2021	Medical records	Predict HTN	Use ML (ANN, decision tree, random forest, gradient boosting) to characterise HTN risks (features identified with LASSO & SVM RFE).	43
Kanegae et al. 2020	Medical records	Predict HTN	Use ML (XGBoost & ensemble model) for hypertension risk prediction.	44
López-Martínez et al. 2020	Medical records	Predict HTN	Use ML (ANN) to predict HTN from demographic & clinical features.	45
Marin et al. 2019	Medical records	Predict HTN	Use ML (random forest, SVM, Gaussian Naïve Bayes, logistic regression) to classify hypertension from medical data.	46
Nour et al. 2020	Medical records	Predict HTN	Use ML (random forest, decision tree, LDA, LSVM) to classify hypertension from medical data.	47
Xu et al. 2019	Medical records	Predict HTN	Use ML (ANN, NBC, CART) to predict HTN risk (development & validation of population-specific HTN risk prediction model).	48
Diao et al. 2021	Medical records	Predict secondary HTN	Use ML (XGBoost) to predict aetiology of secondary HTN.	49
Boutilier et al. 2021	Medical records	Risk stratification	Use ML (decision tree, random forest, RL, k-NN, AdaBoost) for risk stratification of HTN & diabetes in resource-limited LMICs.	50
Chunyu et al. 2020	Medical records	Treatment effects	Use ML (LASSO, mean decrease impurity, recursive feature elimination, ensemble models) to find features contributing to treatment response to 5 commonly prescribed anti-HTN drugs.	51
Angelaki et al. 2021	Medical records & ECG	Predict LVH	Use supervised ML (random forest) to detect abnormal LVG before onset of LVH from ECG & basic clinical parameters from 528 normotensive & hypertensive patients.	52
Gupta et al. 2021	Medical records & imaging	Predict HTN in pregnancy	Use ML (CNN) to predict HTN from placental ultrasound images in pregnancy.	53
Koshimizu et al. 2020	Medical records (clinical trial)	BP variability	Use ML (DNN) to predict BP variability from PREDICT trial data.	54
Esmaelpoor et al. 2020	Medical records, PPG	Predict BP	Use DL (DNN) to estimate BP from PPG.	55
Liu et al. 2020	Nutritional data	Predict HTN	Use ML (SVM, decision tree, random forest, MLP, XGBoost) to predict HTN from nutritional intake.	56

Verhaar et al. 2020	Nutritional, microbiome data	Predict BP	Use ML (XGBoost) to investigate association of microbiome & BP.	57
Alghamdi et al. 2020	Oscillometric waveforms	Predict BP	Use supervised ML models (kNN, WkNN, bagged trees) to predict SBP & DBP from oscillometric waveforms from 350 patients.	58
Argha et al. 2020	Oscillometric waveforms	Predict BP	Use DL (LSTM-RNN) to estimate SBP & DBP from oscillometric waveforms.	59
Argha et al. 2019	Oscillometric waveforms	Predict BP	Use DL (DBN-DNN) to estimate SBP & DBP from oscillometric waveforms.	60
Celler et al. 2020	Oscillometric waveforms	Predict BP	Use ML (GMM-HMM) to estimate SBP & DBP from oscillometric waveforms.	61
Magbool et al. 2021	Other (simulated data)	Aortic BP	Use ML (decision tree, random forest, MLR, neural networks) to estimate aortic BP from simulated pulse wave dataset.	62
Singh et al. 2021	Other (unclear)	HTN, ABPM	Use ML (random forest) to predict HTN from clinical features	63
Pulido et al. 2019	Other (unclear)	Predict HTN	Use ML (MNN) to classify HTN from BP data.	64
Chowdhury et al. 2020	PPG	Predict BP	Use ML (SVR, GPR, regression trees, ensemble trees) & linear regression to determine BP from PPG.	65
Fujita et al. 2019	PPG	Predict BP	Use partial least-squares regression to estimate BP from PPG.	66
Maher et al. 2021	PPG	Predict BP	Use ML (SVM, ANN) to estimate BP from PPG.	67
Mejía-Mejía et al. 2021	PPG	Predict BP	Use ML (k-NN, SVM, ANN) to classify HTN and predict BP from PPG.	68
Chen et al. 2019	Pulse transit time	Realtime BP	Use ML (SVR) to continuously monitor BP from pulse transit time measurements.	69
Huttunen et al. 2019	Pulse transit time, simulated data	BP, aortic BP	Train ML model (Gaussian process regression) on simulated patient data for BP prediction from PTT.	70
Duan et al. 2019	Medical records (clinical trial)	Treatment effects	Use ML (X-learner) & logistic regression to predict treatment effect size of intensive & standard anti-HTN therapy.	71
Tsoi et al. 2020	Medical records (clinical trial)	BP variability	Use ML (K-means clustering, Partitioning Around Medoids, spectral clustering, Ward's method, Expectation Maximization) to cluster BP variability into groups.	72

Ankışhan et al. 2020	Speech recordings	Predict BP	Use ML (CNN, SVM/SVR, MLR) to predict BP from speech recordings from 86 subjects.	73
Chiang et al. 2019	Wearable technology	Personalised treatment	Use ML (random forest) to predict BP from wearable tech data & historical BP readings.	74
El Attaoui et al. 2021	Wearable technology	Realtime BP	Present a wireless medical sensor network with wireless BP sensing and ML (decision tree, kNN, NBC) to monitor BP in real time (for both patients & physicians).	75
Huang et al. 2019	Wearable technology	Realtime BP	ML (random forest, gradient boosting, adaptive boosting regression models) with wearable pulse wave sensor	76
Guthrie et al. 2019	Wearable technology	Treatment effects	Use ML (random forest) to develop digital biomarkers for digital therapeutic treatment response.	77
Zhang et al. 2020	Wearable technology, bioimpedance	Predict BP	Use ML (DANN) to estimate beat-to-beat BP from 5mins of bioimpedance data.	78