European Journal of Clinical Pharmacy
Print ISSN: 2385-409X, **Online ISSN**: 3105-0409
Website: https://farmclin.com/

Machine Learning for Neonatal Survival Prediction: A Systematic Review and Meta-analysis with Implications for Subphenotype-Based Modeling

Article History:

Name of Author:
Restu Octasila[1], Besral[2], Puji Laksmini[3], Siti Dariyani[4], Darin Zahra[5]

Affiliation: *[1]Faculty of Public Health, Universitas Indonesia, Jawa Barat, Indonesia, restu.octasila@gmail.com*
[2]Department of Biostatistic and Population, Universitas Indonesia, Jawa Barat, Indonesia, besral@ui.ac.id
[3]Faculty of Public Health, Universitas Indonesia, Jawa Barat, Indonesia, pujilaksmini@gmail.com
[4]Faculty of Public Health, Universitas Indonesia, Jawa Barat, Indonesia, sitidariyani82@gmail.com
[5]D3 Midwifery Study Program, Banten High School Of Health Sciences, Banten, Indonesia, darinzhr8@gmail.com

Corresponding Author:
Restu Octasila
Email:restu.octasila@gmail.com

Received: 09-04-2026
Revised: 18-04-2026
Accepted: 21-04-2026
Published: 30-04-2026

Abstract: Background: Introduction: Machine learning algorithms have shown promise in predicting the survival among neonates. Yet, current models are predominantly trained on heterogeneous populations and lack any clinically relevant subgroup discovery. The combined use of unsupervised clustering and predictive modelling could enhance prognostic accuracy and clinical utility. **Methods:** A systematic review and meta-analysis following PRISMA 2020 guidelines was performed, and it is registered in PROSPERO (CRD420261279521). A scoping review was conducted searching PubMed, Scopus, and EMBASE for articles from 2015 to 2025 on machine learning systems predicting infant survival. Rob's risk of bias was evaluated with probast. The AUROC values in the studies were assessed, and a meta-analysis using random effects was performed. **Results:** Twelve of the eligible studies met the inclusion criteria, and nine were pooled in the meta-analysis. AUROC values were reported between 0.811 and 0.952, suggesting overall strong discrimination performance with diverse algorithms such as logistic regression, random forests, neural networks, and ensemble models. Multimodal predictors integrating perinatal features, biomarkers, electronic health records information, and physiological measurements were able to maintain the predictive performance. Nevertheless, many studies constructed homogeneous population-based models without specific subphenotype characterization. The rating of risk of bias showed some concerns, especially about the external validation and risk of overfitting. **Conclusions:** Machine learning models are highly discriminative for neonatal survival prediction, but methodological limitations and population heterogeneity constitute major challenges. Integrating clustering within a predictive framework might either achieve better precision, interpretability, or translational utility. An empirical testing of subphenotype-based modeling approaches should be a high priority for future work.

Keywords: Neonatal Mortality; Infant Survival; Machine Learning; Cluster Analysis Prognosis.

INTRODUCTION

Congenital melanocytic nevi (CMN) are pigmented cutaneous lesions present at birth, resulting from the proliferation of neural-crest–derived melanocytes within the skin. Their clinical appearance ranges from small, well-circumscribed macules to large, geographically extensive plaques(1). The size-based classification is clinically significant, with *giant* congenital melanocytic nevi (GCMN) typically defined as lesions measuring more than 20 cm in projected adult size or covering more than 2% of the body surface area in newborns representing the most severe end of the spectrum. GCMN carries both cosmetic and psychosocial consequences for families, as well as important medical considerations, including risk of melanoma, neurocutaneous melanosis, and associated structural anomalies(2).

The bathing-trunk or garment-type distribution of GCMN, which involves the lower trunk, gluteal region, and lower limbs, is one of the more dramatic presentations. These lesions often coexist with multiple satellite nevi and may be associated with leptomeningeal melanocytosis(3). Nevertheless, the majority of affected neonates remain neurologically asymptomatic at birth. While the dermatological manifestations of GCMN are well known, the coexistence of additional congenital soft-tissue masses or vascular malformations is uncommon and less frequently reported in literature(4).

Vascular anomalies in neonates encompass a wide heterogeneity of lesions, ranging from high-flow arteriovenous malformations to benign low-flow venous or lymphatic malformations. Their clinical presentation varies widely depending on anatomical site, flow characteristics, and depth of involvement(5). Low-flow vascular malformations tend to be soft, compressible, and slow-growing, whereas solid, firm, progressively enlarging pedunculated masses at birth are unusual and may mimic lipomas, neurofibromas, hamartomas, teratomas, or fibrous tumors. When such masses coexist with GCMN, the diagnostic complexity increases, particularly in distinguishing between melanocytic proliferation and unrelated soft-tissue tumors(6).

