



Paper Type: Original Article

Embedding Neonatal Mortality Prediction into Perinatal Workflows: A Machine-Learning Approach from the IMaN Registry

Mobina Batebi¹, Mahsa Qeytasi², Moslem Habibi^{1*} , Abbas Habibelahi³ 

¹ Department of Industrial Engineering, Sharif University of Technology, Tehran, Iran; mobina.batebi@ie.sharif.edu; mhabibi@sharif.edu.

² Department of Industrial Engineering, Faculty of Industrial and systems engineering, Tarbiat Modares University, Tehran, Iran; mahsa.qeytasi@outlook.com.

³ Tehran University of Medical Sciences, Tehran, Iran; ahabibelahi@yahoo.com.

Citation:

Received: 20 June 2025

Revised: 25 November 2025

Accepted: 15 February 2026

Batebi, M., Qeytasi, M., Habibi, M., & Habibelahi, A. (2026). Embedding neonatal mortality prediction into perinatal workflows: A machine-learning approach from the IMaN registry. *Annals of Healthcare Systems Engineering*, 3(2), 124-133.


Abstract


Neonatal mortality remains a major challenge in resource-limited settings, where delayed recognition of high-risk cases and inconsistent clinical decisions hinder timely and targeted interventions, which are essential for reducing preventable deaths. In response, this study developed and evaluated Machine-Learning (ML) models to predict neonatal death using maternal and neonatal features collected both before and after delivery. To this end, guided by the CRISP-DM data-mining framework, we analyzed a dataset of 7,214 births (5,000 survivors and 2,214 deaths) from 2021–2022, derived from routinely collected records in the Iranian Maternal and Neonatal (IMaN) registry. As a result, among the data-mining models—Random Forest, Extreme Gradient Boosting (XGBoost), and Support Vector Machine—trained with imbalance-sensitive techniques, XGBoost achieved the best performance (ROC-AUC = 0.967, PR-AUC = 0.940). Feature importance analysis identified gestational age (importance = 0.179) and birth weight (0.109) as the dominant predictors, followed by nervous system malformations (0.035), musculoskeletal malformations (0.033), high-risk delivery indicators (0.032), and other congenital malformations (0.031). The contribution of this study is in twofolds, first, these findings demonstrate that accurate, real-time prediction of neonatal mortality is achievable. Second, beyond a prognostic tool, the final model can serve as an operational lever within neonatal services; when embedded into a clinical Decision Support System (DSS), it can enhance early risk detection, improve triage accuracy, facilitate timely NICU preparedness, and strengthen overall process reliability and system performance in resource-limited care settings.

Keywords: Neonatal mortality, Improving neonatal services system, Machine learning, Clinical decision support system, Prediction of neonatal mortality.

1 | Introduction

Neonatal mortality, defined as the death of an infant within the first 28 days of life, remains a major global public health concern. In 2023, an estimated 2.3 million children died during the first month of life

 Corresponding Author: mhabibi@sharif.edu

 <https://doi.org/10.22105/ahse.v3i2.64>



Licensee System Analytics. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0>).

worldwide¹. The Neonatal Mortality Rate (NMR) is widely recognized as a core indicator for assessing the performance and equity of national health systems [1], [2]. Recent estimates for the Islamic Republic of Iran indicate an NMR of 8 deaths per 1,000 live births in 2022, compared with 3.3 deaths per 1,000 live births in the European Union—approximately 59% lower than Iran’s rate [3], [4]. Another study marked provincial variation within Iran, with NMR values ranging from 3.1 (2.6–3.7) to 10.0 (9.2–10.8) deaths per 1,000 live births. These disparities underscore the need for predictive, preventive, and system-level approaches that can both identify high-risk neonates and inform targeted improvements in care delivery [5].

