Machine Learning for Predicting Stillbirth: A Systematic Review

Qingyuan Li[1] · Pan Li[2] · Junyu Chen[3] · Ruyu Ren[3] · Ni Ren[3] · Yinyin Xia[3]

Abstract

Stillbirth is a major global issue, with over 5 million cases each year. The multifactorial nature of stillbirth makes it difficult to predict. Artificial intelligence (AI) and machine learning (ML) have the potential to enhance clinical decision-making and enable precise assessments. This study reviewed the literature on predictive ML models for stillbirth highlighting input characteristics, performance metrics, and validation. The PubMed, Cochrane, and Web of Science databases were searched for studies using AI to develop predictive models for stillbirth. Findings were analyzed qualitatively using narrative synthesis and graphics. Risk of bias and the applicability of the studies were assessed using PROBAST. Model design and performance were discussed. Eight studies involving 14,840,654 women with gestational ages ranging from 20 weeks to full term were included in the qualitative analysis. Most studies used neural networks, random forests, and logistic regression algorithms. The number of predictive features varied from 14 to 53. Only 50% of studies validated the models. Cross-validation was commonly employed, and only 25% of studies performed external validation. All studies reported area under the curve as a performance metric (range 0.54–0.9), and five studies reported sensitivity (range, 60– 90%) and specificity (range, 64 − 93.3%). A stacked ensemble model that analyzed 53 features performed better than other models (AUC = 0.9; sensitivity and specificity > 85%). Available ML models can attain a considerable degree of accuracy for prediction of stillbirth; however, these models require further development before they can be applied in a clinical setting.

Abbreviations

AI	Artificial intelligence
ML	Machine learning
WHO	World Health Organisation
ENAP	Every Newborn Action Plan
ROC	Receiver operating characteristic
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
LR	Logistic regression
ANN	Artificial neural networks
RF	Random forest
DT	Decision tree
SVM	Support vector machine
SE	Stacked ensemble
CDC	Centers for Disease Control and Prevention
NYC	New York City Department of Health and Mental Hygiene

✉ Yinyin Xia
100118@cqmu.edu.cn

1 Department of Clinical Medicine, International Medical College of Chongqing Medical University, Yixueyuan Road No.1, Yuzhong District, Chongqing 400016, China

2 Department of Clinical Medicine, Southwest Medical University, Zhongshan Road, No.319 Section 3, Luzhou 646000, China

3 School of Public Health, Chongqing Medical University, Yixueyuan Road No.1, Yuzhong District, Chongqing 400016, China

Background

The World Health Organisation (WHO) defines stillbirth as the death of an infant after more than 28 weeks of gestation, but before or during labor [1, 2]. The personal, emotional, and financial impact of stillbirth have a profound effect on parents, healthcare providers, and society [3, 4]. In an effort to end preventable stillbirth, in 2014, the WHO Every Newborn Action Plan (ENAP) set a target of ≤ 12 stillbirths per

1000 total births for every country by 2030. Despite this, there are over 5 million cases of stillbirth globally each year, and the incidence of stillbirth remains particularly high in middle and low-income countries, with rates reaching 22.8 stillbirths per 1000 [5] total births in some regions [1].

The etiology of stillbirth is multifactorial, but interventions aimed at prevention are effective [6]. Globally, there is no standardized system of investigating and reporting stillbirth, and available information is classified using numerous and disparate systems [7]. There is an unmet need to collect quality information on the causes of stillbirth to inform predictive models of stillbirth [1, 8].

Simple linear statistics lack the capacity to model complex problems such as stillbirth. Advances in the processing power, memory, and storage of computers and the widespread avaiability of rich datasets have led to the application of artificial intelligence (AI) and machine learning (ML) in healthcare to improve risk prediction [9]. ML can integrate vast and heterogeneous datasets and identify patterns and correlations. ML algorithms are created and trained to make classifications or predictions [10, 11]. The performance of an ML model may be explained by its accuracy, assessed using metrics such as sensitivity, specificity, predictive value, probability ratio, and the area under the receiver operating characteristic (ROC) curve (AUC) [12].

ML has the potential to improve early disease prediction, diagnosis, and treatment in maternal-fetal medicine [13], and has been used to assess fetal well-being and predict and diagnose gestational diabetes, pre-eclampsia, preterm birth and fetal growth restriction [14]. Predictive modes established using ML can inform clinical decision-making but should not be used to make definitive diagnoses. An increasing number of models are being constructed to screen and monitor pregnancies and detect those at high-risk of stillbirth; however, there is no comprehensive systematic review of the latest advances in this field [15]. This study reviewed the literature on predictive ML models for stillbirth, highlighting input characteristics, performance metrics, and validation. Findings should improve care relevant to stillbirth.

Methods

This study is reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [16]. The protocol that incudes these analyses was registered on the PROSPERO database under the ID CRD42022380270.

To account for the disparate systems used for classifying stillbirth globally, for the purpose of this study, stillbirth was defined as the death of an infant after 20 weeks of gestation, but before or during labor or at birth.

The PubMed, Cochrane, and Web of Science databases were searched for studies using AI (e.g. ML and deep learning) to develop predictive models for stillbirth. Search terms included "stillbirth", "fetal death", "mortality", "death", "artificial intelligence", "machine learning", "deep learning", "predictive models" (Please see Supplementary Table 1).

Inclusion criteria were: (1) use of a ML technique to develop and/or validate a predictive model for stillbirth; (2) the model integrated data for ≥ 2 variables (features) for stillbirth prediction; and (3) comprehensive performance assessment was conducted. Exclusion criteria were: (1) reviews, meta-analyses, conference abstracts, letters or comments; (2) studies conducted in animals; (3) studies with a sample size < 100 stillbirths as machine learning requires a certain sample size for model training [17]; (4) studies investigating the molecular mechanisms underlying stillbirth; or (5) studies that did not propose a predictive model for stillbirth.