Antenatal ultrasonography has significantly improved prenatal detection of structural fetal anomalies. However, small or superficially located cutaneous lesions particularly vascular malformations confined to the subcutaneous plane may remain undetected due to fetal position, limited contrast resolution, and technical constraints(7). Thus, even in pregnancies with normal mid-trimester anomaly scans, neonates may present with unexpected external masses at birth. Early postnatal imaging, including ultrasound with Doppler and MRI, plays a crucial role in delineating lesion extent, flow characteristics, tissue components, and relationship with underlying structures(8).

The simultaneous occurrence of a giant bathing-trunk nevus with a large pedunculated low-flow vascular mass is extremely rare and sparsely documented. Such presentations pose diagnostic, therapeutic, and psychosocial challenges for clinicians and caregivers. Surgical intervention becomes necessary when complications such as rapid enlargement, tension on overlying skin, ulceration, or cosmetic concerns arise(9). This case highlights the importance of meticulous clinical evaluation, detailed imaging, and coordinated multidisciplinary management in neonates presenting with complex congenital cutaneous and soft-tissue lesions.

METHODS

This research was conducted in compliance with the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) 2020 statement. This systematic review protocol is registered in PROSPERO (registration number: CRD42026127 9521). Some small amendments were made to the title of this document - no changes were made to the overall focus or content of this piece of research. By doing so, this review paper will first review and critically assess the existing scientific literature regarding the use of machine learning (ML) techniques for predicting infant survival/death, followed by evaluating the possibility of combining clustering–based analytical strategies with machine learning prediction models, thereby enhancing neonatal prognostic modeling. We conducted a systematic literature search using three databases (PubMed, Scopus, and EMBASE) to identify articles between January 1, 2015, and December 31, 2025, with English language restriction. The search strategy was constructed and included thesauri/controlled vocabulary (MeSH and Emtree) as well as free terms used focusing on newborn outcomes, infant survival/mortality–explicitly defined as “infant mortality”, “neonatal mortality” or “infant survival” - and any AKA related approaches: machine leaning techniques (“machine learning,” “artificial intelligence", "deep learning", "neural network", "random forest”) in line of risk strata fication such as clustering(“clustering”), unsupervised-learningor subphenotype – term alternative use for ("subphenotype").

The search from the databases provided a total of 2,596 articles (PubMed = 1,103; Scopus = 1,203, and EMBASE = 291). 2,494 articles were screened following removal of duplicates and exclusion of administrative records. Title and abstract were

screened by two independent reviewers to search for predictions of IM/S COMLS using machine learning. Overall, 2,338 articles were excluded due to non-generalizability or ML approach, and 36 otherwise; not meeting language and outcome definitions. Complete texts of 120 articles were then read to assess eligibility. A total of 108 articles were excluded for reasons including inadequate sample size, no follow-up studies, or inappropriate design (Figure 1). Ultimately, twelve studies fulfilled all the criteria for inclusion and were included in the qualitative synthesis. Out of 12 studies, the AUROC values were 9 for quantitative discrimination, or had enough data and could be included in the meta-analysis. The PRISMA 2020 flow diagram describes the study selection procedure (Fig. 1).

We included studies reporting quantitative performance measures (AUROC or AUC) on machine learning methods/models for survival/mortality prediction in infants published between 2015 and 2025, that were written in English. Editorial, comments, or news narrative reviews for discussion of background information, brief reports that did not clearly state the methodology, and were not applied ML approach, those studies had been excluded from our review. Data were extracted to a standardised form, which included the type of study, characteristics of the research population (size and description), number of predictor variables studied , types of machine learning algorithms used, validation methods used, and model performance outcomes. The process of full text extraction was performed independently by two reviewers, and consensus was reached for any disagreements. PROBAST (Prediction model Risk Of Bias Assessment Tool) was used for Quality Appraisal of the participants and predictors, outcome, and analysis domains. We both used narrative synthesis to summarise patterns of population heterogeneity and methodological variation, and quantitative synthesis of AUROCs from the nine studies that reported them to obtain an overall quantitative summary estimate of discriminative performance of machine learning models in predicting survival amongst neonates. Overall, the generalisation of our meta-analysis should be considered exploratory rather than confirmatory due to a number of the included studies being non-informative for full variance. A random-effects model was used to allow for anticipated clinical and methodological diversity between studies.