Evidence from Iran and other middle-income settings suggests that higher neonatal mortality is closely associated with contextual and organizational factors. Provinces with lower socioeconomic development, limited access to advanced NICU services, and lower levels of maternal education tend to exhibit significantly higher mortality rates [6]. Managerial gaps—such as constrained neonatal service capacity, suboptimal use of surfactant therapy, and fragmented referral pathways—contribute to avoidable adverse outcomes [7]. At the micro-process level, workflow variability, limited decision-support tools, and the absence of integrated dashboards or optimized Electronic Health Record (EHR) functions hinder timely recognition and management of critical neonatal conditions, including necrotizing enterocolitis and severe respiratory distress [8]. Collectively, these findings indicate that beyond clinical causes alone, process inefficiencies and delays in decision-making play a crucial role in determining neonatal outcomes.

Accordingly, interest has increased in applying Artificial Intelligence (AI) and Machine-Learning (ML) approaches to strengthen neonatal care, particularly in contexts where conventional quality-improvement strategies have shown limited impact when used alone [9]. These methods offer a means to overcome existing challenges by facilitating the early detection of high-risk newborns and enabling timely clinical responses. This study seeks to develop and evaluate ML models for predicting neonatal mortality in Iran using routinely collected data from the Iranian Maternal and Neonatal (IMaN) registry. In addition to examining predictive performance, the study positions neonatal mortality modeling as an operational tool to enhance process reliability, guide resource allocation, and improve overall system efficiency—consistent with modern priorities in healthcare process improvement and operational excellence.

In this domain, prior research has shown strong potential for ML-based neonatal outcome prediction. One study reported that Extreme Gradient Boosting (XGBoost) achieved excellent discriminatory performance ($AUC \approx 0.97$), successfully identifying over 90% of neonatal deaths within the highest 5% of predicted risk scores [10]. Likewise, DeepSurv-based survival analysis demonstrated an AUC of 0.95, outperforming Cox proportional hazards models and random survival forests in forecasting mortality among very low birth weight infants [11]. Nationwide investigations of very low birth weight preterm neonates have similarly documented robust performance for ML models such as random forest and XGBoost, with AUROC values ranging from 0.81 to 0.90 for predicting early mortality and severe intraventricular hemorrhage [12]. A Korean neonatal network study further found that random forest, artificial neural networks, and logistic regression achieved comparable predictive power ($AUROC \approx 0.84$), whereas support vector machines performed less effectively [13]. Expanding beyond general mortality, an XGBoost model developed for infants with congenital diaphragmatic hernia yielded an AUC of 0.87, underscoring the broader utility of ML across heterogeneous neonatal conditions [14].

Within Iran, notable progress has also been made. One study utilized more than 1.4 million IMaN registry records to classify live births and various stillbirth categories using a stacked ensemble of decision-tree, gradient-boosting, logistic-regression, random-forest, and SVM models; the two-stage architecture achieved high accuracy and AUC values around 0.90, outperforming individual algorithms [15]. Another investigation

¹ <https://www.who.int/>

applied deep learning to over 113,000 NICU admissions, demonstrating that a Convolutional Neural Network (CNN)—constructed by converting numerical features into image-like matrices—outperformed conventional ML classifiers and achieved 98% accuracy, emphasizing the promise of deep learning for neonatal mortality prediction [16]. Complementary work has employed Multi-Criteria Decision-Making (MCDM) methods to prioritize systemic neonatal challenges, providing structured guidance for optimizing resource allocation and policy development [17].

