# Potential applications and performance of machine learning techniques and algorithms in clinical practice: A systematic review

Ezekwesiri Michael Nwanosike[1], Barbara R Conway[1], Hamid A Merchant[1], Syed Shahzad Hasan[1,2]

[1] Department of Pharmacy, School of Applied Sciences, University of Huddersfield, Queensgate Huddersfield HD1 3DH, West Yorkshire, United Kingdom

[2]School of Biomedical Sciences & Pharmacy, University of Newcastle, Callaghan, Australia

**Corresponding Author:** Syed Shahzad Hasan, PhD, Department of Pharmacy, School of Applied Sciences, University of Huddersfield, Queensgate, HD1 3DH, Huddersfield, West Yorkshire, United Kingdom; s.hasan@hud.ac.uk

**Keywords**

Machine learning, clinical studies, electronic health records (EHRs), clinical practice, model deployment, AUROC, prediction, classification

**Abstract** = 261 words

**Main Text** (introduction to conclusion, including figures & summary point) = 4000 words

## Abstract

### Purpose

The advent of clinically adapted machine learning algorithms can solve numerous problems ranging from disease diagnosis and prognosis to therapy recommendations. This systematic review examines the performance of machine learning (ML) algorithms and evaluates the progress made to date towards their implementation in clinical practice.

### Methods

Systematic searching of databases (PubMed, MEDLINE, Scopus, Google Scholar, Cochrane Library and WHO Covid-19 database) to identify original articles published between January 2011 and October 2021. Studies reporting ML techniques in clinical practice involving humans and ML algorithms with a performance metric were considered.

### Results

Of 873 unique articles identified, 36 studies were eligible for inclusion. The XGBoost (extreme gradient boosting) algorithm showed the highest potential for clinical applications (n=7 studies); this was followed jointly by random forest algorithm, logistic regression, and the support vector machine, respectively (n=5 studies). Prediction of outcomes (n=33), in particular Inflammatory diseases (n=7) received the most attention followed by cancer and neuropsychiatric disorders (n = 5 for each) and Covid-19 (n=4). Thirty-three out of the thirty-six included studies passed more than 50% of the selected quality assessment criteria in the TRIPOD checklist. In contrast, none of the studies could achieve an ideal overall bias rating of 'low' based on the PROBAST checklist. In contrast, only three studies showed evidence of the deployment of ML algorithm(s) in clinical practice.

### Conclusions

ML is potentially a reliable tool for clinical decision support. Although advocated widely in clinical practice, work is still in progress to validate clinically adapted ML algorithms. Improving quality standards, transparency, and interpretability of ML models will further lower the barriers to acceptability.

### Keywords

Machine learning; clinical studies; electronic health records (EHRs); clinical practice; model deployment; AUROC; prediction; COVID-19.

## 1. Introduction

Since the Food and Drug Administration (FDA) approval of machine learning (ML) algorithms for decision making without explicit programming, the push for ML applications in clinical practice has grown rapidly [1]. Today, data are generated at a high velocity and in massive volumes, making it impossible for manual handling and analysis; hence the term 'big data' is often used to describe it. Big data can be obtained from various sources within the scope of clinical practice, for instance, Electronic Health Records (EHR), diagnostic imaging, etc. [2,3]. In addition, there is also an abundance of wearables technologies such as sensors, mobile phone applications, and medical devices that constantly capture and transmit the data to the relevant databases [4,5,6].

Clinical practice has greatly benefited from the boom in machine learning applications [7]. Current clinical decision support tools (with built-in machine learning algorithms) have shown remarkable potential depending on the algorithms selected: disease classification or profiling (diagnosis); risk stratification (prediction); dose recommendation (estimation); pattern recognition; precision treatment, and health monitoring, etc. [8].

Machine learning relies on the science of artificial intelligence - a hybrid of computer science and statistics. Numerous ML algorithms (supervised and unsupervised learning algorithms) have been devised for clinical applications: Random forests, Support vector machines, Naïve Bayesian classifiers, Neural Networks, Decision Trees, ensemble algorithms, and various other proprietary algorithms [9,10,11]. The workflow for the clinical application of a typical ML algorithm is described by Radakovich et al. [12]. First, input data (e.g., from electronic health records) must be pre-processed [4]. This depends on the model used and the research question posed. It usually involves selecting relevant features (e.g., body mass index - BMI). This is often done empirically or handled by an algorithm itself. Pre-processing could also be done iteratively, making changes based on the predictor's effect on model performance.

After pre-processing, a suitable algorithm is selected and trained (relevant data split into training and testing sets). In essence, the model is generated with the training data and then used to predict outcomes associated with test data (which represents held out data). This process is iteratively repeated to improve model generalizability (multi-fold cross-validation). In most cases, some data is reserved (unseen by model) to avoid overfitting variables [13]. For clinical practice, model building is incomplete without a clear explanation of how the predictions are made (transparency), estimation of uncertainty associated with the prediction; and external validation results from clinical studies (ideally prospective randomised control trials). It is important to note that external validation better guarantees generalisability than internal validation of an ML model [14,15]. By and large, the available ML algorithms have demonstrated remarkable clinical utility but are not yet established in clinical practice to replace conventional CDSS tools.

Therefore, this systematic review (SR) critically appraises the strengths and weaknesses of various ML strategies and algorithms being proposed for clinical applications and their ability to transition from early feasibility studies to real-life clinical practice, particularly its application in disease management. In addition, the review seeks to identify key areas in the pipeline (value chain) that need further research and improvement to implement ML in clinical practice successfully.

## 2. Methods

### 2.1. Scope of review: eligibility criteria

This systematic review was completed in line with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines and comprised original research with clear outcomes and design (retrospective cohort study, multisite or longitudinal studies, randomised clinical trial), studies examining the potential or actual deployment of ML model(s), and studies whose data were extracted from electronic health records or clinic registries. Studies related to radiomics, out-of-scope artificial intelligence (including unsupervised learning), medical devices, surgery, secondary research articles (reviews), grey literature, etc., were excluded from this review. The primary investigator (EMN) applied the eligibility criteria to examine original journal articles published in English that met the selection criteria, (a) machine learning tools, and (b) possible clinical applications were included. The search was restricted to studies involving humans and original research articles.

### 2.2. Information sources

The databases searched were PubMed (United States National Library of Medicine), MEDLINE (Medical Literature Analysis and Retrieval System Online, or MEDLARS Online), Scopus, Google Scholar, Cochrane Library databases, and WHO Covid-19 database, from January 2011 until October 2021. Secondary references made up the remaining articles gathered. The databases were searched between August 2021 and November 2021. The reference list of included articles and review papers was also searched to identify relevant articles. The articles identified and selected through this method were subjected to the same eligibility evaluation.

### 2.3. Searching

The search strategy identified original research on clinical outcomes associated with machine learning tools. An extensive literature search was carried out using different search term combinations. The search terms used were: machine learning tools, machine learning applications, machine learning skills, clinical practice, clinical research, clinical decision making, clinical decision support, risk prediction, model development, model deployment, prediction model, pharmacy, patient care, treatment, and healthcare (refer **supplementary materials, Table S1**). The search period was limited to the last ten years (2011 – 2021). Only articles written in English were included; the study subjects were humans (male and female) of all ages.

### 2.4. Data extraction

With the aid of a systematic review software (*Rayyan*), the articles collected were initially screened (based on title and abstract) by two independent reviewers (EMN, SSH). Consensus resolved disagreements, and a third person was called upon when necessary. The information from all the retrieved studies was collected after an in-depth reading and extracted using a table developed by the lead investigator (EMN) and verified by the reviewer (SSH). Extracted information from studies is summarised in **Table 1**. The data extracted included authors, year, country, study design and population, machine learning algorithm or best performing model, potential applications, evidence of clinical deployment, external validation, access to codes, and outcomes**.**

### 2.5. Outcome measures

For the studies included for the systematic review, the outcome measures used to gauge the performance of the algorithms were Area Under the Receiver Operating Characteristic (AUROC) curve, accuracy, specificity, and sensitivity.

### 2.6. Quality assessment

The TRIPOD (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis) checklist for prediction model development was used to assess the quality of the studies included for systematic review [16]. The TRIPOD statement outlines the criteria for reporting studies that involve machine learning algorithm(s) (multivariate models). Notably, the quality assessment criteria used to score the articles included in the systematic review were adapted from the 20-item revised TRIPOD checklist (4a-16 with 5c and eight omitted) used by Wang et al. [17]. Alongside the TRIPOD checklist, the prediction model risk of bias assessment tool (PROBAST) [18] was also used to judge the quality (**Table S2**). The signalling questions, domains, overall ratings of bias, and applicability used as criteria in the tool were adapted from Li et al. [19].