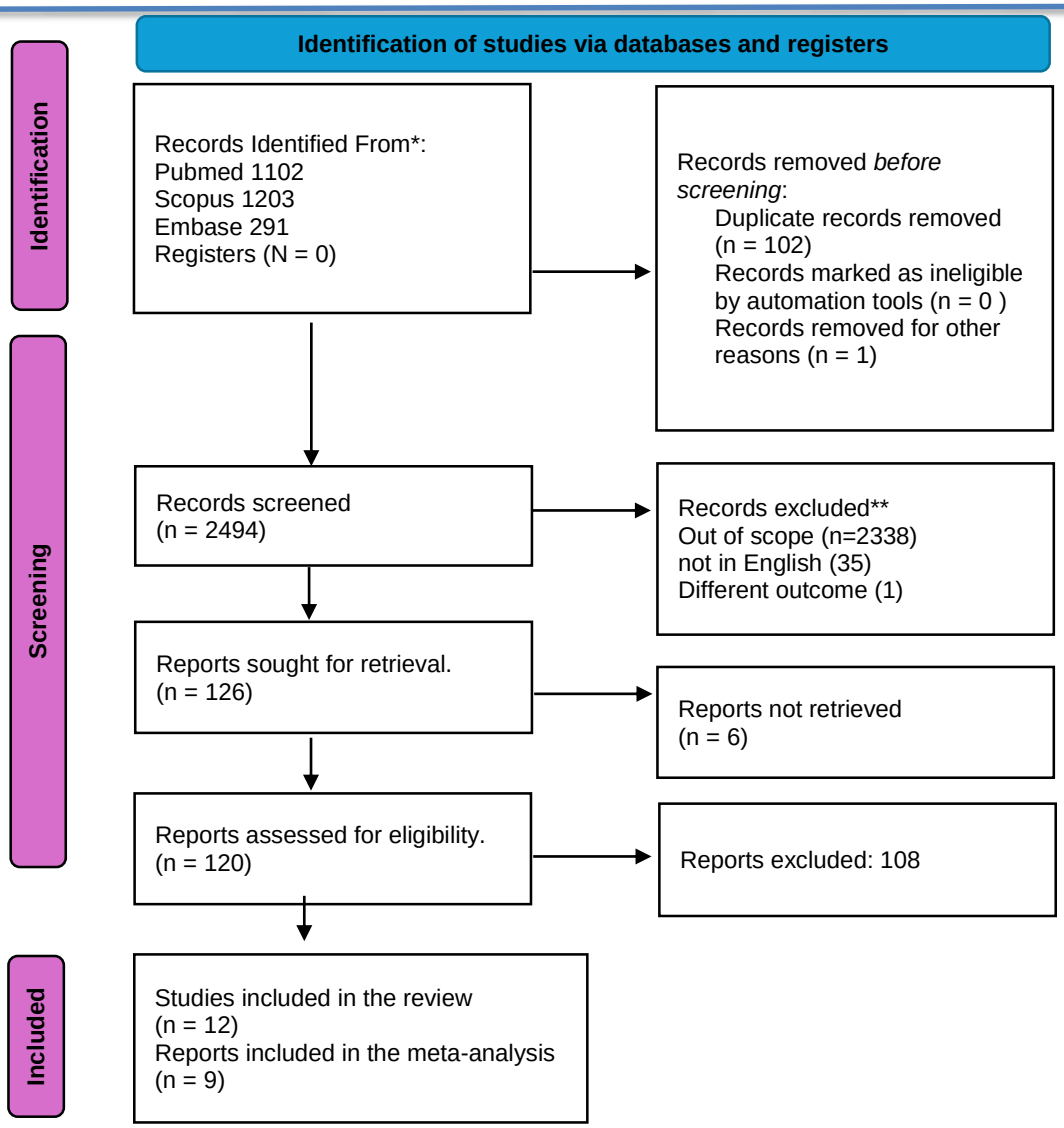

Figure 1. PRISMA 2020 flow diagram

RESULTS

The review identified 12 studies meeting the inclusion criteria, all of which were eligible for qualitative synthesis and had eciprocal reproduction data available in [9]. Discrimination performance of the machine learning models was consistently good to excellent across studies, with reported AUROC values from 0.81 to 0.97. [5–8,20,21]. However, all of these methods are relatively generic , and none have tried to bring together unsupervised clustering-based sub-phenotype identification.

Oh et al. were the two biomarker-based studies. [5] showed that lactate levels within the first 12 hours of life were a significant contributor to the tree-based models (AUC 0.811), indicating a high-risk metabolic phenotype existed in their population. These findings are also consistent with the literature on the added value of biomarkers to clinical prediction models [22].