Despite these advances, several gaps remain. Many neonatal ML studies focus primarily on predictive performance (e.g., AUC) without explicitly addressing how model outputs are integrated into operational processes, triage rules, or resource planning—particularly in resource-limited settings such as parts of Iran. Among existing studies, the closest to our work is [16] which has also used IMaN registry to predict neonatal mortality. This study distinguishes itself from the aforementioned study through several key methodological and conceptual advances. While both leverage the IMaN registry, the prior study employs a CNN on post-birth NICU data and uses a custom folding technique for data balancing. In contrast, we implement a dual-strategy imbalance-aware pipeline that combines strategic under-sampling of the majority class with algorithm-level class weighting, preserving the natural data distribution while enhancing sensitivity to neonatal deaths—a critical consideration for clinical deployment. Furthermore, whereas prior work evaluates models primarily via accuracy and F-score, we optimize and assess performance using clinically meaningful metrics such as PR-AUC, F1-score, and balanced accuracy, and incorporate threshold tuning to refine model outputs for improved precision-recall trade-offs. Our approach focuses on interpretable models like XGBoost applied to information, enabling point-of-care risk assessment and early intervention. Methodologically, our framework is explicitly designed for seamless integration into clinical workflows, emphasizing triage support, NICU preparedness, and resource allocation in resource-limited settings.

This study presents five key contributions. First, it is the inaugural research to construct a neonatal mortality prediction model using the national Iranian IMaN registry, encompassing 7,114 births and 163 clinically validated features—notable for its comprehensiveness compared to earlier studies, which utilized a significantly smaller set of predictors. Methodologically, it introduces a novel dual-strategy, imbalance-aware framework that integrates the retention of all neonatal deaths, controlled sampling of survivors, model-level class weighting, Average Precision (AP)-based hyperparameter tuning, and F1-score-based threshold optimization into a complete pipeline. Theoretically, it develops a clinically interpretable risk structure linking maternal, obstetric, and neonatal characteristics to mortality, validated through pediatric expert consultation. Operationally, it formally frames ML prediction as a lever for enhancing process reliability, NICU preparedness, triage accuracy, and workflow standardization in resource-limited clinical settings. Finally, in terms of clinical insight, it identifies a dominant set of predictors—including gestational age, birth weight, congenital anomalies, and intrapartum risks—that collectively account for over 55% of the model's feature importance, providing actionable targets for frontline clinical intervention.

2 | Methods

This study utilized a retrospective observational design using routinely collected data from the IMaN registry for the year 1400 (2021–2022). The IMaN registry collects standardized maternal, obstetric, and neonatal information from hospitals and maternity centers all over Iran. After screening all available birth records, a total of 7,114 neonates were included in the analysis, consisting of 2,114 neonatal deaths and 5,000 survivors. Neonatal death—defined as death within the first 28 days after birth—was considered the primary outcome. In the IMaN registry, outcomes are recorded in the variable OutN. Based on this variable, a binary variable for neonatal death was created, where a value of 1 indicated neonatal death (OutN = 4) and a value of 0 indicated survival beyond the neonatal period. This binary variable served as the target label for all predictive modeling procedures. During the initial data inspection and variable screening, all variables available in the raw registry were systematically evaluated.

During the initial data inspection and variable screening, all variables available in the raw registry were systematically evaluated. Variables with $\geq 80\%$ missingness were excluded from further analysis. Following this step, a curated list of clinically relevant variables—representing maternal characteristics, obstetric history, pregnancy risk factors, labor characteristics, and newborn parameters—was retained. In total, 163 candidate features were included for model development, covering both pre-delivery and post-delivery information related to the mother and the neonate. To ensure clinical validity, feature selection was performed in close consultation with a pediatric specialist, and variables lacking clinical relevance or interpretability were removed based on expert judgment.

2.1 | Handling Missing Data and Data Encoding

The retained variables exhibited minimal missingness (generally $< 5\%$), reflecting the high completeness of key clinical fields in the registry. To preserve sample size and avoid bias, missing values were imputed rather than excluded. Median imputation was applied to numerical variables (e.g., maternal age, parity), while the most frequent category was used for categorical variables (e.g., maternal education, mode of delivery). All imputation procedures were fitted exclusively on the training set and subsequently applied to the test set to prevent information leakage.

Predictor variables were categorized as numeric/Boolean or categorical. Numeric and Boolean features captured continuous or binary clinical measures, whereas categorical features represented nominal descriptors such as maternal education, occupation, insurance status, and delivery mode. Categorical variables were transformed using one-hot encoding, and all preprocessing steps—imputation and encoding—were integrated into a unified scikit-learn pipeline via a ColumnTransformer. This structure ensured consistent and leakage-free preprocessing across training and test subsets.