## 3. Results

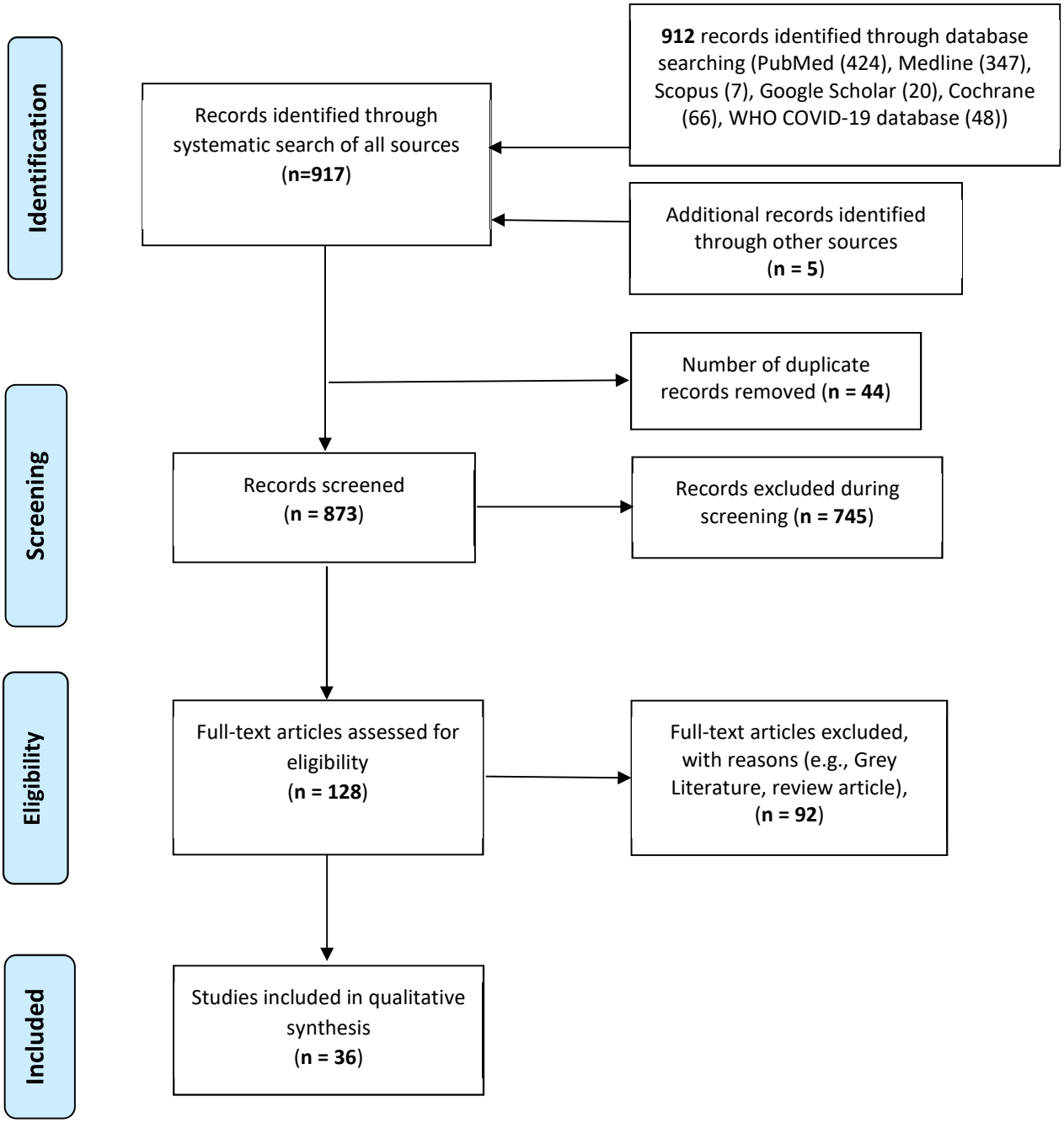


**Figure 1:** PRISMA flow diagram describing the screening process for articles collected.

### 3.1. Study Selection

The unique articles identified were 873; 745 articles were excluded after screening (**see supplementary Table S1 for details**). Based on eligibility criteria, 128 articles remained for assessment. All the studies used structured or semi-structured data, except two (2) studies that relied on imaging data. Finally, 36 articles were selected for the systematic review (**Figure 1**). It is important to note that quantitative analysis (e.g., meta-analysis) could not be carried out due to significant inconsistent performance metrics reported across the studies.

### 3.2. Quality assessment

The quality assessment was completed using a 20-item revised TRIPOD checklist and PROBAST tool. Based on the TRIPOD checklist, two studies, Bazila & Ponniah [42], and Veeranki et al. [34], received the lowest score of 40%. In contrast, three studies, Deshmukh & Merchant [48], Waljee et al. [29], and Shouval et al. [24] received the highest percentage score of 85%. None of the studies met the full criteria (**Figure 2**). Based on the PROBAST tool, none of the studies attained the ideal overall (concerns for) bias rating of low, nor (concerns for) applicability rating of low (**Table S2**). The studies had ratings ranging from moderate to high: Turki and Wang [30], Diao et al. [31], Veeranki et al. [34], Bora et al. [46], Xiao et al. [41], Brinker et al. [45], Kouchaki et al. [39], and Bazila and Ponniah et al. [42] (7 studies) showed serious concern for bias rating; Turki and Wang [30], Veeranki et al. [34], Kouchaki et al. [39], Bazila and Ponniah et al. [42], Brinker et al. [45] and Bora et al. [46] (6 studies showed serious overall concern for applicability rating).

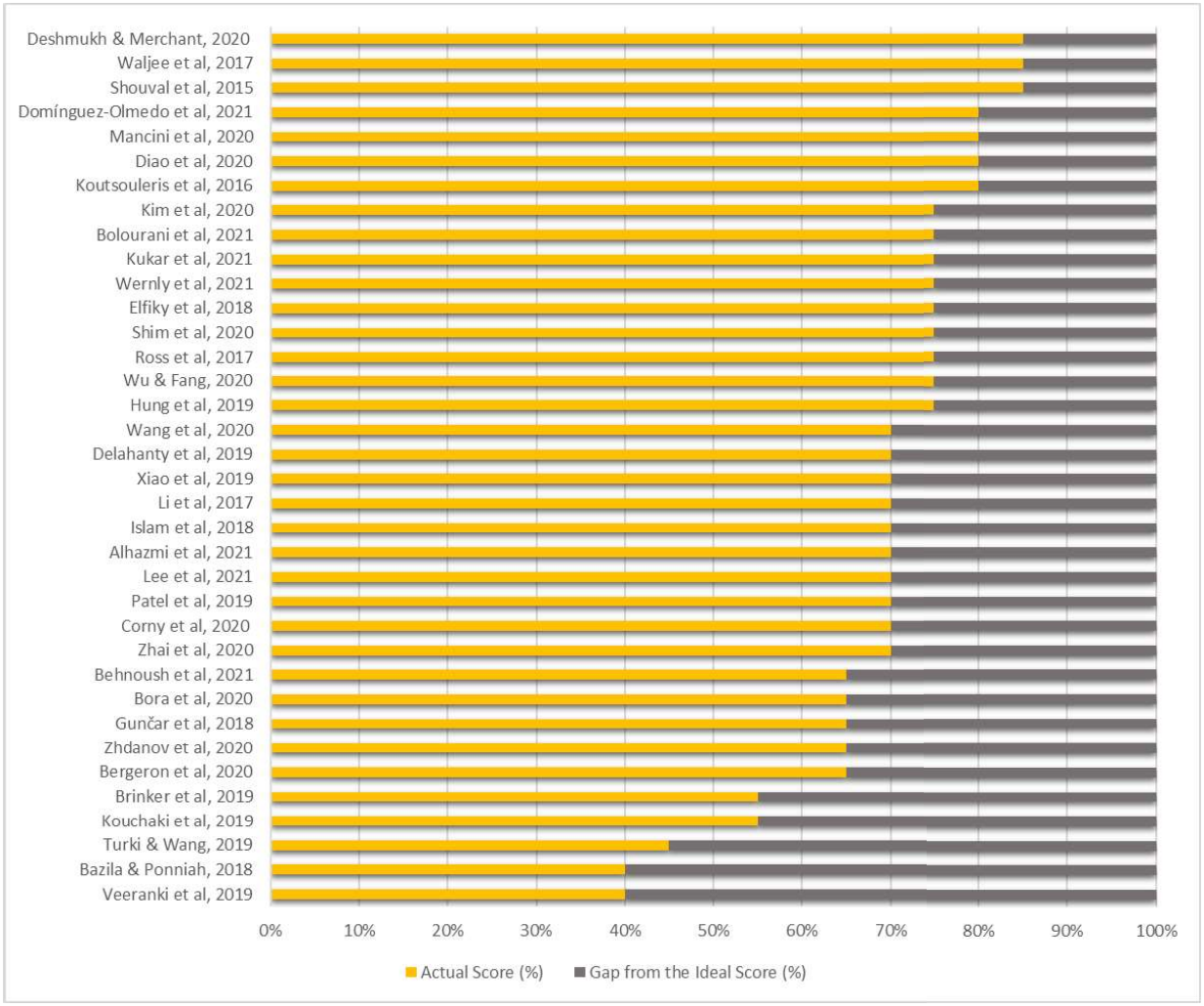


**Figure 2:** Quality assessment of included studies using TRIPOD checklist.

Three studies met 85% of the selected criteria in the TRIPOD statement. These are considered to have the highest quality. It is important to note that although all publications from 2011-2021 were included in the study and initially screened, the articles published during 2015-2021 (post-TRIPOD) met the inclusion criteria and therefore assessed against the TRIPOD checklist that was established in 2015. It was, therefore, not possible to compare the pre-TRIPOD and post-TRIPOD studies. It is also worth noting