Duplicate references were eliminated, and two reviewers (Q.L. and P.L.) independently screened titles and abstracts and reviewed the remaining full text articles to determine which studies met the inclusion criteria.

The two reviewers independently extracted information pertaining to the construction of predictive models for stillbirth from the included studies, including number of models created, features used and their contribution to model prediction, and use of hyperparameter optimization/tuning, internal and/or external validation, and calibration analysis. Full details of the features used in the predictive models are provided in Supplementary Table 2.

The two reviewers independently assessed risk of bias and the applicability of included studies using PROBAST, a tool that consists of 20 signaling questions in four domains: participants, predictors, outcomes, and analysis [18].

Discrepancies between reviewers were resolved through consultation with a third reviewer (J-Y.C.). Findings were analyzed qualitatively using narrative synthesis and graphics generated by Revman and R version 4.2.3.

Results

Study Selection

The initial search identified 455 studies. Titles and abstracts were screened and 368 studies were excluded, including 128 duplicates and 240 studies that did not meet the inclusion criteria. The full text of 49 studies were reviewed, and 8 studies were included in the qualitative analysis **(Fig. 1).**
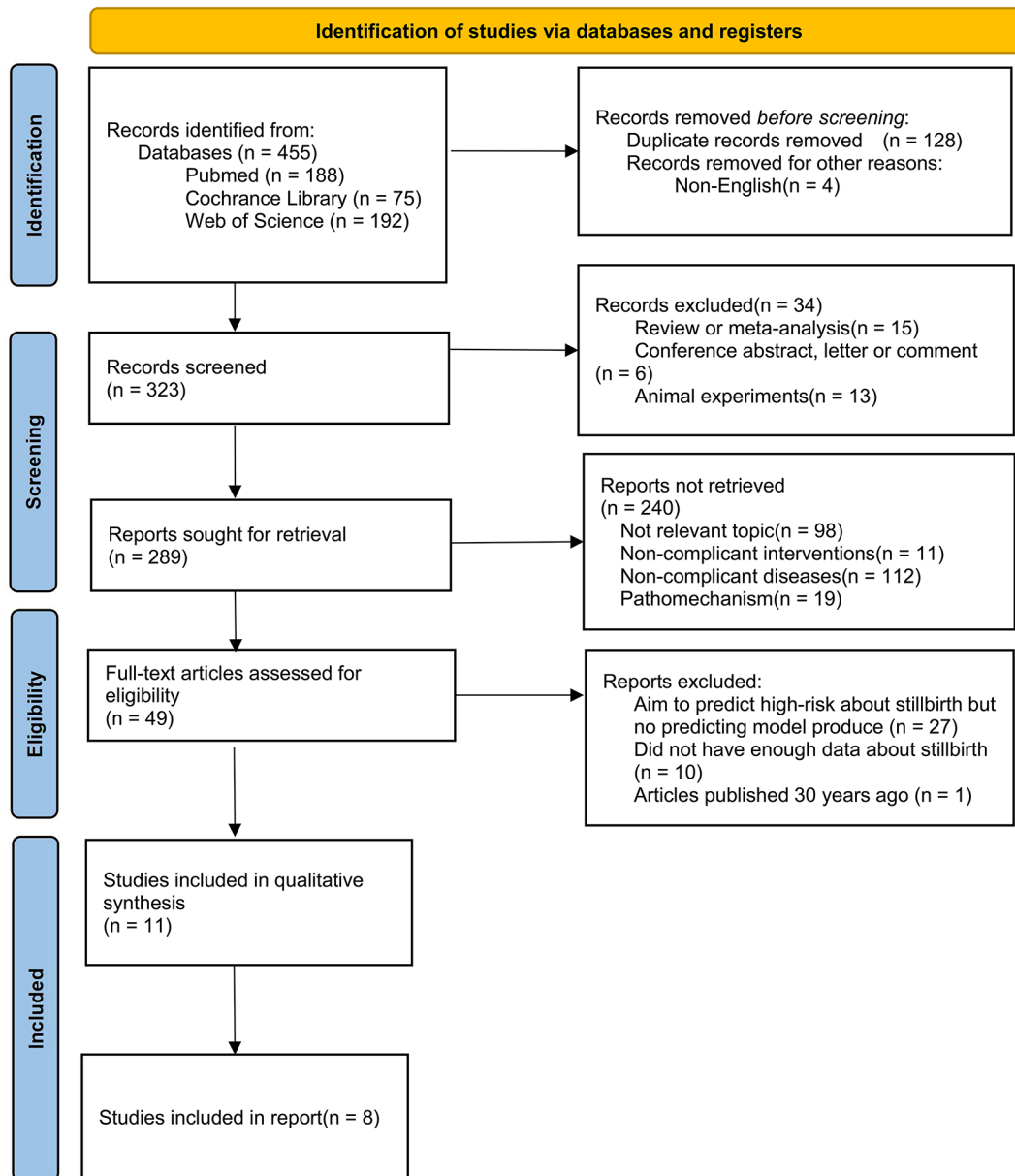
Fig. 1 PRISMA flow diagram of study selection

Study Characteristics

The characteristics of the included studies are summarized in Table 1. These studies described 84 (range 1–30) predictive models for stillbirth created using 22 ML algorithms applied across various datasets.