Yang et al. [7] and Podda et al. [14], when comparing to only simpler perinatal factors (which included gestational age and birth weight) as dominant classifiers (NB GEOCS24, 25; AUC up to 0.90). Teh et al. vs Longitudinal method of Na et al. [21] (AUROC 0.932–0.973) and the phase adaptive approach proposed in [17] achieved comparable performance to Methods, but to our knowledge are not directly comparable. [6](td-AUC=0.838) indirectly supports the heterogeneity temporality in neonates.

1The EHR-based research work by Li et al. [8](AUROC 0.91) illustrates the advantage of using maternal and

neonatal data together to detect differences in risk not captured by traditional models. Brahma & Mukherjee [13] expose another factor, which are the social determinants of 6 Pop World JOURNAL. Therefore, they also have a link with infantile mortality (in a context of poverty). Their study provides an integrated view , not merely at the biological but also at the structural level of any heterogeneity. Do et al. compared various algorithms. [20]. Overall, while our focus is not on clustering, the empirical risk of biomarkers, phase of disease, or social determinants and natural history process all support this requirement for population segmentation prior to predictive model development [5–8,13–15,20,21]. A summary of the most relevant results is reported in the table below.

Table 1. Characteristics of Included Studies

No	Author (Year)	Country/Data Source	Study Design	Population (n)	Outcome	ML Method(s)	Best Performance
1	Oh et al. (2025)	Korea, single NICU	Retrospective cohort	168 very low birth weight infants	D7 & D30 mortality	RF, GBM, LightGBM	AUC 0.811 (GBM)
2	Smith et al. (2024)	USA/Canada PHN	Multicenter cohort	549 HLHS infants	5-year transplant-free survival	ML ensemble + SHAP	td-AUC 0.838
3	Yang et al. (2024)	Taiwan Neonatal Network	Nationwide cohort	7,471 VLBW infants	Early mortality	LR, NN	AUC 0.81–0.90
4	Li et al.	USA, MIMIC-III	Retrospective database	459 infants (37 deaths)	NICU survival	Random Forest	AUROC 0.91
5	Na et al. (2023)	Korean Neonatal Network	Nationwide cohort	15,790 very low birth weight infants	In- hospital mortality	MLP + ensemble	AUROC 0.932–0.973
6	Mfateneza et al. (2022)	Rwanda DHS	Cross-sectional	8,000+ records	Infant mortality	RF, SVM, DT	AUROC 0.842 (RF)
7	Leigh et al. (2022)	USA	Retrospective cohort	689 preterm infants	BPD-free survival	Ensemble ML	AUROC 0.921 / 0.899
8	Do et al. (2022)	Korean Neonatal Network	Nationwide cohort	7,472 very low birth weight infants	Mortality	LR, ANN, RF, SVM	AUROC 0.845 (ANN)
9	Brahma & Mukherjee (2022)	India DHS	National survey	100,000+ births	Neonatal & infa	LASSO, RF, Boosting	AUPRC > LR
10	Rinta-Koski et al. (2018)	Finland NICU	Cohort	598 very low birth weight infants	In- hospital mortality	Gaussian Process	AUC 0.948
11	Podda et al. (2018)	Italian Neonatal Network	Multicenter cohort	23,747 (development), 5,810 (test)	Survival prediction	Neural Network	AUROC 0.914
12	Chen et al. (2017)	Canada	Retrospective	48 infants	3-month survival	CART	Accuracy 83

Quality assessment was performed using the PROBAST tool across 4 domains (participants, predictors, outcome, analysis). Many studies were at low risk of bias in the areas of participant, predictor, and outcome domains, indicating appropriate targeting of population, definition of predictors, and specification of outcomes. But the fears were mostly an analytic worry. Five studies were rated to have an overall high risk of bias, mainly due to inadequate external validation or overfitting, and a lack of information on calibration and the binary outcome data. Several of the studies were classified to have unclear risk as a result of incomplete reporting on methodological details regarding model validation and hyperparameter tuning. Validation: Only two studies received a low risk of bias rating overall (robust validation, clear reporting). Risk of Bias for the included studies is shown in the table.