that the TRIPOD checklist was established during an era when exploratory research on potential clinical applications of machine learning was gaining momentum. For this reason, the majority of the PRE-2015 publications referenced in this systematic review were review articles and included only a few original research studies.

### 3.3. Study characteristics

Most of the studies included in the review were conducted in the USA (39%), followed by China (19%). Asia and North America were the predominant regions where the majority of studies were carried out (44% Asia, 42% USA), followed by Europe (36%) and only one study from Australia. Of the publications included in the SR, 7 (19%) were prospective study designs, whereas 29 (81%) were retrospective.

Meanwhile, **Figure 3** shows the yearly breakdown of the number of articles on the review topic collected by the authors prior to the articles' screening process. Within the ten years under review, 85% of the articles identified were published from 2017 to 2021. There was a significant increase in the number of publications related to clinical applications of ML from 2017. The sharpest growth was observed after 2017 showing increased enthusiasm for machine learning in clinical practice.

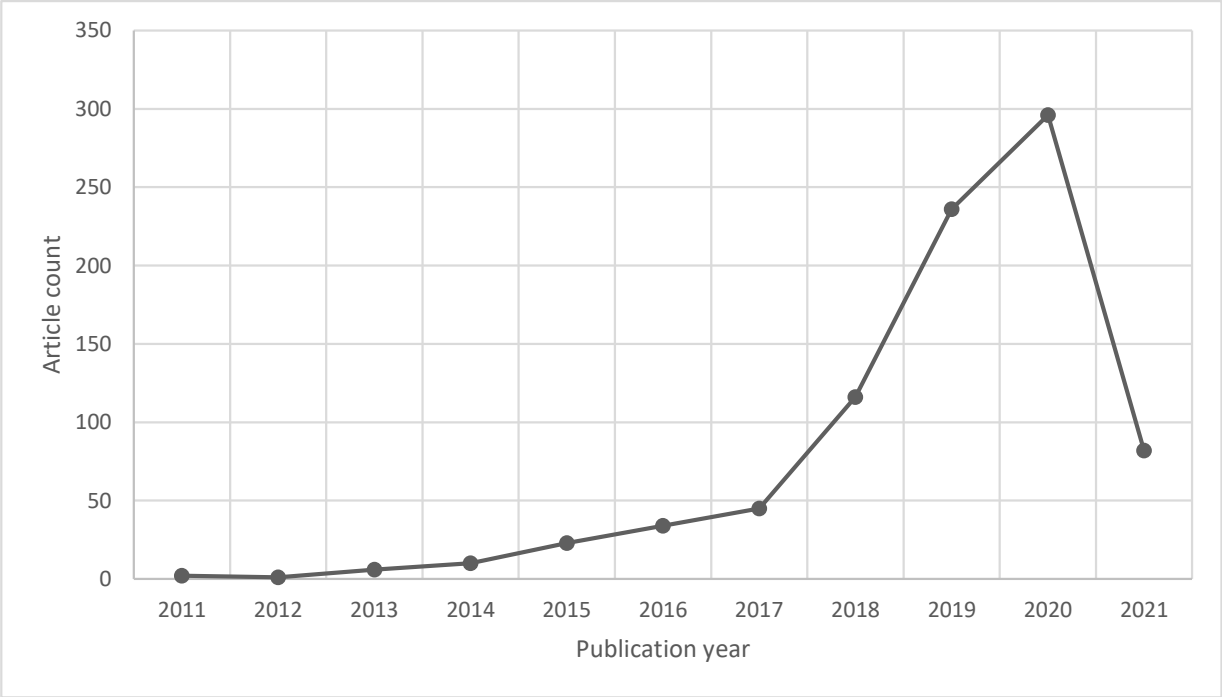


**Figure 3**: Scatterplot of article count vs. publication year (the publications for 2021 (82) were counted up to October)

### 3.4. ML applications intended for clinical practice

The systematic review captured a wide variety of ML applications for potential clinical use (**Table 1**). Most ML applications lie in the outcome prediction domain (92% or 33 studies) that focuses mainly on inflammatory diseases, including sepsis (seven studies), cancer and neuropsychiatric diseases (five studies each), Covid-19 (4 studies), haematological diseases (three studies), cardiovascular diseases, and infectious diseases (two studies each). However, few examples demonstrated the strict application of ML in classification tasks (8% or three studies). Of 36 studies, 35 involved disease-related or clinical outcomes (including mortality); only one study demonstrated the application of ML in predicting

prescription errors. It is important to note that the demarcation between prediction and classification is unclear, implying that the models appeared to perform both prediction and classification. It is surprising to note that among the 36 studies appraised, only three studies, Corny et al., [23], Kukar et al. [52], and Kim et al., [55], presented the evidence of deployment in clinical practice; 7 studies showed clear evidence of external validation, while four studies made their codes openly accessible.

For possible clinical applications, XGBoost (extreme gradient boosting algorithm) was the top-performing ML algorithm found to be the most popular among the studies sampled in the systematic review (7 studies). Next to this is Random Forest, Logistic Regression, and Support Vector Machine, respectively (5 studies each). This is followed by Gradient-boosting Decision Trees (4 studies), Bayesian algorithm (3 studies), and Artificial Neural Networks (2 studies). On the other hand, modified decision/regression trees convoluted neural networks, deep learning systems, and other proprietary algorithms (DSaaS, Predictor Pursuit, *Adabag,* and Survival Quilts) were among the ML models least featured.

### 3.5. Performance metrics

Of 36 publications reviewed, 34 specified feature variables (median 33; range 5 – 5930). All the studies specified some form of performance metric (outcome measure). Of 36 publications, 30 specified the AUROC curve of the best algorithm used for their respective studies: median 0.85; range 0.70 – 0.99. The remaining publications used other performance measures like accuracy, specificity, sensitivity, etc. Of the 30 studies, 20 (67%) had an AUROC curve ≥ of 0.80, the threshold for the acceptable performance (**Table 2**). All the studies indicated a sample size that varied from 24 – 2,759,529; median sample size = 4091. No apparent correlation was observed between model performance and the sample size; however, surprisingly, all algorithms included in the SR failed to justify the use of sample size. This may also explain why item 8 was omitted from the revised checklist by Wang et al. [17], adapted for this systematic review. By and large, ALL studies included in the SR, except three, met at least 50% of the TRIPOD criteria (20 items).

## 4. Discussion

### 4.1. Main findings

Over the review period, there was a rising trend in deploying ML algorithms for clinical applications (**Figure 2**). The review suggested that the XGBoost, RF algorithm, Logistic regression, and support vector machine were popular algorithms potentially deployed for clinical applications. Notably, Wang et al. [17] recently carried out a systematic review of the ML algorithm by domain-specific clinical application and concluded that RF was the most popular algorithm.

Random forest is an ensemble algorithm that handles a large set of predictor variables (high dimensionality of feature space), which are split through multiple independent decision trees, sampled (randomised), and aggregated in a manner that minimises bias and optimises performance in terms of output. The output could be nominal (regression) or categorical (classification). Key among several advantages of the algorithm, as highlighted by Del Parigi et al. [56], Gunčar et al. [33], and Kouchaki et al. [39], is the efficient computation speed and the fact that it handles missing data well (common in EHR). Minimal hyperparameter tuning/feature pre-processing is required for optimal performance. A major drawback of the RF algorithm is the increasing computational costs as the datasets become larger.