2.2 | Final Analytic Dataset

After cleaning and imputation, the final dataset included $N = 7,114$ neonates: 2,114 neonatal deaths (29.7%) and 5,000 survivors (70.3%). This distribution mirrors a moderately imbalanced, where positive cases (deaths) are underrepresented but sufficient for supervised learning. Descriptive statistics revealed maternal ages ranging from 13–57 years (mean = 29.9, SD = 6.4). Among surviving neonates, 45.3% were delivered vaginally and 54.7% by cesarean section. In contrast, neonatal deaths occurred more frequently following vaginal delivery (60.3%) compared with cesarean section (39.7%). This pattern suggests that mode of delivery may be associated with neonatal outcome, potentially reflecting underlying clinical indications, maternal risk profiles, or complications prompting the choice of delivery method (*Fig. 1*). *Fig. 2* demonstrates a clear inverse relationship between gestational age and mortality risk. Neonates in the earliest gestational age category (< 28 weeks) experienced the highest mortality rate at approximately, with rates progressively declining across categories to the lowest mortality in the ≥ 37 weeks group. This strong physiological gradient validates the fundamental biological plausibility of our predictive features and highlights gestational age as a dominant risk factor in neonatal mortality prediction. A total of 25 variables were included in the analysis to examine their interrelationships and potential associations with neonatal mortality. A correlation matrix was constructed to visualize the degree and direction of relationships among these variables (*Fig. 3*). This analysis helped identify clusters of correlated predictors and provided insights into how maternal, fetal, and perinatal characteristics jointly contribute to neonatal outcomes.

2.3 | Train–Test Split

The dataset was randomly partitioned into an 80% training set ($n = 5,691$) and a 20% independent test set ($n = 1,423$), stratified by the outcome variable (stratify=y). This ensured proportional representation of classes ($\approx 30\%$ deaths, 70% survivors) in both subsets, facilitating unbiased evaluation.

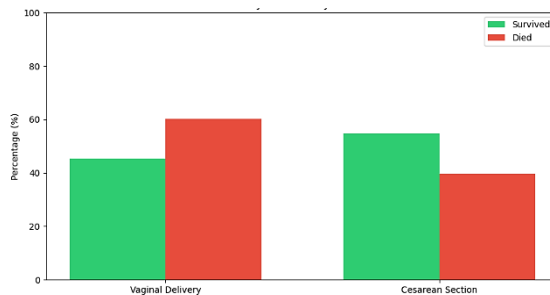


Fig. 1. Mode of delivery stratified by neonatal outcome.

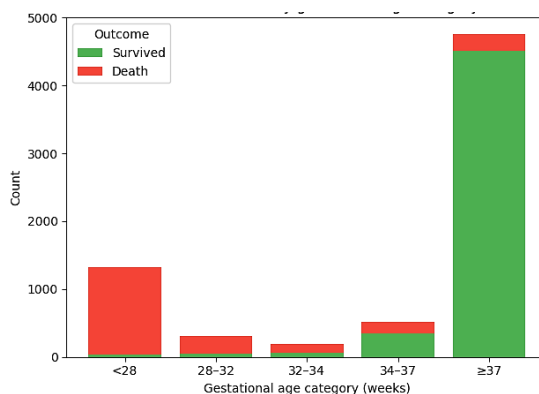


Fig. 2. NMR by gestational age.