All studies explicitly outlined their inclusion and exclusion criteria. 62.5% of the studies reported on methods for handling missing data; of these, 80% of studies omitted cases with missing data. All studies reported AUC as a performance metric for the models. Five studies provided information on the sensitivity and specificity of the models [19–24]. Two studies gave detection rates for a 5% and 10%

false positive rate [24, 25]. One study reported the positive likelihood ratio, negative likelihood ratio, positive predictive value, and negative predictive value [25].

Risk of Bias

The risk of bias assessment based on the PROBAST tool is shown in Supplementary Table 3. Three studies had a "high risk", of bias, one study had a "low risk" of bias, and four studies had an "unclear" risk of bias.

Table 1 Characteristics of included studies

Reference	Geographic area	Time range	Sample size	Target population	Final number of included features	Handling of missing values	Validation	Types of prediction models	Prediction models, n
Åmark 2018 [19]	Sweden	2009 to 2016	Total = 145,319; overweight=45,859; normal(18.5–24.9)=96,067	obesity women	22	Calculate median BMI and stillbirth incidence as a proxy	Internal validation: K-fold cross-validation	LR	1
Khatibi 2021 [20]	Iranian	2016/04/01 to 2017/01/01	Total = 1,431,597 live birth=1,415,623 stillbirth=5502	Births at 28 weeks or > 28 weeks	53	Remove cases where the timing of stillbirth is unclear	Internal validation: K-fold cross-validation; external validation	DT, GBC, LR, RF, SVM, SE	24
Koivu_Sairanen_2020 [45]	American	2013 to 2016(CDC), 2014 to 2016(NYC)	CDC: normal =11,901,611, early stillbirth=7924, late stillbirth=8310; NYC: normal=266,419, early stillbirth=139, late stillbirth=110	Births > 22weeks	26	Not reported	Internal validation: 10-fold CV and bootstrapping; external validation	LR, DNN, SELU, LGBM, averaged ensemble, WA ensemble	24
Kumar 2022 [21]	South-east Asian region	2015/11 to 2020/12	Experimental group=265,Control group=288	Hypertensive Women and births > 20 weeks	15	Not reported	Internal validation: K-fold cross-validation	LR	1
Malacova et al_2020 [25]	Western Australia	1980/01/01 to 2015/12/31	Total = 964,813 stillbirth=7788	>20week OR weight<400 g	33	Yes	Internal validation: 10-fold CV	LR, DT, RF, XGBoost, MLP	30
Meng 2021 [22]	Shandong	2011/09 to 2020/11	Total = 106 stillbirth=16	Acute Fatty Liver of Pregnancy	41	Patients with severely incomplete data were excluded from the research	Internal validation	LR, MELD	2
Wu 2019 [23]	Shanghai, China	2011/09 to 2017/05	Total = 338 stillbirth=6	SLE and baby born without signs of life or > 28 weeks	22	Yes	Internal validation:10-fold CV and two tailed	LR	1
Yerlikaya et al_2016 [24]	UK	2006/03 to 2015/10	Total = 113,415 live-births=113,019 stillbirth=396	Births > 24weeks	14	Not reported	Internal validation	LR	1

Characteristics of Predictive Models

The most common ML algorithms included in the predictive models were logistic regression (LR), artificial neural networks (ANN), and random forest (RF) [26]. Less than 50% of the studies conducted model calibration, hyperparameter tuning, and external validation.

Across studies, 226 predictive features of stillbirth were identified, including 154 distinct features. Finally, 15 to 53 features were analyzed in the predictive models. Certain features emerged as potential predictors of stillbirth (Fig. 2). The top five predictors were age, parity, hypertension, smoking, and miscarriage. Other features that predicted > 50% of stillbirths included maternal body mass index, place of birth, maternal education, previous stillbirth, gestational diabetes, gestational hypertension, diabetes, and intrauterine growth restriction.

The performance of all models was assessed using AUC. The performance of the machine learning models by dataset are summarized as a box plot (Fig. 3). Mean AUC is shown by the line that divides the box into two parts. The whiskers represent the standard deviation. Outliers are depicted as black circles. The predictive models had a mean AUC > 0.7 (range, 0.54–0.9). When reported, sensitivity (range, 60–90%) and specificity (range 64–93.3%) of most models was

> 70%. Notably, the stacked ensemble (SE) model proposed by Khatibi et al. (2021) [15] achieved an AUC of 0.9 and sensitivity and specificity > 85% (Figs. 4 and 5).