On the other hand, gradient boosting algorithms, another ensemble algorithm, perform better on larger datasets and have the advantages of RF coupled with shallower architecture, less bias, and computation time if tuned correctly [37,21]. In addition, several simple DT assembled to ensure decision-making (performance) improves iteratively as the tree grows [44]. However, the model is more prone to overfitting when compared to random forests (which employ bagging) but is more suitable for high bias/unbalanced datasets. The bias or variance trade-off is a key consideration for boosting or bagging technique(s) to avoid mislabelling test data.

Among several algorithms applied in clinical practice, SVM and Logistic Regression are noteworthy. The SVM predicts outcomes by using optimal hyperplane to classify highly dimensional features (i.e., a vast array of data points) mathematically transformed by Kernel function(s). This makes the model highly accurate and robust, ideal for heterogeneous datasets common in medical records [47,57]. Though powerful, the model is hard to interpret and usually requires a lot of parameter tuning [14]. The Logistic Regression, as summarised by Kouchaki et al. [39], accurately predicts categorical outcomes by using a different approach (decision weights), and predictors are mapped in a way that minimises overfitting. Its efficiency and ease of interpretation make it a popular choice for clinical use. However, the model cannot handle non-linear problems for which SVM is the best choice.

There is significant work still required before the implementation of ML into the clinical workflow can be realised. There were inconsistent reporting performance metrics for the ML algorithms in the studies included in the systematic review. For example, 83% of the included studies reported AUROC, while 47% reported accuracy. This is coupled with a lack of strict adherence to quality assessment checklists (e.g., TRIPOD and PROBAST).

### 4.2. Potential role in neuropsychiatry and mental health

Support vector machine algorithm has been used to classify patients with schizophrenia [58] accurately. Similarly, Kim & Na [59] described the potential application of machine learning in individual-level classifications of mood disorders. This makes a strong case for ML use in computer-aided diagnosis in real-world clinical practice.

An RF algorithm had been deployed in a feasibility study to diagnose depression in primary care [60]. Kautzky et al. [61] combined supervised and unsupervised ML algorithms to predict antidepressant treatment response. Meanwhile, the RF algorithm was also used to improve the prediction of response to medication in treatment-resistant depression [62]. Finally, Icten et al. [63] identified reliable predictors for early detection of Parkinson's disease with the aid of machine learning (XGBoost).

### 4.3. Potential role in infectious diseases

Few researchers have addressed the problem of multidrug-resistant UTIs, including Mancini et al. [47]_ see **Table S2**. In their review, Anahtar et al. [64] hypothesised that ML could solve the global threat of antimicrobial resistance (AMR) by predicting AMR using genomics, unravelling the mechanistic basis of antibiotic effect (to develop better anti-infectives), and antibiotic stewardship based on data from the EHR.

Currently, the COVID-19 pandemic has received global attention from researchers, policymakers, and the public due to its devastating impact on socioeconomic and health systems. As a result, machine learning has been identified as a valuable tool for waging a successful war on pandemic potential viruses (**Table 3**). In addition, the clinical research literature has identified potential applications of machine learning techniques [65,66,67] that demonstrated superiority over the conventional classification benchmarks.

### 4.4. Potential role in cancer

Early diagnosis, therapy, and prognosis remain a challenge for most forms of cancer, including haematological malignancies [12]. However, ML can apply pattern detection (diagnosis) and outcome prediction from patient tumour samples and clinical metadata [10]. Also, sensitivities to different therapies in acute myeloid leukaemia (AML) have been predicted based on DNA and RNA sequencing and *in vitro* drug sensitivity testing [12].

ML has catalysed the implementation of personalised medicine by making sense of big data. Personal health records were used to build a model capable of predicting the 5-year breast cancer personalised risk that can be a cost-effective and non-invasive risk stratification tool for early detection and prevention of breast cancer [14]. ML have also shown invaluable application in precision cardio-oncology, as demonstrated by Zhou et al. [80].

### 4.5. Potential role in diabetes mellitus

Del Parigi et al. [56] showed that ML (Random Forest) could be used to determine the predictors of the highest improvement in glycemic control for an individual patient from the data pooled from randomized controlled trials. In addition, Fujihara et al. [81] have shown that ML (Neural Networks) can aid clinicians to decide, with high prediction accuracy, when to initiate an insulin regimen for their patients with Type 2 diabetes. Crutzen et al. [82] employed a collection of algorithms to identify patients with type 2 diabetes at an increased risk of hypoglycaemia in primary care.

### 4.6. Potential role in cardiovascular health

Cardiovascular diseases are the major cause of global morbidity and mortality [83]. A range of ML algorithms can diagnose, classify, and predict heart failure [84] with an immensely complex pathophysiology. ML, therefore, can help uncover unseen predictor variables associated with the

disease outcomes. Moreover, ML can combine the decision weights of multiple algorithms, including multiple feature variables from complex datasets, to reliably predict an individual's response to medication [85].

Gibson et al. [86] compared the ensemble model to traditional methods to predict adverse events among ACS (acute coronary syndrome) patients. The ensemble model was found superior in stratifying (and/or predicting) thrombotic and bleeding risks, respectively. Finally, using atrial fibrillation as a case study, Siontis et al. [87] provided a concise summary of possible machine learning applications to cardiovascular medicine. For example, when deciding whether oral anticoagulation is required in high-risk patients without AF in stroke risk classification.

### 4.7. Limitations of machine learning techniques and this study

Very few authors [88] have cited the actual deployment of ML in clinical workflows. This systematic review conveys what the authors deem to be a fair representation of the recent trends in clinical applications of machine learning; not the full potential or actual ML applications could have been captured (i.e., the actual count of included studies are fewer than the total literature published, Figure 1). Therefore, selection bias is inevitable, and the publication bias cannot be ruled out from the review.

ML is widely lauded for savings costs, increased efficiency, throughput, accuracy, and precision. However, several questions have been raised on reliability, ethics, transparency, and generalisability. Increased ethical concerns and mounting debate may hinder ML's mainstream adoption in clinical practice. The risk of algorithmic bias is pervasive; therefore, emphasis is laid on the quality of data collected from the source [89]. ML algorithms primarily rely on the data used to create the model; poor data quality, therefore, negatively affects the outcome measures and may lead to inappropriate healthcare interventions in clinical practice [90]. Also, complex models built by massive datasets are very difficult to interpret and explain. This black box nature of the popular machine learning algorithms is very challenging for its application in real-world settings [7].

Only a few prospective studies to validate ML models for improved clinical outcomes have successfully been carried out. Therefore, there is an urgent need for more real-world clinical studies to validate existing models [52], for instance, in cardiovascular therapeutics [89]. Perhaps, ML models are only ideal for high-dimensionality problems where traditional models are overwhelmed with the perpetuity of data points. ML is not suitable for all clinical scenarios and should be limited to a decision support tool at best, not a substitute for human clinical judgement [91].

### 4.8. Implications for practice and research

From this systematic review, it can be inferred that ML algorithms are yet to be deployed successfully in clinical practice. Many feasibility studies have not been successfully transitioned to clinical practice as yet. Moreover, acceptable quality standards for real-world applications have not been fully met. It is hoped that the heightened pace of external validation added to the normal (current) pace of model development and internal validation will aid the translation to the clinical practice in the near future.

It is well known that future toxicity and safety challenges (associated with increased use of medicines and patient diversity) stress the limits of traditional statistical techniques and pharmacometrics. Therefore, massive amounts of newly available heterogeneous data present an opportunity for adding machine learning to improve drug safety (perhaps effectiveness) at a much lower computing cost.

For ML to live up to its hype, optimism must be tempered with caution. Care must be taken to avoid overreliance on algorithms, especially those that do not give clinically meaningful outputs. There is plenty of scope for improvement in this regard.

**Summary points**

**What is already known on the topic:**

- ML algorithms are becoming more versatile: There is an ongoing expansion in therapeutic or clinical use cases.
- The performance standards of ML algorithms (e.g., AUC and accuracy) in terms of prediction and classification have improved over the past decade, albeit it is still primarily based on internal validation protocols.
- External validation is acknowledged as the key translational gap in clinical practice, and attempts are being made to bridge this gap.
- The full potential of ML is yet to be appreciated; consistency and reliability standards need to be addressed to translate ML models into global clinical practice successfully.

**What this study adds:**

- A comprehensive review of studies exploring the use of Machine Learning in diverse areas of clinical therapeutics
- A concise assessment of eligible studies on clinically oriented ML algorithms against established quality standards for its implementation and reporting
- An up-to-date survey of the extent of ML algorithms' deployment in clinical practice

## References

1. Char DS, Abràmoff MD, Feudtner C. Identifying Ethical Considerations for Machine Learning Healthcare Applications. *Am J Bioeth AJOB*. 2020;20(11):7–17.