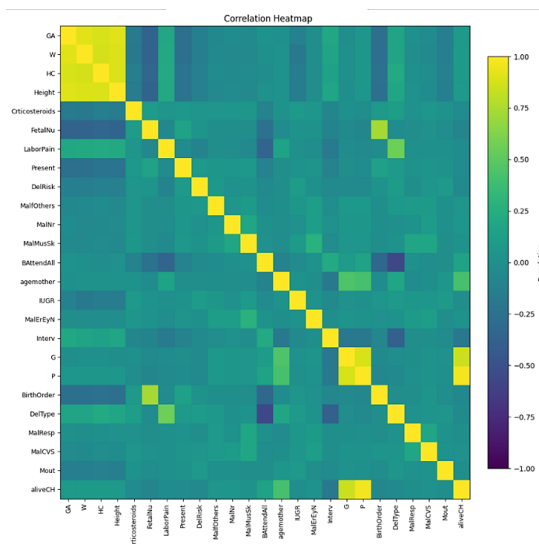


Fig. 3. Correlation heatmap of the 25 selected features.

2.4 | Handling Class Imbalance

Neonatal mortality prediction presents a class-imbalance challenge, as adverse outcomes constitute a minority of births. In this study, neonatal deaths accounted for 29.7% of the dataset (2,114 deaths and 5,000 survivors), creating a moderately imbalanced classification problem. To prevent bias toward the majority class and preserve sensitivity to high-risk neonates, several complementary imbalance-handling strategies were implemented. First, all neonatal death cases were retained, while a controlled random sample of surviving neonates was selected, avoiding synthetic oversampling and preserving the natural data structure.

Second, stratification was applied throughout the workflow—during train–test splitting and across all cross-validation folds—to maintain consistent class proportions and ensure unbiased performance estimation. Third, each model incorporated algorithm-specific class-balancing mechanisms. Random Forest used class-weighted learning to increase the penalty for misclassifying deaths, enhancing minority-class responsiveness. XGBoost employed positive-class weighting aligned with the survivor-to-death ratio to strengthen gradient updates for high-risk cases. For SVM, class-weighted margin optimization imposed higher costs on errors involving neonatal deaths, improving minority-class detection despite the model’s non-probabilistic nature. Together, these strategies provided a coherent imbalance-aware framework that supported reliable and clinically meaningful prediction of neonatal mortality.

2.5 | Optimization Using Imbalance-Sensitive Metrics

Because the dataset was highly imbalanced, conventional metrics such as accuracy and ROC-AUC were not reliable indicators of minority-class performance. Model hyperparameters were therefore optimized using the AP score derived from the precision–recall curve and averaged across stratified cross-validation folds. AP offers a more discriminative assessment for rare outcomes by jointly reflecting false positives through precision and false negatives through recall, and by summarizing performance across all possible thresholds. This approach ensured that model tuning prioritized the accurate detection of neonatal deaths.

2.6 | Threshold Optimization for Probabilistic Models

Both the Random Forest and XGBoost models generated probability estimates, enabling post-training threshold optimization. Predicted probabilities were evaluated across the full range using precision–recall curves and corresponding F1-scores, and the threshold maximizing the F1-score was selected as the optimal operating point. This process was particularly important for achieving adequate clinical sensitivity, as prioritizing recall reduced the likelihood of missing high-risk neonates while maintaining clinically meaningful precision. For the Support Vector Machine, which outputs decision-function scores rather than probabilities, an equivalent threshold-sweeping procedure was applied to identify the F1-maximizing cut-point.

3 | Results

Three ML models were evaluated on an independent test set of 1,423 neonates (1,000 survivors and 423 deaths). Model performance was examined using both threshold-independent metrics (ROC-AUC, PR-AUC) and threshold-dependent metrics (precision, recall, F1-score, balanced accuracy). All three models exhibited strong discriminative ability, with ROC-AUC values ranging from 0.961 to 0.967 and PR-AUC values from 0.918 to 0.940. The Random Forest model achieved an accuracy of 0.927 and a balanced accuracy of 0.911 at the default threshold (0.50); applying the optimal F1-maximizing threshold (0.428) increased the F1-score to 0.884, improved recall to 0.891, and reduced false negatives from 55 to 46 without loss of precision. XGBoost demonstrated the strongest baseline performance, with an accuracy of 0.935, ROC-AUC of 0.967, and PR-AUC of 0.940; at its optimal threshold (0.717), it reached an F1-score of 0.895, with precision and recall of 0.918 and 0.872, achieving the best overall balance among the models. The SVM model initially yielded high precision but lower recall; however, threshold adjustment (−0.053) substantially improved minority-class detection, raising the F1-score to 0.896, recall to 0.887, and precision to 0.906, with a balanced accuracy of 0.924—representing the largest performance gain following threshold optimization. *Table 1* provides a detailed comparison of model performance before and after threshold adjustment.