Discussion

Stillbirth remains a significant public health concern worldwide. Mothers who have previously experienced stillbirth are at a considerably higher risk of recurrence in subsequent pregnancies. Given the wide-reaching impact of stillbirth, accurately predicting this condition is of paramount importance. However, conventional methods for stillbirth prediction, such as statistical modeling, may be insufficient due to limitations in their ability to account for the multifactorial determinants of stillbirth and the complex interactions between potential risk factors [27, 28]. Unlike statistical modeling, ML-based predictive models do not require a priori selection of predictors. Instead, they are capable of automatically and thoroughly exploring the complex associations and interactions among potential risk factors and outcomes. This approach allows the outcome to be investigated in great depth and facilitates the identification of new insights in complex systems [29–34]. Therefore, ML-based
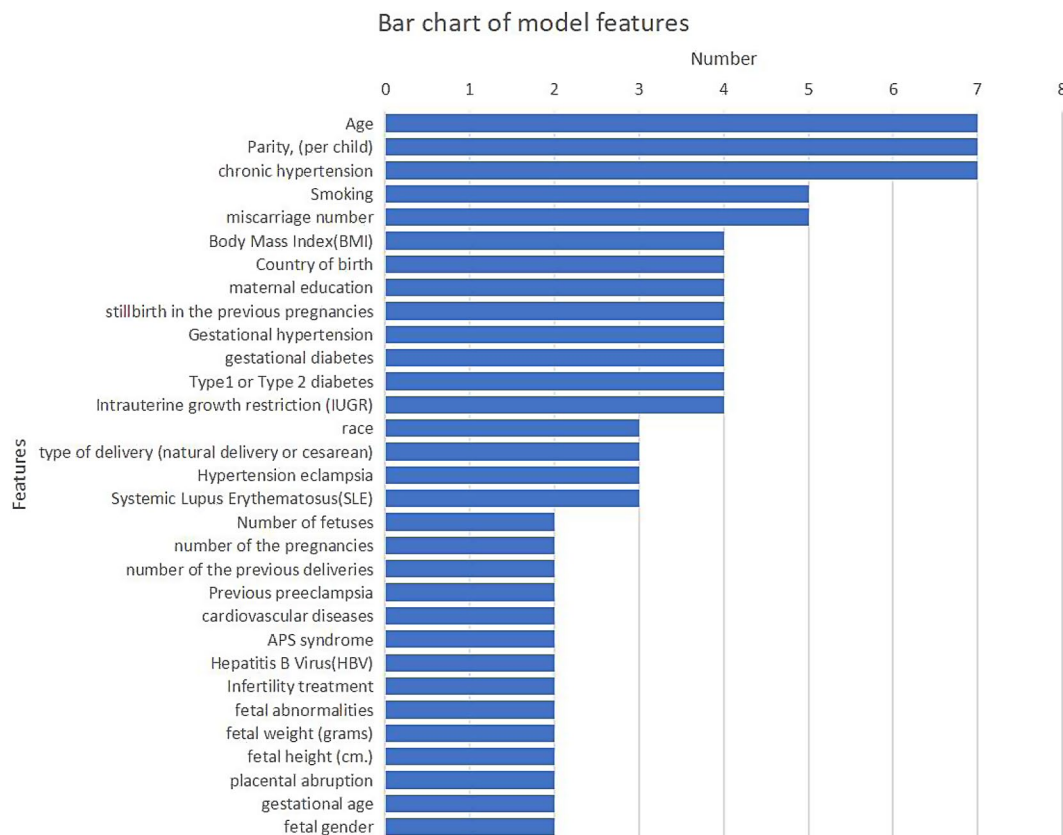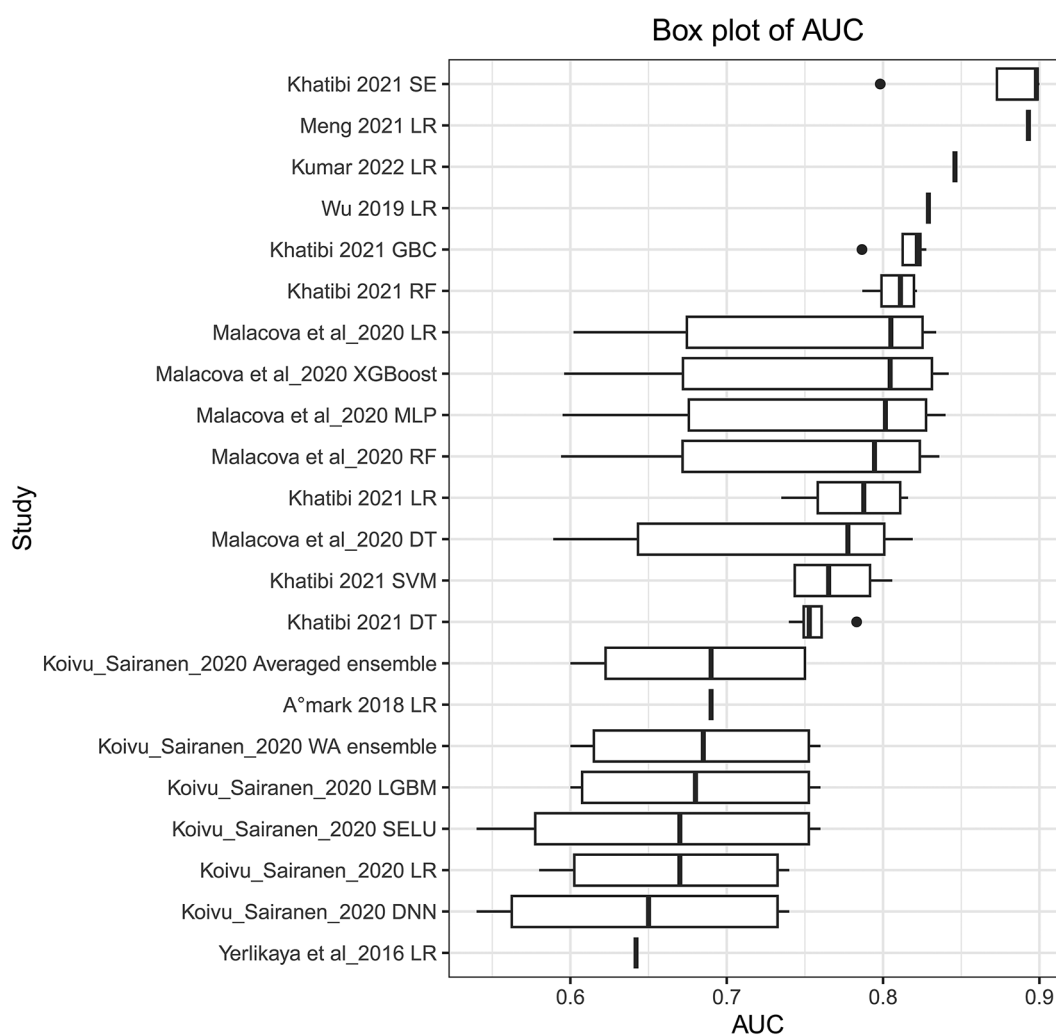

Fig. 2 Bar chart of model features. BMI: Body Mass Index; SLE: Systemic Lupus Erythematosus; HBV: Hepatitis B Virus

Box plot of AUC

Fig. 3 Box plot of AUC

predictive models represent a promising tool for enhancing the accuracy and effectiveness of stillbirth prediction [35].