2. Ngiam KY, Khor IW Big data and machine learning algorithms for healthcare delivery. *Lancet Oncol*. 2019;20(5):e262–73

3. Damiati SA. Digital Pharmaceutical Sciences. AAPS PharmSciTech. 2020;21(6):206.

4. Luo G. MLBCD: a machine learning tool for big clinical data. *Health Inf Sci Syst*. 2015; 28(3):3.

5. Kubota KJ, Chen JA, Little MA. Machine learning for large-scale wearable sensor data in Parkinson's disease: Concepts, promises, pitfalls, and futures. *Mov Disord Off J Mov Disord Soc*. 2016;31(9):1314–26.

6. Goulooze SC, Zwep LB, Vogt JE, Krekels EHJ, Hankemeier T, van den Anker JN, et al. Beyond the Randomized Clinical Trial: Innovative Data Science to Close the Pediatric Evidence Gap. *Clin Pharmacol Ther*. 2020;107(4):786–95.

7. Hong S, Lee S, Lee J, Cha WC, Kim K. Prediction of Cardiac Arrest in the Emergency Department Based on Machine Learning and Sequential Characteristics: Model Development and Retrospective Clinical Validation Study. *JMIR Med Inform*. 2020;8(8):e15932.

8. Goecks J, Jalili V, Heiser LM, Gray JW. How Machine Learning Will Transform Biomedicine. *Cell*. 2020;181(1):92–101.

9. Jayatilake SMDAC, Ganegoda GU. Involvement of Machine Learning Tools in Healthcare Decision Making. *J Healthc Eng*. 2021;2021:6679512.

10. Jagga Z, Gupta D. Machine learning for biomarker identification in cancer research - developments toward its clinical application. *Pers Med*. 2015;12(4):371–87.

11. Lo Vercio L, Amador K, Bannister JJ, Crites S, Gutierrez A, MacDonald ME, et al. Supervised machine learning tools: a tutorial for clinicians. *J Neural Eng* [Internet]. 2020; Available from: http://search.ebscohost.com/login.aspx?direct=true&db=cmedm&AN=33036008&site=ehost-live

12. Radakovich N, Nagy M, Nazha A. Machine learning in haematological malignancies. *Lancet Haematol*. 2020;7(7):e541–50.

13. Rashidi P, Edwards DA, Tighe PJ. Primer on machine learning: utilization of large data set analyses to individualize pain management. *Curr Opin Anaesthesiol*. 2019;32(5):653–60.

14. Stark GF, Hart GR, Nartowt BJ, Deng J. Predicting breast cancer risk using personal health data and machine learning models. *PloS One*. 2019;14(12):e0226765.

15. Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. *N Engl J Med*. 2019;380(14):1347-1358. doi: 10.1056/NEJMra1814259. PMID: 30943338.

16. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD Statement. *Circulation*. 2020;131(2):211-9. https://www.equator-network.org/reporting-guidelines/tripod-statement/

17. Wang W, Kiik M, Peek N, Curcin V, Marshall IJ, Rudd AG, et al. A systematic review of machine learning models for predicting outcomes of stroke with structured data. *PloS One*. 2020;15(6):e0234722.

18. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, Reitsma JB, Kleijnen J, Mallett S; PROBAST Group†. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. Ann Intern Med. 2019 Jan 1;170(1):51-58. doi: 10.7326/M18-1376. PMID: 30596875.

19. Li J, Zhou Z, Dong J, Fu Y, Li Y, Luan Z, Peng X. Predicting breast cancer 5-year survival using machine learning: A systematic review. PLoS One. 2021;16(4):e0250370.

20. Hung M, Voss MW, Rosales MN, Li W, Su W, Xu J, et al. Application of machine learning for diagnostic prediction of root caries. *Gerodontology*. 2019;36(4):395–404.

21. Zhai Q, Lin Z, Ge H, Liang Y, Li N, Ma Q, et al. Using machine learning tools to predict outcomes for emergency department intensive care unit patients. *Sci Rep*. 2020;10(1):20919.

22. Bergeron MF, Landset S, Zhou X, Ding T, Khoshgoftaar TM, Zhao F, et al. Utility of MemTrax and Machine Learning Modeling in Classification of Mild Cognitive Impairment. *J Alzheimers Dis JAD*. 2020;77(4):1545–58.

23. Corny J, Rajkumar A, Martin O, Dode X, Lajonchere J-P, Billuart O, et al. A machine learning-based clinical decision support system to identify prescriptions with a high risk of medication error. *J Am Med Inform Assoc JAMIA* [Internet]. 2020; Available from: https://www.cochranelibrary.com/central/doi/10.1002/central/CN-02192671/full

24. Shouval R, Labopin M, Bondi O, Mishan-Shamay H, Shimoni A, Ciceri F et al. Prediction of Allogeneic Hematopoietic Stem-Cell Transplantation Mortality 100 Days After Transplantation Using a Machine Learning Algorithm: A European Group for Blood and Marrow Transplantation Acute Leukemia Working Party Retrospective Data Mining Study. *J Clin Oncol.* 2015;33(28):3144-51. doi: 10.1200/JCO.2014.59.1339. Epub 2015 Aug 3. PMID: 26240227.

25. Patel A, Gan K, Li A, Weiss J, Nouraie S, Tayur S, et al. Machine learning algorithms in predicting hospital readmissions in sickle cell disease. *Blood* [Internet]. 2019;134. Available from: https://www.cochranelibrary.com/central/doi/10.1002/central/CN-02048753/full

26. Zhdanov A, Atluri S, Wong W, Vaghei Y, Daskalakis ZJ, Blumberger DM, et al. Use of Machine Learning for Predicting Escitalopram Treatment Outcome From Electroencephalography Recordings in Adult Patients With Depression. *JAMA Netw Open*. 2020;3(1):e1918377.

27. Wu Y, Fang Y. Stroke Prediction with Machine Learning Methods among Older Chinese. *Int J Environ Res Public Health*. 2020;17(6).

28. Koutsouleris N, Kahn RS, Chekroud AM, Leucht S, Falkai P, Wobrock T, et al. Multisite prediction of 4-week and 52-week treatment outcomes in patients with first-episode psychosis: a machine learning approach. *Lancet Psychiatry*. 2016;3(10):935–46.

29. 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*. 2017;24(1):45–53.

30. Turki T, Wang JTL. Clinical intelligence: New machine learning techniques for predicting clinical drug response. *Comput Biol Med*. 2019;107:302–22.

31. Diao X, Huo Y, Yan Z, Wang H, Yuan J, Wang Y, et al. An Application of Machine Learning to Etiological Diagnosis of Secondary Hypertension: Retrospective Study Using Electronic Medical Records. *JMIR Med Inform*. 2021;9(1):e19739.

32. Ross MK, Yoon J, van der Schaar A, van der Schaar M. Discovering Pediatric Asthma Phenotypes on the Basis of Response to Controller Medication Using Machine Learning. *Ann Am Thorac Soc*. 2018;15(1):49–58.

33. Gunčar G, Kukar M, Notar M, Brvar M, Černelč P, Notar M, et al. An application of machine learning to haematological diagnosis. *Sci Rep*. 2018;8(1):411.

34. Veeranki SPK, Hayn D, Jauk S, Quehenberger F, Kramer D, Leodolter W, et al. An Improvised Classification Model for Predicting Delirium. *Stud Health Technol Inform*. 2019;264:1566–7.

35. Lee C, Light A, Alaa A, Thurtle D, van der Schaar M, Gnanapragasam VJ. Application of a novel machine learning framework for predicting non-metastatic prostate cancer-specific mortality in men using the Surveillance, Epidemiology, and End Results (SEER) database. *Lancet Digit Health* [Internet].

2021; Available from: http://search.ebscohost.com/login.aspx?direct=true&db=cmedm&AN=33549512&site=ehost-live

36. Alhazmi A, Alhazmi Y, Makrami A, Masmali A, Salawi N, Masmali K, et al. Application of artificial intelligence and machine learning for prediction of oral cancer risk. *J Oral Pathol Med Off Publ Int Assoc Oral Pathol Am Acad Oral Pathol* [Internet]. 2021; Available from: http://search.ebscohost.com/login.aspx?direct=true&db=cmedm&AN=33394536&site=ehost-live

37. Shim J-G, Kim DW, Ryu K-H, Cho E-A, Ahn J-H, Kim J-I, et al. Application of machine learning approaches for osteoporosis risk prediction in postmenopausal women. *Arch Osteoporos*. 2020;15(1):169.

38. Islam MM, Wu C-C, Poly TN, Yang H-C, Li Y-CJ. Applications of Machine Learning in Fatty Live Disease Prediction. *Stud Health Technol Inform*. 2018;247:166–70.

39. Kouchaki S, Yang Y, Walker TM, Sarah Walker A, Wilson DJ, Peto TEA, et al. Application of machine learning techniques to tuberculosis drug resistance analysis. *Bioinforma Oxf Engl*. 2019;35(13):2276–82.