Table 1. Performance comparison of ML models at default and optimized thresholds for neonatal mortality prediction.

Model	Threshold	Precision	Recall	F1-score	Accuracy	Notes
Random Forest	0.50	0.882	0.870	0.876	0.9105	Default threshold
Random Forest	0.4282	0.877	0.891	0.8839	0.9191	F1-max threshold
XGBoost	0.50	0.886	0.898	0.892	0.9247	Default threshold
XGBoost	0.7171	0.918	0.872	0.8945	0.9197	F1-max threshold
SVM (RBF)	0.00	0.957	0.728	0.827	0.8571	Default margin
SVM (RBF)	-0.0535	0.906	0.887	0.8961	0.9238	F1-max threshold

4 | Discussion

This study developed and evaluated three supervised ML models—Random Forest, XGBoost, and Support Vector Machine—to predict neonatal mortality using routinely collected data from the national IMaN registry. Feature-importance analysis consistently highlighted gestational age, birth weight, anthropometric measures (head circumference and height), congenital anomalies, Intrauterine Growth Restriction (IUGR), and intrapartum risk factors as the most influential predictors—variables that are available before or at the time of delivery, enabling real-time clinical decision support. Among the evaluated models, XGBoost achieved the highest overall predictive performance (F1 = 0.8945; balanced accuracy = 0.9197) and the strongest discrimination (ROC-AUC = 0.967; PR-AUC = 0.940) (Figs. 4 and 5). It also maintained the most favorable precision–recall trade-off, with high precision (0.918) and clinically acceptable recall (0.872), and demonstrated superior stability across cross-validation folds. XGBoost’s ability to capture complex nonlinear interactions among key obstetric and fetal variables further strengthened its performance. Taken together, these advantages—high accuracy, robust generalization, interpretability of contributions, and real-time feasibility—justify the selection of XGBoost as the final model for integration into a neonatal Decision Support System (DSS). Importantly, compared with prior neonatal ML studies that typically rely on a narrower subset of predictors, our work incorporates a substantially broader feature space (163 clinically curated variables) covering maternal, fetal, obstetric, and process-related characteristics. Despite this expanded and more heterogeneous feature set, the model achieved discriminative performance comparable to the best published results, including CNN-based NICU models that operate on post-birth data [17]. Although certain deep-learning models in previous studies report slightly higher accuracy, they were trained on more limited or highly specialized feature sets, whereas our model operates on a substantially larger, clinically diverse feature space and maintains far greater interpretability—an advantage that enhances clinical trust and supports meaningful medical decision-making. This demonstrates both the robustness of the modeling pipeline and the value of leveraging richer information for early risk assessment. Feature importance analysis of the final XGBoost model revealed that gestational age (importance = 0.179) and birth weight (0.109) were by far the most influential predictors, together accounting for nearly 29% of the model’s decision-making. Nervous system malformations (0.035) and musculoskeletal malformations (0.033) ranked third and fourth, followed by high-risk delivery indicators (0.032) and other congenital anomalies (0.031). The top 10 features alone contributed more than 55% of total importance, confirming that a small, clinically intuitive set of variables drives most of the predictive signal. Our model’s top predictors align closely with established risk factors in neonatal mortality literature, including gestational age, birth weight, maternal education/sociodemographic status, and Congenital anomalies [5, 10, 18, 19]. This divergence may reflect dataset-specific emphases or methodological differences. Overall, our model reinforces core clinical determinants while enriching the feature set with detailed familial risks, potentially addressing gaps in socioeconomic and health dimensions underrepresented in the literature.