Despite numerous studies exploring predictive models for stillbirth, there are a lack of systematic evaluations consolidating and assessing the effectiveness of these models [11, 36–38]. To fill this gap, the present study undertook the first systematic qualitative analysis of published predictive ML models for stillbirth. Findings identified the strengths and limitations of existing models and opportunities for improvements.

Selection of Predictive Features

ML algorithms such as decision tree (DT), support vector machine (SVM), and RF can account for non-linear and high dimensional relationships, which may lead to better predictive performances over traditional prediction methods that involve statistical modeling [9, 39, 40]. When using ML algorithms, selecting appropriate datasets and predictive

factors are crucial for model development, training and validation and achieving optimal results [9].

This qualitative analysis revealed the selection of predictive factors for stillbirth varied across published ML models. The influence of factors such as maternal age, number of births, history of miscarriage, infectious diseases during pregnancy, and smoking status were consistent with the results of clinical studies. Preexisting conditions such as hypertension, diabetes, and obesity were significantly associated with stillbirth. As the choice of features can significantly impact the accuracy and efficiency of a predictive ML model, there is a need to select the right input features and achieve a balance between accuracy and data limitations [39–42]. The characteristics of the target population should be considered. For example, different predictive factors may be relevant for women with underlying chronic conditions (e.g., systemic lupus erythematosus, obesity, hypertension, and acute fatty liver) compared with healthy women [21–23].

Fig. 4 Model sensitivity
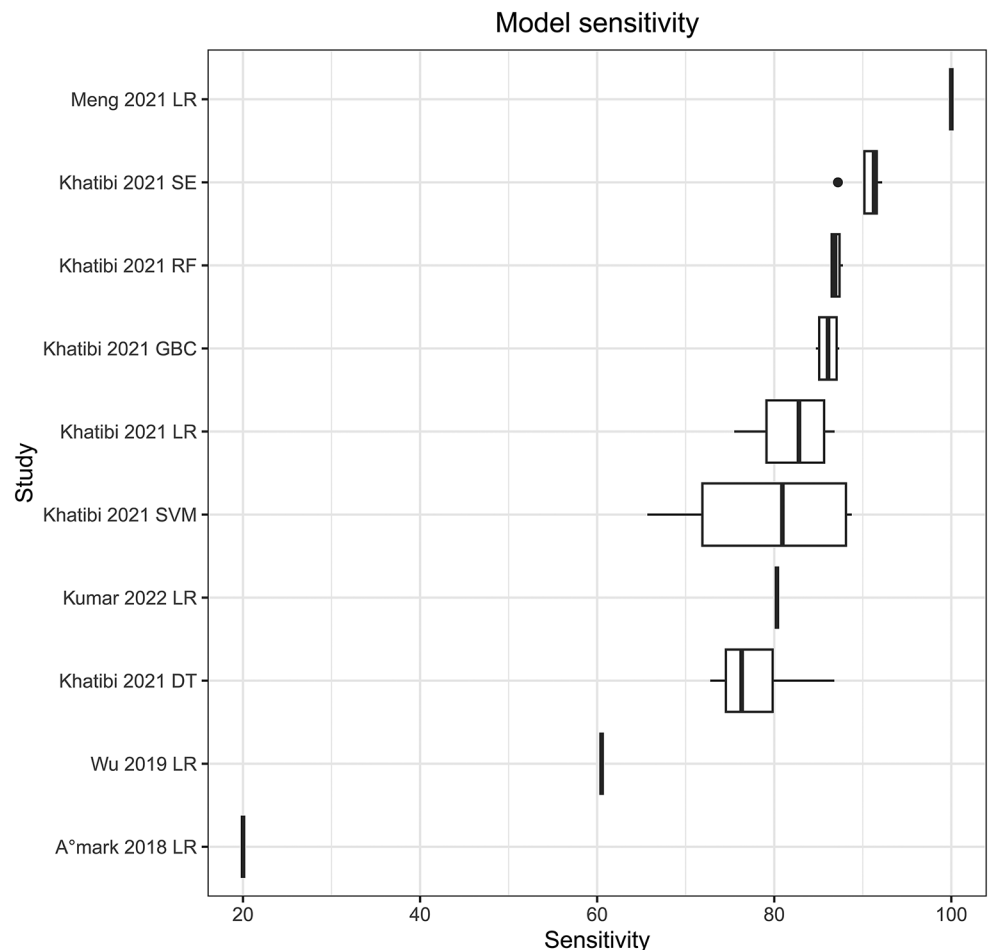
The models evaluated in this review included a range of features. The least number of features were used by Yerlikaya et al. (2016) ($n = 14$ features) and Kumar et al. (2022) ($n = 15$ features). Both studies constructed singular LR models and achieved AUCs of 0.642 and 0.846, respectively. The most features ($n = 53$) were used in the stacked ensemble (SE) model proposed by Khatibi (2021) [15], potentially making its application impractical in a clinical setting.