Table 2. Risk of Bias Assessment Using PROBAST

No	Study	Participants	Predictors	Outcome	Analysis	Overall Risk
1	Oh et al. (2025)	Low	Low	Low	High	High
2	Smith et al. (2024)	Low	Low	Low	Unclear	Unclear

3	Yang et al. (2024)	Low	Low	Low	Unclear	Unclear
4	Li et al. (2024)	Low	Low	Low	High	High
5	Na et al. (2023)	Low	Low	Low	Low	Low
6	Mfateneza et al. (2022)	Unclear	Low	Low	High	High
7	Leigh et al. (2022)	Low	Low	Low	Unclear	Unclear
8	Do et al. (2022)	Low	Low	Low	Low	Low
9	Brahma & Mukherjee (2022)	Unclear	Low	Low	High	High
10	Rinta-Koski et al. (2018)	Low	Low	Low	Unclear	Unclear
11	Podda et al. (2018)	Low	Low	Low	Unclear	Unclear
12	Chen et al. (2017)	Unclear	Low	Low	High	High

Prediction accuracy of infant survival by machine learning models is good or near perfect because the pooled AUROC scores are high, as reported in the current meta-analysis. Across AUROCs for individual outcomes (0.811–0.952), a range of algorithmic approaches (e.g., logistic regression to ensemble neural networks) may be capable of excellent discrimination in the newborn population. However, high heterogeneity in sample size of the studies and differences in study design and population do not permit generaliz ation on this good diagnostic performance.

There was no obvious asymmetry in the funnel plot, indicating that there is no strong evidence of publication bias caused by small-study effects. However, the interpretation should be cautious as few studies and significant methodological heterogeneity existed. Interpretation of causality regarding asymmetrical distribution patterns. With regard to models intended for prediction, differences in outcome definitions may be present, as well as the duration between baseline and follow-up (frequency imbalance) or a non-random selection of predictors to enter into the model

Table 3. Meta-analysis Dataset with Statistical Transformation

Study		**Sample Size (n)**	**AUROC**	**logit (AUROC)**	**Approx. SE(AUROC)**	**Variance (SE^2)**
Oh et al. (2025)		168	0.811	1.4565	0.03021	0.0009124
Smith et al. (2024)		549	0.838	1.6434	0.01573	0.0002473
Yang et al.(2024)		7471	0.855	1.7744	0.00407	0.0000166
Li et al. (2024)		459	0.910	2.3136	0.01336	0.0001784
Na et al. (2023)		15,790	0.952	2.9874	0.00170	0.0000029
Mfateneza et al. (2022)		8000	0.842	1.6732	0.00408	0.0000166
Leigh et al. (2022)		689	0.910	2.3136	0.01090	0.0001189
Do et al. (2022)		7472	0.845	1.6959	0.00419	0.0000175
Rinta-Koski et al. (2018)		598	0.948	2.9031	0.00908	0.0000824
Podda et al. (2018)		23747	0.914	2.3635	0.00182	0.0000033

The MID (moderate) AUROCs of the meta-analysis give evidence to conclude that performance on the prediction of survival in newborns using machine learning models is acceptable and mostly good to very good. The broad variation of individual AUROCs (0.811–0.952), I believe, is an illustration that there are algorithm designs ranging from logistic regression to ensemble neural networks, that could have very good classification performance in this population of newborns. However, the very heterogeneous sample size between studies and study designs makes the perfect performance not interpretable in general.

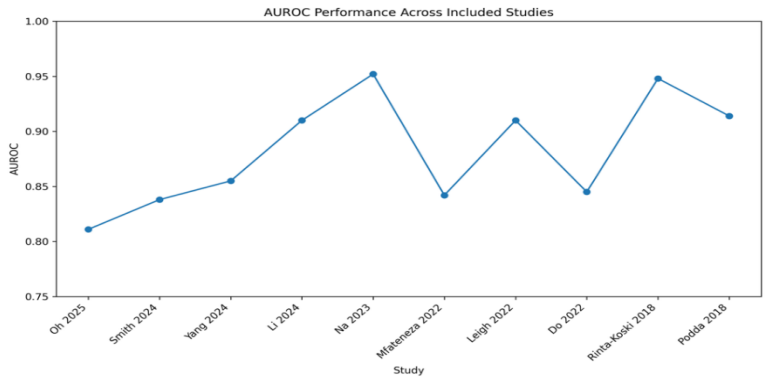

Fig 2. AUROC diagram

Studies with a larger sample size from national networks that demonstrate consistently better performance (ie, stable point estimates and small SEs) will be given more weight in random effects models. In contrast, estimates from small samples are more variable and have broader confidence intervals , indicating that performance estimates could be unstable. This observation is consistent with a statistical theory that discrimination estimates are plagued by the work of overfitting for small sample sizes, and in particular when the number of predictors is not large compared to the number of events.

This can be argued from the perspective of risk bias PROBAST. Prospective studies show most studies to be low risk in participant and outcome, methodological weaknesses are more frequently identified in the analysis domain. Some research lacks a model calibration, external validation , or a more in-depth explanation of how they addressed the missing data and class imbalance. This is important, as high discrimination (AUROC) does not necessarily imply good generalization to external populations. Hence, although the pool AUROC suggests encouraging performance, performance overestimation is not completely ruled out.