40. Li C, Yao Z, Zhu M, Lu B, Xu H. Biopsy-Free Prediction of Pathologic Type of Primary Nephrotic Syndrome Using a Machine Learning Algorithm. *Kidney Blood Press Res*. 2017;42(6):1045–52.

41. Xiao J, Ding R, Xu X, Guan H, Feng X, Sun T, et al. Comparison and development of machine learning tools in the prediction of chronic kidney disease progression. *J Transl Med*. 2019;17(1):119.

42. Banu A B, Thirumalaikolundusubramanian P. Comparison of Bayes Classifiers for Breast Cancer Classification. *Asian Pac J Cancer Prev APJCP*. 2018;19(10):2917–20.

43. Elfiky A, Pany M, Parikh R, Obermeyer Z. Development and Application of a Machine Learning Approach to Assess Short-term Mortality Risk Among Patients With Cancer Starting Chemotherapy. *JAMA Netw Open*. 2018;1(3):e180926-.

44. Delahanty RJ, Alvarez J, Flynn LM, Sherwin RL, Jones SS. Development and Evaluation of a Machine Learning Model for the Early Identification of Patients at Risk for Sepsis. *Ann Emerg Med*. 2019;73(4):334–44.

45. Brinker TJ, Hekler A, Enk AH, Berking C, Haferkamp S, Hauschild A, Weichenthal M, Klode J, Schadendorf D, Holland-Letz T, von Kalle C, Fröhling S, Schilling B, Utikal JS. Deep neural networks are superior to dermatologists in melanoma image classification. *Eur J Cancer*. 2019;119:11-17. doi: 10.1016/j.ejca.2019.05.023. Epub 2019 Aug 8. PMID: 31401469.

46. Bora A, Balasubramanian S, Babenko B, Virmani S, Venugopalan S, Mitani A et al. Predicting the risk of developing diabetic retinopathy using deep learning. *The Lancet Digital Health*. 2020; 3(1): E10-E19. doi:10.1016/s2589-7500(20)30250-8

47. Mancini A, Vito L, Marcelli E, Piangerelli M, De Leone R, Pucciarelli S, et al. Machine learning models predicting multidrug resistant urinary tract infections using 'DsaaS'. *BMC Bioinformatics*. 2020;21(Suppl 10):347.

48. Deshmukh F & Merchant SS. Explainable machine learning model for predicting GI bleed mortality in the Intensive Care Unit. *The American Journal of Gastroenterology*. 2020; 115(10): 1657–1668. https://doi.org/10.14309/ajg.0000000000000632

49. Wang S, Pathak J, & Zhang Y. Using electronic health records and machine learning to predict postpartum depression. *Studies in Health Technology and Informatics*. 2019; 264; 888–892. https://doi.org/10.3233/SHTI190351

50. Wernly B, Mamandipoor B, Baldia P, Jung C, & Osmani V. Machine learning predicts mortality in septic patients using only routinely available ABG variables: a multi-centre evaluation. *International journal of medical informatics*. 2021;145:104312. https://doi.org/10.1016/j.ijmedinf.2020.104312

51. Behnoush B, Bazmi E, Nazari SH, Khodakarim S, Looha MA, & Soori H. Machine learning algorithms to predict seizure due to acute tramadol poisoning. *Human & Experimental Toxicology*. 2021; 960327121991910. Advance online publication. https://doi.org/10.1177/0960327121991910

52. Kukar M, Guncar G, Vovko T, Podnar S, Cernelc P, Brvar M, et al. COVID-19 diagnosis by routine blood tests using machine learning. Sci Rep. 2021 May;11(1):10738–10738.

53. Domínguez-Olmedo JL, Gragera-Martínez Á, Mata J, Pachón Álvarez V. Machine Learning Applied to Clinical Laboratory Data in Spain for COVID-19 Outcome Prediction: Model Development and Validation. J Med Internet Res. 2021 Apr;23(4):e26211–e26211.

54. Bolourani S, Brenner M, Wang P, McGinn T, Hirsch JS, Barnaby D, et al. A Machine Learning Prediction Model of Respiratory Failure Within 48 Hours of Patient Admission for COVID-19: Model Development and Validation. J Med Internet Res. 2021 Jan;23(2):e24246–e24246.

55. Kim H-J, Han D, Kim J-H, Kim D, Ha B, Seog W, et al. An Easy-to-Use Machine Learning Model to Predict the Prognosis of Patients With COVID-19: Retrospective Cohort Study. J Med Internet Res. 2020 Oct;22(11):e24225–e24225.

56. Del Parigi A, Tang W, Liu D, Lee C, Pratley R. Machine Learning to Identify Predictors of Glycemic Control in Type 2 Diabetes: An Analysis of Target HbA1c Reduction Using Empagliflozin/Linagliptin Data. *Pharm Med*. 2019;33(3):209–17.

57. Basile AO, Yahi A, Tatonetti NP. Artificial Intelligence for Drug Toxicity and Safety. *Trends Pharmacol Sci.* 2019;40(9):624–35.

58. Veronese E, Castellani U, Peruzzo D, Bellani M, Brambilla P. Machine learning approaches: from theory to application in schizophrenia. *Comput Math Methods Med*. 2013;2013:867924.

59. Kim Y-K, Na K-S. Application of machine learning classification for structural brain MRI in mood disorders: Critical review from a clinical perspective. *Prog Neuropsychopharmacol Biol Psychiatry*. 2018;80(Pt B):71–80.

60. Souza Filho EM de, Veiga Rey HC, Frajtag RM, Arrowsmith Cook DM, Dalbonio de Carvalho LN, Pinho Ribeiro AL, et al. Can machine learning be useful as a screening tool for depression in primary care? *J Psychiatr Res*. 2021;132:1–6.

61. Kautzky A, Möller H-J, Dold M, Bartova L, Seemüller F, Laux G, et al. Combining machine learning algorithms for prediction of antidepressant treatment response. *Acta Psychiatr Scand*. 2021;143(1):36–49.

62. Kautzky A, Dold M, Bartova L, Spies M, Vanicek T, Souery D, et al. Refining Prediction in Treatment-Resistant Depression: results of Machine Learning Analyses in the TRD III Sample. *J Clin Psychiatry*

[Internet]. 2018;79(1). Available from: https://www.cochranelibrary.com/central/doi/10.1002/central/CN-01952309/full

63. Icten Z, Burns S, Menzin J. AI2 predictors of parkinson disease in a medicare population: an application of machine learning in early disease detection. *Value Health*. 2020;23:S1-.

64. Anahtar MN, Yang JH, Kanjilal S. Applications of machine learning to the problem of antimicrobial resistance: An emerging model for translational research. *J Clin Microbiol* [Internet]. 2021; Available from: http://search.ebscohost.com/login.aspx?direct=true&db=cmedm&AN=33536291&site=ehost-live

65. Assaf D, Gutman Y, Neuman Y, Segal G, Amit S, Gefen-Halevi S, et al. Utilization of machine-learning models to accurately predict the risk for critical COVID-19. *Intern Emerg Med*. 2020;15(8):1435–43.

66. Zimmerman A, Kalra D. Usefulness of machine learning in COVID-19 for the detection and prognosis of cardiovascular complications. *Rev Cardiovasc Med*. 2020;21(3):345–52.

67. Xu M, Ouyang L, Han L, Sun K, Yu T, Li Q, et al. Accurately Differentiating Between Patients With COVID-19, Patients with Other Viral Infections, and Healthy Individuals: Multimodal Late Fusion Learning Approach. *J Med Internet Res*. 2021;23(1):e25535.

68. Lichtner G, Balzer F, Haufe S, Giesa N, Schiefenhövel F, Schmieding M, et al. Predicting lethal courses in critically ill COVID-19 patients using a machine learning model trained on patients with non-COVID-19 viral pneumonia. Sci Rep. 2021 Jun;11(1):13205–13205.

69. Fernandes FT, de Oliveira TA, Teixeira CE, Batista AF de M, Dalla Costa G, Chiavegatto Filho ADP. A multipurpose machine learning approach to predict COVID-19 negative prognosis in São Paulo, Brazil. Sci Rep. 2021 Feb;11(1):3343–3343.

70. Halasz G, Sperti M, Villani M, Michelucci U, Agostoni P, Biagi A, et al. A Machine Learning Approach for Mortality Prediction in COVID-19 Pneumonia: Development and Evaluation of the Piacenza Score. J Med Internet Res. 2021 May;23(5):e29058–e29058.