Missing Data

Among the 8 studies included in this qualitative synthesis, there was limited reporting on the handling of missing data. Of the 5 studies that reported on the management of missing data, 4 studies omitted data with ambiguous records, including cases with imprecise timing of stillbirth [20, 22, 23, 25]. This may have led to bias and loss of accuracy [43]. Alternative approaches to handling missing data will ensure reliable and robust predictions [43]. Among the published ML models, alternative approaches included imputing missing data utilizing records from hospital visits as supplementary data sources [25], or computing medians and prevalence of

stillbirth for pregnancies with missing data on any of the predictive variables [19].

Validation

One notable issue with the studies included in this qualitative analysis was the relative lack of validation techniques employed, with only 50% of studies utilizing model validation, and even fewer performing external validation. In the absence of scrupulous and impartial external validation, ML may generate erroneous high-risk predictions by incorrectly capturing the interconnectedness of features [44]. It is imperative for future studies to conduct comprehensive validation of their models to ensure accuracy and reliability.

Of the 8 studies reviewed, cross-validation was the most commonly employed method for validation. While all studies used internal validation, only two studies also used external datasets. The predictive model of stillbirth constructed by Koivu [45] was developed using two datasets, namely Centers for Disease Control and Prevention (CDC) and New York City Department of Health and Mental Hygiene (NYC). The external validation of the model utilized the NYC dataset in combination with a subset of the CDC

Fig. 5 Model specificity
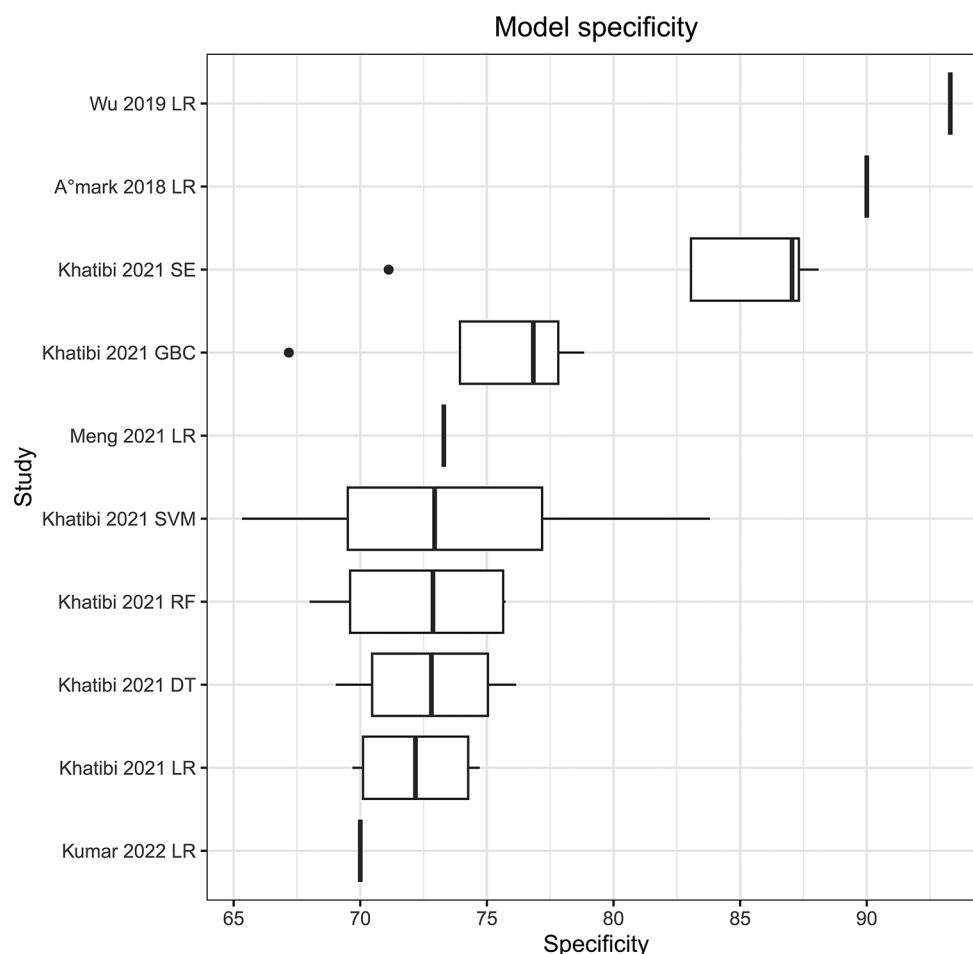
dataset, consisting of 31,429 pregnancies. In the model constructed by Khatibi [20], the dataset used in external validation included stillbirth cases that occurred between 2011 and 2018 in hospitals across different provinces in Iran. This external dataset was obtained from a different registry than the one used for model development. The feature selection process and preprocessing techniques applied during model development were also applied to the external validation dataset to ensure consistency. The remaining 6 studies did not report external validation. Yerlikaya [24] evaluated the performance of their predictive model of stillbirth across different gestational ages by analyzing data from a different cohort in California, to some extent validating its generalizability to stillbirth rates. Kumar [21] assessed the performance of their model using 30% of the data as a separate test set.

Overall, the included literature lacked sufficient reporting on external validation, indicating a need for future ML predictive models to recognize this as a necessity. Without comprehensive assessment using a diverse collection of datasets, the external validity of a model may be limited, potentially hindering its effectiveness when applied to data

collection systems from different patient populations and regions [44, 46].

Model Performance

This qualitative analysis focused on AUC as a metric of the performance of the published predictive models for stillbirth. AUC is widely used for assessing the discriminative ability of ML-based prediction models [47]. An AUC close to 0.6 represents low discriminative ability, while an AUC close to 1 indicates high discriminative ability [48]. The majority of the models evaluated in this qualitative synthesis had an AUC between 0.7 and 0.9, and none were considered highly discriminative, highlighting the need to optimize the performance of existing predictive models for stillbirth before they can assist in clinical decision-making.