No evidence of obvious asymmetry was detected in funnel plot analysis, revealing that there is no strong publication bias of small-study effects. However, we ought to be careful to interpret the effect as the number of included studies was small and there appeared to be considerable methodological diversity. In the context of prediction models, variation in outcome definitions, follow-up time, choice of predictors, and validation approach can lead to asymmetrical distributions despite the absence of evidence for systematic publication bias.

DISCUSSION

This systematic review and meta-analysis highlights that ML models consistently achieve a high level of discrimination for neonatal survival prediction in a variety of datasets and clinical environments. Nevertheless, the evidence also indicates significant heterogeneity in predictor choice, statistical modeling, and validation approaches. Most importantly, the majority of published models have been trained using population-based methods without direct recognition of clinically relevant subphenotypes, which could indicate that high predictive accuracy itself may not capture completely model stability or transportability in diverse neonatal populations.

High-performance discrimination in neonatal prediction models is in line with publications from clinical machine learning , generally, where ensemble models, neural networks, and tree-based methods were shown to have a high predictive accuracy in datasets related to healthcare [1–4]. Many of the studies in this review also revealed that core perinatal factors , such as gestational age, birth weight, and APGAR score, remained crucial predictors, only for a limited time to come [5,7,14,22]. Biomarker model of risk, including the role of early-life lactate levels, has also been reported to improve mortality prediction in preterm infants [5]. The inclusion of maternal features along with electronic health records (EHRs) and long-term physiologic data illustrates the rising prevalence of multimodal prediction modeling in clinical medicine [2,8].

Answering the Research Questions

The current review aimed to answer three pre-defined research questions on population heterogeneity, model specialization, and the predictive determinants of neonatal survival modelling.

RQ1: Are clinically relevant subgroups of neonates inducible by an unsupervised clustering approach?
The results of this review have uncovered significant biological, clinical, temporal, and social variation among neonatal populations in terms of metabolic biomarkers, treatment-phase dynamics, physiological monitoring signals, and structural determinants of

health [5,6,13,21]. Although clustering-based modeling workflows were seldom directly applied in the neonatal studies investigated, robust evidence from other medical disciplines suggests that unsupervised clustering can effectively be used to identify clinically relevant subphenotypes associated with differential outcomes and treatment responses [9,10,16]. Collectively, available evidence supports the conceptual soundness and potential clinical relevance of clustering-based subphenotype identification for use as a preparatory step in neonatal prognostic model-building.

RQ2: Are subphenotype-specific prediction models superior to generic population-wide models?
The included literature lacked any direct comparison between clustering-based subphenotype models and general neonatal prediction models. Collecting indirect evidence, however, suggests that predictive models built in more homogenous clinical cohorts (e.g., disease-specific populations or longitudinal phase-stratified cohorts) have better stability and discrimination properties [6,21]. On the other hand, models that were trained on more diverse neonatal cohorts combined across sites during these trials may demonstrate good discrimination within yet have relatively high risk for overfitting and inferior external generalizability [18,19]. These data imply that population stratification and subphenotype-specific modeling could enhance predictive performance, while prospective confirmatory studies testing these approaches are yet to be conducted.
RQ3: What are the most important predictors, algorithms, and modeling attributes for neonatal survival prediction?

Gestational age, birth weight, Apgar scores (at 5 min), metabolic biomarkers such as lactate, maternal clinical factors, and physiological data from the electronic health record were among the most commonly important predictors across studies [5,7,8,14,22]. There was an indication in population-based data that a similar risk associated with mortality could be caused by factors related to socioeconomic status [13]. With respect to modeling methodologies, several ensemble tree-based models, neural networks, and mixed types of machine learning architectures often achieved excellent predictive performance [5,8,21]. However, more parsimonious models like logistic regression sometimes can perform comparably in structured or fairly homogeneous populations [7,20]. Model interpretability tools and techniques, such as SHAP-based explanations, were highlighted as crucial for improving clinical transparency and promoting responsible use of machine learning prediction systems [11].

Together, these results suggest that the 98 For IUPS (https://www.iups.org/) meeting Abstracts evidence base for prediction of neonatal survival will be best served by well-performing core clinical predictors in conjunction with contextually appropriate algorithm choice and modeling frameworks amenable to population heterogeneity.

Recommendations for Future Research
Further work is required to validate external generalization of the clustering-then-predict framework on larger sample , multicenter datasets via empirical validation. Validation across population and healthcare environments will be necessary to determine the stability of the generated spp and regulation performance consistency. More robust methods should also be applied to handle the imbalanced class and missing data in future studies, as neonatal mortality is a relatively rare event and clinical information is usually incompletely collected. Methods, such as over-sampling followed by under-sampling, cost-sensitive or probabilistic learning, might be applied in order to improve the predictive capability.