71. Stachel A, Daniel K, Ding D, Francois F, Phillips M, Lighter J. Development and validation of a machine learning model to predict mortality risk in patients with COVID-19. 2021 May;28(1). Available from: https://dx.doi.org/10.1136/bmjhci-2020-100235

72. Jimenez-Solem E, Petersen TS, Hansen C, Hansen C, Lioma C, Igel C, et al. Developing and validating COVID-19 adverse outcome risk prediction models from a bi-national European cohort of 5594 patients. Sci Rep. 2021 Feb;11(1):3246–3246.

73. Kivrak M, Guldogan E, Colak C. Prediction of death status on the course of treatment in SARS-COV-2 patients with deep learning and machine learning methods. Comput Methods Programs Biomed. 2021 Jan;201:105951–105951.

74. Schöning V, Liakoni E, Baumgartner C, Exadaktylos AK, Hautz WE, Atkinson A, et al. Development and validation of a prognostic COVID-19 severity assessment (COSA) score and machine learning models for patient triage at a tertiary hospital. J Transl Med. 2021 Feb;19(1):56–56.

75. Patel D, Kher V, Desai B, Lei X, Cen S, Nanda N, et al. Machine learning based predictors for COVID-19 disease severity. Sci Rep. 2021 Feb;11(1):4673–4673.

76. Marcos M, Belhassen-García M, Sánchez-Puente A, Sampedro-Gomez J, Azibeiro R, Dorado-Díaz P-I, et al. Development of a severity of disease score and classification model by machine learning for hospitalized COVID-19 patients. PLoS One. 2021 Apr;16(4):e0240200–e0240200.

77. Navlakha S, Morjaria S, Perez-Johnston R, Zhang A, Taur Y. Projecting COVID-19 disease severity in cancer patients using purposefully-designed machine learning. BMC Infect Dis. 2021 May;21(1):391–391.

78. Magunia H, Lederer S, Verbuecheln R, Gilot BJ, Koeppen M, Haeberle HA, et al. Machine learning identifies ICU outcome predictors in a multicenter COVID-19 cohort. Crit Care. 2021 Aug;25(1):295–295.

79. Dai W, Ke P-F, Li Z-Z, Zhuang Q-Z, Huang W, Wang Y, et al. Establishing Classifiers With Clinical Laboratory Indicators to Distinguish COVID-19 From Community-Acquired Pneumonia: Retrospective Cohort Study. J Med Internet Res. 2021 Feb;23(2):e23390–e23390.

80. Zhou Y, Hou Y, Hussain M, Brown S-A, Budd T, Tang WHW, et al. Machine Learning-Based Risk Assessment for Cancer Therapy-Related Cardiac Dysfunction in 4300 Longitudinal Oncology Patients. *J Am Heart Assoc*. 2020;9(23):e019628.

81. Fujihara K, Matsubayashi Y, Harada Yamada M, Yamamoto M, Iizuka T, Miyamura K, et al. Machine Learning Approach to Decision Making for Insulin Initiation in Japanese Patients with Type 2 Diabetes (JDDM 58): Model Development and Validation Study. *JMIR Med Inform*. 2021;9(1):e22148.

82. Crutzen S, Nagaraj SB, 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*. 2020;e3426.

83. Roth JA, Battegay M, Juchler F, Vogt JE, Widmer AF. Introduction to Machine Learning in Digital Healthcare Epidemiology. *Infect Control Hosp Epidemiol*. 2018;39(12):1457–62.

84. Olsen CR, Mentz RJ, Anstrom KJ, Page D, Patel PA. Clinical applications of machine learning in the diagnosis, classification, and prediction of heart failure. *Am Heart J*. 2020;229:1–17.

85. MacGregor RM, Guo A, Masood MF, Cupps BP, Ewald GA, Pasque MK, et al. Machine Learning Outcome Prediction in Dilated Cardiomyopathy Using Regional Left Ventricular Multiparametric Strain. *Ann Biomed Eng*. 2021 ;49(2):922–32.

86. Gibson WJ, Nafee T, Travis R, Yee M, Kerneis M, Ohman M, et al. Machine learning versus traditional risk stratification methods in acute coronary syndrome: a pooled randomized clinical trial analysis. *J Thromb Thrombolysis*. 2020;49(1):1–9.

87. Siontis KC, Yao X, Pirruccello JP, Philippakis AA, Noseworthy PA. How Will Machine Learning Inform the Clinical Care of Atrial Fibrillation? *Circ Res*. 2020;127(1):155–69.

88. Sendak MP, D' Arcy J, Kashayap S, Gao M, Nichols M, Corey K, Ratliff W et al. A Path for Translation of Machine Learning Products into Healthcare Delivery. EMJ Innovations [Internet]. 2020;. Available from: https://www.emjreviews.com/innovations/article/a-path-for-translation-of-machine-learning-products-into-healthcare-delivery/

89. Russak AJ, Chaudhry F, De Freitas JK, Baron G, Chaudhry FF, Bienstock S, et al. Machine Learning in Cardiology-Ensuring Clinical Impact Lives Up to the hype. *J Cardiovasc Pharmacol Ther.* 2020;25(5):379–90.

90. Noseworthy PA, Attia ZI, Brewer LC, Hayes SN, Yao X, Kapa S, et al. Assessing and Mitigating Bias in Medical Artificial Intelligence: The Effects of Race and Ethnicity on a Deep Learning Model for ECG Analysis. *Circ Arrhythm Electrophysiol*. 2020;13(3):e007988.

91. Shouval R, Fein JA, Savani B, Mohty M, Nagler A. Machine learning and artificial intelligence in haematology. *Br J Haematol*. 2021;192(2):239–50.