Overall, the predictive models had a mean AUC > 0.7. When reported, sensitivity and specificity of most models was > 70%, with the exception of the LR model constructed by Amark (2018) [19]. Sensitivity or specificity was highest for the LR models constructed by Meng (2021) [22] or Wu (2019) [23], respectively.

The SE model proposed by Khatibi (2021) [20] performed better than other models, achieving an AUC of 0.9 and sensitivity and specificity > 85%.

Future Perspectives

This review performed a qualitative analysis of 8 studies that included a large sample ($n = 14,840,654$) of pregnant women. Findings should be interpreted with caution as the analysis was associated with some limitations, including the need to combine studies with and without limited reporting of handling of missing data and those that did and did not conduct external validation into a single outcome. However results demonstrate the potential of ML to improve the prediction of stillbirth by analyzing large datasets to identify cases of stillbirth in populations with medical disorders and infections during pregnancy, and in populations without obvious high-risk conditions [22].

Existing predictive models for stillbirth created using ML algorithms require further development before being applied in clinical care. Optimal decision thresholds are challenging to assign due to variability in the classification of stillbirths and miscarriages across regions. Model performance has not been consistently evaluated, making comparisons between models difficult. Some studies report AUC and others rely on sensitivity and specificity. In particular, models have not been validated in external datasets, which is crucial when assessing model performance. External validation provides an understanding of the importance of individual features, and can allow comparisons between models, which is essential when multiple models claim high performance on their respective training datasets. A high AUC on the training or validation set does not guarantee a model will perform similarly on new, unseen data. Training data may cause overfitting. Training and validation datasets may represent a specific patient population or setting at a certain point in time and have inherent biases due to sampling methods and data collection procedures. External validation helps ensure a model generalizes well to new data and is robust and applicable in diverse populations or settings over time. This allows healthcare stakeholders, including researchers, clinicians, and regulatory agencies, confidence that the model can be applied in real-world scenarios.

Conclusion

This qualitative analysis revealed that available ML models can attain a considerable degree of accuracy for prediction of stillbirth; however, these models require further development before they can be applied in a clinical setting, including improved approaches to handling missing data

and external validation. The optimization of predictive ML models for stillbirth should ensure precise support of individual patients and contribute to the wellbeing of mothers, children, and society.

Registration and Protocol

1. The protocol of this review was registered on the PROSPERO database under the ID CRD42022380270.
2. The review protocol can be accessed at: https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=380270.

Declarations

Ethics Approval Not applicable. This article is a review, no animal or human experiment was conducted in this study.

Consent to Participate Not applicable.

Consent for Publication Not applicable.

Competing Interests The authors declare that they have no competing interests.

References

1. Lawn JE, Blencowe H, Waiswa P, Amouzou A, Mathers C, Hogan D, et al. Stillbirths: rates, risk factors, and acceleration towards 2030. Lancet. 2016;387:587–603.
2. Management of stillbirth. Obstetric care consensus No 10. Obstet Gynecol. 2020;135:e110–32.
3. Robinson GE. Pregnancy loss. Best practice & research clinical obstetrics & Gynaecology. 2014;28:169–78.
4. Campbell HE, Kurinczuk JJ, Heazell A, Leal J, Rivero-Arias O. Healthcare and wider societal implications of stillbirth: a population-based cost-of-illness study. BJOG. 2018;125:108–17.

5. Hug L, You D, Blencowe H, Mishra A, Wang Z, Fix MJ, et al. Global, regional, and national estimates and trends in still-births from 2000 to 2019: a systematic assessment. Lancet. 2021;398:772–85.

6. Flenady V, Koopmans L, Middleton P, Frøen JF, Smith GC, Gibbons K, et al. Major risk factors for stillbirth in high-income countries: a systematic review and meta-analysis. Lancet. 2011;377:1331–40.

7. Reinebrant HE, Leisher SH, Coory M, Henry S, Wojcieszek AM, Gardener G, et al. Making stillbirths visible: a systematic review of globally reported causes of stillbirth. BJOG. 2018;125:212–24.

8. Fretts RC. Etiology and prevention of stillbirth. Am J Obstet Gynecol. 2005;193:1923–35.

9. Deo RC. Mach Learn Med Circulation. 2015;132:1920–30.

10. Darcy AM, Louie AK, Roberts LW. Machine learning and the profession of medicine. JAMA. 2016;315:551–2.

11. Panesar SS, D'Souza RN, Yeh FC, Fernandez-Miranda JC. Machine learning versus logistic regression methods for 2-year mortality prognostication in a small. Heterogen Glioma Database World Neurosurg X. 2019;2:100012.

12. Šimundić AM. Measures of diagnostic accuracy. Basic Definitions Ejifcc. 2009;19:203–11.

13. Iftikhar P, Kuijpers MV, Khayyat A, Iftikhar A, DeGouvia De Sa M. Artificial intelligence: a new paradigm in obstetrics and gynecology research and clinical practice. Cureus. 2020;12:e7124.

14. Arain Z, Iliodromiti S, Slabaugh G, David AL, Chowdhury TT. Machine learning and disease prediction in obstetrics. Curr Res Physiol. 2023;6:100099.

15. Ahmed Z, Mohamed K, Zeeshan S, Dong X. Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. Database (Oxford). 2020;2020.

16. Urrútia G, Bonfill X. [PRISMA declaration: a proposal to improve the publication of systematic reviews and meta-analyses]. Med Clin (Barc). 2010;135:507–11.

17. Rajput D, Wang WJ, Chen CC. Evaluation of a decided sample size in machine learning applications. BMC Bioinformatics. 2023;24:48.

18. Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. Ann Intern Med. 2019;170:W1–33.

19. Åmark H, Westgren M, Persson M. Prediction of stillbirth in women with overweight or obesity-A register-based cohort study. PLoS ONE. 2018;13:e0206940.

20. Khatibi T, Hanifi E, Sepehri MM, Allahqoli L. Proposing a machine-learning based method to predict stillbirth before and during delivery and ranking the features: nationwide retrospective cross-sectional study. BMC Pregnancy Childbirth. 2021;21:202.

21. Kumar M, Ravi V, Meena D, Chopra K, Nain S, Puri M. Predictive model for late stillbirth among antenatal hypertensive women. J Obstet Gynaecol India. 2022;72:96–101.

22. Meng Z, Fang W, Meng M, Zhang J, Wang Q, Qie G, et al. Risk factors for maternal and fetal mortality in acute fatty liver of pregnancy and new predictive models. Front Med (Lausanne). 2021;8:719906.

23. Wu J, Zhang WH, Ma J, Bao C, Liu J, Di W. Prediction of fetal loss in Chinese pregnant patients with systemic lupus erythematosus: a retrospective cohort study. BMJ Open. 2019;9:e023849.

24. Yerlikaya G, Akolekar R, McPherson K, Syngelaki A, Nicolaides KH. Prediction of stillbirth from maternal demographic and pregnancy characteristics. Ultrasound Obstet Gynecol. 2016;48:607–12.

25. Malacova E, Tippaya S, Bailey HD, Chai K, Farrant BM, Gebremedhin AT, et al. Stillbirth risk prediction using machine learning for a large cohort of births from Western Australia, 1980–2015. Sci Rep. 2020;10:5354.

26. Stoltzfus JC. Logistic regression: a brief primer. Acad Emerg Med. 2011;18:1099–104.

27. Ishak M, Khalil A. Prediction and prevention of stillbirth: dream or reality. Curr Opin Obstet Gynecol. 2021;33:405–11.

28. Jardine J. Risk prediction for stillbirth: further to go. BJOG. 2023;130:1071.

29. Heo J, Yoon JG, Park H, Kim YD, Nam HS, Heo JH. Machine learning-based model for prediction of outcomes in acute stroke. Stroke. 2019;50:1263–5.

30. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. J Med Internet Res. 2016;18:e323.

31. Alanazi HO, Abdullah AH, Qureshi KN. A critical review for developing accurate and dynamic predictive models using machine learning methods in medicine and health care. J Med Syst. 2017;41:69.

32. Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. BMC Med Res Methodol. 2019;19:64.

33. Ota R, Yamashita F. Application of machine learning techniques to the analysis and prediction of drug pharmacokinetics. J Control Release. 2022;352:961–9.

34. Kausch SL, Moorman JR, Lake DE, Keim-Malpass J. Physiological machine learning models for prediction of sepsis in hospitalized adults: an integrative review. Intensive Crit Care Nurs. 2021;65:103035.

35. Khan K, Ahmad W, Amin MN, Ahmad A, Nazar S, Alabdullah AA. Compressive strength estimation of steel-fiber-reinforced concrete and raw material interactions using advanced algorithms. Polymers (Basel). 2022;14:3065.

36. Shung D, Simonov M, Gentry M, Au B, Laine L. Machine learning to Predict outcomes in patients with acute gastrointestinal bleeding: a systematic review. Dig Dis Sci. 2019;64:2078–87.

37. Manz CR, Chen J, Liu M, Chivers C, Regli SH, Braun J, et al. Validation of a machine learning algorithm to predict 180-day mortality for outpatients with cancer. JAMA Oncol. 2020;6:1723–30.

38. Steyerberg EW, Mushkudiani N, Perel P, Butcher I, Lu J, McHugh GS, et al. Predicting outcome after traumatic brain injury: development and international validation of prognostic scores based on admission characteristics. PLoS Med. 2008;5:e165 (discussion e).

39. Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. Nat Methods. 2018;15:233–4.

40. Cai J, Luo J, Wang S, Yang S. Feature selection in machine learning: a new perspective. Neurocomputing. 2018;300:70–9.

41. Spooner A, Chen E, Sowmya A, Sachdev P, Kochan NA, Trollor J, et al. A comparison of machine learning methods for survival analysis of high-dimensional clinical data for dementia prediction. Sci Rep. 2020;10:20410.

42. Bennett M, Hayes K, Kleczyk EJ, Mehta R. Similarities and differences between machine learning and traditional advanced statistical modeling in healthcare analytics. arXiv preprint arXiv:220102469. 2022.

43. Nijman S, Leeuwenberg AM, Beekers I, Verkouter I, Jacobs J, Bots ML, et al. Missing data is poorly handled and reported in prediction model studies using machine learning: a literature review. J Clin Epidemiol. 2022;142:218–29.

44. Siontis GC, Tzoulaki I, Castaldi PJ, Ioannidis JP. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. J Clin Epidemiol. 2015;68:25–34.

45. Koivu A, Sairanen M, Airola A, Pahikkala T. Synthetic minority oversampling of vital statistics data with generative adversarial networks. J Am Med Inf Assoc. 2020;27:1667–74.

46. Naimi AI, Balzer LB. Stacked generalization: an introduction to super learning. Eur J Epidemiol. 2018;33:459–64.

47. Caetano SJ, Sonpavde G, Pond GR. C-statistic: a brief explanation of its construction, interpretation and limitations. Eur J Cancer. 2018;90:130–2.

48. Steyerberg EW, Harrell FE Jr., Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J Clin Epidemiol. 2001;54:774–81.