In addition, interpretability should be designed carefully from the beginning rather than being an afterthought at the end of analysis. The utilization of tools such as SHAP or other explainable AI modalities is expected to improve clinician confidence around, and integration of predictive model output into clinical practice and health policy. Prototyping of real-time decision support systems leveraging physiologic data is another field that is growing, especially to identify early deterioration and for resource utilization in the NICU. Nevertheless, the rollout of such systems needs to be evaluated prospectively, and the impact on clinical endpoints (mortality, morbidity) as well as service delivery needs to be tested. If this research agenda is pursued, it could render a historical focus on an integrative subphenotype-based approach more than just a conceptual model and thereby transform the construct into an implementable, adaptable, and clinically meaningful system of prognostication in neonatal medicine.

CONCLUSION

The current review aimed to answer three pre-defined research questions on population heterogeneity, model specialization, and the predictive determinants of neonatal survival modelling.
RQ1: Are clinically relevant subgroups of neonates inducible by an unsupervised clustering approach?
The results of this review have uncovered significant biological, clinical, temporal, and social variation among neonatal populations in terms of metabolic biomarkers, treatment-phase dynamics, physiological monitoring signals, and structural determinants of health [5,6,13,21].

Although clustering-based modeling workflows were seldom directly applied in the neonatal studies investigated, robust evidence from other medical disciplines suggests that unsupervised clustering can effectively be used to identify clinically relevant subphenotypes associated with differential outcomes and treatment responses [9,10,16]. Collectively, available evidence supports the conceptual soundness and potential clinical relevance of clustering-based subphenotype identification for use as a preparatory step in neonatal prognostic model-building.

RQ2: Are subphenotype-specific prediction models superior to generic population-wide models?

The included literature lacked any direct comparison between clustering-based subphenotype models and general neonatal prediction models. Collecting indirect evidence, however, suggests that predictive models built in more homogenous clinical cohorts (e.g., disease-specific populations or longitudinal phase-stratified cohorts) have better stability and discrimination properties [6,21]. On the other hand, models that were trained on more diverse neonatal cohorts combined across sites during these trials may demonstrate good discrimination within yet have relatively high risk for overfitting and inferior external generalizability [18,19]. These data imply that population stratification and subphenotype-specific modeling could enhance predictive performance, while prospective confirmatory studies testing these approaches are yet to be conducted.

RQ3: What are the most important predictors, algorithms, and modeling attributes for neonatal survival prediction?

Gestational age, birth weight, Apgar scores (at 5 min), metabolic biomarkers such as lactate, maternal clinical factors, and physiological data from the electronic health record were among the most commonly important predictors across studies [5,7,8,14,22]. There was an indication in population-based data that a similar risk associated with mortality could be caused by factors related to socioeconomic status [13]. With respect to modeling methodologies, several ensemble tree-based models, neural networks, and mixed types of machine learning architectures often achieved excellent predictive performance [5,8,21]. However, more parsimonious models like logistic regression sometimes can perform comparably in structured or fairly homogeneous populations [7,20]. Model interpretability tools and techniques, such as SHAP-based explanations, were highlighted as crucial for improving clinical transparency and promoting responsible use of machine learning prediction systems [11].

Together, these results suggest that the 98 For IUPS (https://www.iups.org/) meeting Abstracts evidence base for prediction of neonatal survival will be best served by well-performing core clinical predictors in conjunction with contextually appropriate algorithm choice and modeling frameworks amenable to population heterogeneity.