**Table 1:** Summary table showing performance of ML tools in included studies

| Authors, year | Country | Sample size | Features | AUC | Accuracy | ML tool(s) | Potential Application | Evidence of Clinical Deployment | External validation | Code accessible |
|---|---|---|---|---|---|---|---|---|---|---|
| Hung et al, 2019[19] | USA | 5135 | 37 | 0.99 | 0.97 | SVM | Diagnostic prediction of dental caries | No | No | No |
| Zhai et al, 2020 [20] | China | 1706 | 75 | 0.85 | 0.84 | XGBoost | Prediction of mortality in emergency care setting | No | No | No |
| Bergeron et al,2020 [21] | USA & China | 259 | 29 | 0.91 | | NB | Patient care, diagnostic support, and classification aid for patients with cognitive impairment (e.g., dementia) | No | No | No |
| Corny et al, 2020 [22] | France | 10716 | 25 | 0.81 | | GBDT | Clinical Decision Support tool (prescription checking software) for reducing risk of prescribing errors; Improves safety in high-risk patients | Yes | No | No |
| Shouval et al, 2015 [23] | Israel | 28236 | 20 | 0.7 | | ADT | Prediction of survival following allogenic hematopoietic stem cell transplantation (allo-HSCT) | No | No | No |
| Patel et al, 2019 [24] | USA | 446 | 481 | 0.74 | | LR | Predicting hospital admissions due to sickle cell disease | No | No | No |
| Zhdanov et al, 2020 [25] | Canada | 122 | 127 | | 0.82 | SVM | Predicting escitalopram treatment outcome in adult patients with depression (EEG features used as predictors) | No | No | No |
| Wu & Fang, 2020 [26] | China | 1131 | 15 | 0.71 | 0.78 | RF | Reliable stroke prediction using demographic, lifestyle, and clinical data | No | No | No |
| Koutsouleris et al, 2016 [27] | Germany, USA, Netherlands, Austria | 334 | 189 | | 0.74 | SVM | Prediction of treatment outcomes in patients with first episode of psychosis | No | No | No |
| Waljee et al, 2017 [28] | USA | 20368 | 26 | 0.88 | | RF | Distinguishing patients at high and low risk of Inflammatory Bowel Disease (IBD) flares at point of care (individualised | No | No | No |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | therapeutic management e.g., with corticosteroids) | | | |
| Turki & Wang, 2019 [29] | USA & Saudi Arabia | 24 | | 0.76 | 0.66 | SVM | Accurate prediction of clinical drug response (sensitivity) of patients with breast cancer, triple-negative breast cancer, and multiple myeloma (among other cancer types) | No | No | Yes |
| Diao et al, 2020 [30] | China | 7532 | 22 | 0.92 | 0.86 | XGBoost | Diagnosis of secondary hypertension | No | No | No |
| Ross et al, 2017 [31] | USA & United Kingdom | 2394 | 656 | 0.85 | | Proprietary | Identification of phenotypes that best predict controller medication response in paediatric asthma patients | No | No | No |
| Gunčar et al, 2018 [32] | Switzerland & Slovenia | 8233 | 61 | 0.86 | | RF | Prediction/diagnosis of haematological diseases based only on blood tests | No | yes | No |
| Veeranki et al, 2019 [33] | Austria | 24927 | 500 | 0.91 | | RF + LR | More effective prediction of delirium compared to other traditional methods | No | No | No |
| Lee et al, 2021 [34] | USA & United Kingdom | 171942 | 33 | 0.83 | | Proprietary | Prediction of mortality due to non-metastatic prostate cancer in men | No | Yes | No |
| Alhazmi et al, 2021 [35] | Saudi Arabia | 73 | 29 | | 0.79 | ANN | Prediction of individualised oral cancer risk | No | No | No |
| Shim et al, 2020 [36] | South Korea | 1792 | 19 | 0.74 | 0.72 | ANN | Osteoporosis risk prediction in routine clinical practice | No | No | No |
| Islam et al, 2018[37] | Taiwan | 994 | 10 | 0.76 | 0.7 | LR | Accurate, timely and effective prediction of fatty liver disease from electronic health records (EHR) | No | No | No |
| Kouchaki et al, 2019[38] | United Kingdom | 13402 | 5919 | 0.92 | | LR | Prompt prediction of drug resistance to *Mycobacterium tuberculosis* (this includes detection of resistance markers) | No | No | No |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Li et al, 2017 [39] | China | 222 | 33 | | 0.62 | RF | Reliable prediction of the pathology type of nephrotic syndrome without the aid of renal biopsy | No | Yes | No |
| Xiao et al, 2019 [40] | China | 551 | 18 | 0.87 | 0.82 | LR | Prediction of chronic kidney disease (CKD) progression using demographic and blood chemistry features during follow up | No | No | No |
| Bazila & Ponniah, 2018 [41] | India | 569 | 32 | | 0.94 | NB+GBDT | Breast cancer detection and classification | No | No | No |
| Elfiky et al, 2018 [42] | USA | 26946 | 5930 | 0.94 | | GBDT | Accurate prediction of short-term mortality risk (from EHRs) among patients initiating cancer chemotherapy | No | Yes | Yes |
| Delahanty et al, 2019 [43] | USA | 2759529 | 217 | 0.97 | | GBDT | Superior risk screening tool for sepsis compared to traditional screening tools | No | No | No |
| Brinker et al, 2019 [44] | Germany | 4202 | | | | CNN | Screening of melanoma or non-melanoma skin cancers based on dermatopathology images | No | No | No |
| Bora et al, 2020 [45] | USA & Thailand | 575431 | 20 | 0.81 | | DL | Early screening of diabetic retinopathy using colour fundus photographs | No | Yes | Yes |
| Mancini et al, 2020 [46] | Italy | 1486 | 5 | 0.74 | 0.72 | Proprietary | Clinical decision support (prediction) tool for treating patients with high risk of acquiring multi-drug resistant urinary tract infections (UTIs) | No | No | No |
| Deshmukh & Merchant, 2020 [47] | USA | 5691 | 34 | 0.85 | | XGBoost | Enhanced risk stratification of patients with gastrointestinal bleeding | No | No | Yes |
| Wang et al, 2020 [48] | USA & China | 9980 | 98 | 0.79 | | SVM | Accurate prediction of post-partum depression (PPD). | No | No | No |
| Wernly et al, 2021 [49] | Austria, Sweden, | 3979 | 23 | 0.88 | | RNN (DL) | Mortality prediction of sepsis patients in ICU setting. | No | No | No |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Germany, Italy | | | | | | | | | |
| Behnoush et al, 2021[50] | Iran | 909 | 25 | 0.71 | 0.69 | NBC | Accurate prediction of seizure occurrence (due to acute tramadol poisoning) within 24 hrs post ED admission | No | No | No |
| Kukar et al, 2021[51] | Slovenia, Switzerland | 5493 | 35 | 0.97 | | XGBoost | Efficient diagnosis of Covid-19 based on routine blood tests. | Yes | No | No |
| Domínguez-Olmedo et al, 2021[52] | Spain | 1823 | 32 | 0.97 | 0.94 | XGBoost | Prediction of severity and mortality risk associated with Covid-19. | No | No | No |
| Bolourani et al, 2021[53] | USA | 11525 | 34 | 0.77 | 0.92 | XGBoost | Efficient triage by prediction of 48-hour respiratory failure in Covid-19 patients | No | Yes | No |
| Kim et al, 2020[54] | Korea Republic | 4787 | 36 | 0.897 | | XGBoost | Predicting the need for Intensive care based on routine patient data | Yes | Yes | No |

**Note:** Ethical approval in line with declaration of Helsinki was given for all the studies. All the models were implemented successfully. **NB**: AUC results within 95% confidence interval

**Table 2:** Summary table showing performance of ML tools in included studies

| S/N | Author(s) | Sample size | Features | AUC | Accuracy | ML tool(s) |
|---|---|---|---|---|---|---|
| 1 | Hung *et al*, 2019 | 5135 | 37 | 0.99 | 0.97 | SVM |
| 2 | Zhai *et al*, 2020 | 1706 | 75 | 0.85 | 0.84 | XGBoost |
| 3 | Bergeron *et al*, 2020 | 259 | 29 | 0.91 | | NB |
| 4 | Corny *et al*, 2020 | 10716 | 25 | 0.81 | | GBDT |
| 5 | Shouval *et al*, 2015 | 28236 | 20 | 0.7 | | ADT |
| 6 | Patel *et al*, 2019 | 446 | 481 | 0.74 | | LR |
| 7 | Zhdanov *et al*, 2020 | 122 | 127 | | 0.82 | SVM |
| 8 | Wu & Fang, 2020 | 1131 | 15 | 0.71 | 0.78 | RF |
| 9 | Koutsouleris *et al*, 2016 | 334 | 189 | | 0.74 | SVM |
| 10 | Waljee *et al*, 2017 | 20368 | 26 | 0.88 | | RF |
| 11 | Turki & Wang, 2019 | 24 | | 0.76 | 0.66 | SVM |
| 12 | Diao *et al*, 2020 | 7532 | 22 | 0.92 | 0.86 | XGBoost |
| 13 | Ross *et al*, 2017 | 2394 | 656 | 0.85 | | Proprietary |
| 14 | Gunčar *et al*, 2018 | 8233 | 61 | 0.86 | | RF |
| 15 | Veeranki *et al*, 2019 | 24927 | 500 | 0.91 | | RF + LR |
| 16 | Lee *et al*, 2021 | 171942 | 33 | 0.83 | | Proprietary |
| 17 | Alhazmi *et al*, 2021 | 73 | 29 | | 0.79 | ANN |
| 18 | Shim *et al*, 2020 | 1792 | 19 | 0.74 | 0.72 | ANN |
| 19 | Islam *et al*, 2018 | 994 | 10 | 0.76 | 0.7 | LR |
| 20 | Kouchaki *et al*, 2019 | 13402 | 5919 | 0.92 | | LR |
| 21 | Li *et al*, 2017 | 222 | 33 | | 0.62 | RF |
| 22 | Xiao *et al*, 2019 | 551 | 18 | 0.87 | 0.82 | LR |
| 23 | Bazila & Ponniah, 2018 | 569 | 32 | | 0.94 | NB+GBDT |
| 24 | Elfiky *et al*, 2018 | 26946 | 5930 | 0.94 | | GBDT |
| 25 | Delahanty *et al*, 2019 | 2759529 | 217 | 0.97 | | GBDT |
| 26 | Brinker *et al*, 2019 | 4202 | | | | CNN |
| 27 | Bora *et al*, 2020 | 575431 | 20 | 0.81 | | DL |
| 28 | Mancini *et al*, 2020 | 1486 | 5 | 0.74 | 0.72 | Proprietary |
| 29 | Deshmukh & Merchant, 2020 | 5691 | 34 | 0.85 | | XGBoost |
| 30 | Wang *et al*, 2020 | 9980 | 98 | 0.79 | | SVM |
| 21 | Wernly *et al*, 2021 | 3979 | 23 | 0.88 | | RNN (DL) |
| 32 | Behnoush *et al*, 2021 | 909 | 25 | 0.71 | 0.69 | NBC |

**Table 3** Summary of potential applications of ML in COVID-19

| s/n | Potential clinical applications | Reference |
|---|---|---|
| 1 | Prognostic assessment | Lichtner et al. [67]; Fernandez et al. [68] |
| 2 | Mortality (risk) prediction | Halasz et al. [69]; Stachel et al. [70]; Jimenez-Solem et al. [71]; Kivrak et al. [72] |
| 3 | Triage of patients and assignment of severity (risk stratification) | Schöning et al. [73]; Patel et al. [74]; Marcos et al. [75]; Navlakha et al. [76]; Magunia et al. [77] |
| 4 | Diagnosis/Differential diagnosis (e.g differentiating community-acquired pneumonia (CAP) from COVID-19) | Dai et al. [78]; Xu et al. [66] |