BIBLIOGRAPHY

1. Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. N Engl J Med 2019;380:1347–58. https://doi.org/10.1056/NEJMra1814259. [DOI].
2. Beam AL, Kohane IS. Big Data and Machine Learning in Health Care. JAMA 2018;319:1317. https://doi.org/10.1001/jama.2017.18391. [DOI].
3. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019;25:44–56. https://doi.org/10.1038/s41591-018-0300-7. [DOI].
4. Liu Y, Chen P-HC, Krause J, Peng L. How to Read Articles That Use Machine Learning : Users' Guides to the Medical Literature. JAMA 2019;322:1806. https://doi.org/10.1001/jama.2019.16489. [DOI].
5. Oh M-Y, Kim S, Kim M, Seo YM, Yum SK. Machine-learning-based evaluation of the usefulness of lactate for predicting neonatal mortality in preterm infants. Pediatr Neonatol 2025;66:310–5. https://doi.org/10.1016/j.pedneo.2024.09.003. [DOI].
6. Smith AH, Gray GM, Ashfaq A, Asante-Korang A, Rehman MA, Ahumada LM. Using machine learning to predict five-year transplant-free survival among infants with hypoplastic left heart syndrome. Sci Rep 2024;14:4512. https://doi.org/10.1038/s41598-024-55285-1. [DOI].
7. MASHRAFI SS AL, Tafakori L, Abdollahian M. Predicting Early Neonatal Mortality using Machine Learning Models 2025. https://doi.org/10.21203/rs.3.rs-7016070/v1. [DOI][PubMed].
8. Li A, Mullin S, Elkin PL. Improving Prediction of Survival for Extremely Premature Infants Born at 23 to 29 Weeks Gestational Age in the Neonatal Intensive Care Unit: Development and Evaluation of Machine Learning Models. JMIR Med Informatics 2024;12:e42271. https://doi.org/10.2196/42271. [DOI][PubMed].
9. Calfee CS, Delucchi K, Parsons PE, Thompson BT, Ware LB, Matthay MA. Subphenotypes in acute respiratory distress syndrome: latent class analysis of data from two randomised controlled trials. Lancet Respir Med 2014;2:611–20. https://doi.org/10.1016/S2213-2600(14)70097-9. [DOI].
10. Seymour CW, Kennedy JN, Wang S, Chang C-

CH, Elliott CF, Xu Z, et al. Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis. JAMA 2019;321:2003. https://doi.org/10.1001/jama.2019.5791. [DOI].
11. Lundberg S, Lee S-I. A Unified Approach to Interpreting Model Predictions 2017. [ArXiv].
12. Das BB, Deshpande SR, Choudhry S, Perumal G. Predicting 1-Year Mortality After Pediatric Heart Transplantation Using Machine Learning. JACC Adv 2026;5:102422. https://doi.org/10.1016/j.jacadv.2025.102422. [DOI] [PubMed].
13. Brahma D, Mukherjee D. Early warning signs: targeting neonatal and infant mortality using machine learning. Appl Econ 2022;54:57–74. https://doi.org/10.1080/00036846.2021.1958141. [DOI].
14. Podda M, Bacciu D, Micheli A, Bellù R, Placidi G, Gagliardi L. A machine learning approach to estimating preterm infants survival: development of the Preterm Infants Survival Assessment (PISA) predictor. Sci Rep 2018;8:13743. https://doi.org/10.1038/s41598-018-31920-6. [DOI].
15. Rinta-Koski O-P, Särkkä S, Hollmén J, Leskinen M, Andersson S. Gaussian process classification for prediction of in-hospital mortality among preterm infants. Neurocomputing 2018;298:134–41. https://doi.org/10.1016/j.neucom.2017.12.064. [DOI].
16. Shah SJ, Katz DH, Selvaraj S, Burke MA, Yancy CW, Gheorghiade M, et al. Phenomapping for Novel Classification of Heart Failure With Preserved Ejection Fraction. Circulation 2015;131:269–79. https://doi.org/10.1161/CIRCULATIONAHA.114.010637. [DOI].
17. Tataranno ML, Vijlbrief DC, Dudink J, Benders MJNL. Precision Medicine in Neonates: A Tailored Approach to Neonatal Brain Injury. Front Pediatr 2021;9. https://doi.org/10.3389/fped.2021.634092. [DOI] [PubMed].
18. Collins GS, Reitsma JB, Altman DG, Moons K. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. BMC Med 2015;13:1. https://doi.org/10.1186/s12916-014-0241-z. [DOI].
19. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. Ann Intern Med 2019;170:51–8. https://doi.org/10.7326/M18-1376. [DOI].
20. Sullivan BA, Moreira AG, McAdams RM, Knake LA, Husain A, Qiu J, et al. Comparing machine learning techniques for neonatal mortality prediction: insights from a modeling competition. Pediatr Res 2025;98:405–11. https://doi.org/10.1038/s41390-024-03773-5. [DOI].
21. Na JY, Jung D, Cha JH, Kim D, Son J, Hwang JK, et al. Learning-Based Longitudinal Prediction Models for Mortality Risk in Very-Low-Birth-Weight Infants: A Nationwide Cohort Study. Neonatology 2023;120:652–60. https://doi.org/10.1159/000530738. [DOI].
22. Riley RD, Hayden JA, Steyerberg EW, Moons KGM, Abrams K, Kyzas PA, et al. Prognosis Research Strategy (PROGRESS) 2: Prognostic Factor Research. PLoS Med 2013;10:e1001380. https://doi.org/10.1371/journal.pmed.1001380. [DOI].