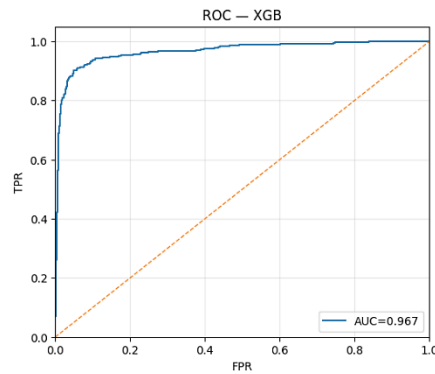


Fig. 4. Receiver Operating Characteristic (ROC) curve for the XGBoost model.

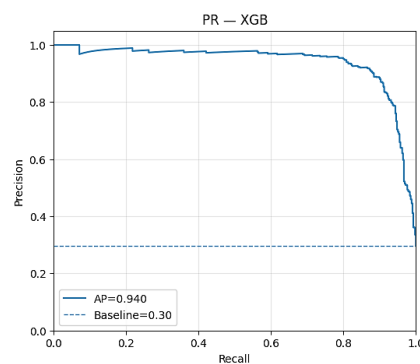


Fig. 5. Precision-Recall curve for the XGBoost model.

Predictive modeling of neonatal mortality enables early identification of high-risk cases and highlights modifiable maternal, socioeconomic, and healthcare-related factors. These insights support targeted interventions, more efficient resource allocation, and evidence-based policy decisions, particularly in low-resource settings. By forecasting risk using routine clinical and demographic variables, ML models facilitate proactive measures—such as maternal education and timely clinical care—that can reduce preventable neonatal deaths. This study demonstrates how ML-based mortality prediction can directly enhance neonatal healthcare processes. Current perinatal workflows often suffer from delays in identifying high-risk neonates, inconsistent triage decisions, and late activation of NICU resources. By providing early and accurate risk estimates, the developed models enable earlier intervention, more efficient resource allocation, and greater standardization across care pathways. Embedding these models within a clinical DSS allows automated risk assessment at admission, real-time alerts to neonatal teams, and timely preparation for critical interventions. This shifts practice from reactive decision-making to a proactive, anticipatory model aligned with Lean healthcare and patient-safety principles. Automated detection of high-risk cases also supports better routing of deliveries to facilities with appropriate NICU capacity, reducing unplanned transfers and improving system-wide operational stability. Moreover, using data-driven predictions helps minimize variability linked to subjective judgment, thereby enhancing reliability and equity in neonatal care.

5 | Conclusion

This study demonstrates that ML models can accurately predict neonatal mortality using routinely collected data. The integration of such models into clinical workflows offers a powerful pathway for health-process improvement, enabling hospitals to deliver safer, faster, and more reliable neonatal care. A neonatal DSS built on these models can standardize risk detection, optimize resource allocation, and support early clinical interventions, ultimately improving survival outcomes.

References

- [1] Osman, M. O., Nur, A. M., Nour, T. Y., Hashi, M. H., & Osman, A. A. (2020). Prevalence and causes of neonatal mortality among neonates admitted in neonatal intensive care unit at Sheik Hassan Yabare Referral Hospital, East Ethiopia 2019. *Science*, 9(1), 11–17. <https://doi.org/10.11648/j.sjcm.20200901.13>
- [2] Mekonnen, T., Tenu, T., Aklilu, T., & Abera, T. (2018). Assessment of neonatal death and causes among admitted neonates in neonatal intensive care unit of Mizan Tepi University Teaching Hospital, Bench Maji Zone, South-West Ethiopia, 2018. *Clin mother child health*, 15(4), 2. <https://doi.org/10.4172/2090-7214.1000305>
- [3] World Bank Group. (2016). *Mortality rate, neonatal (per 1,000 live births) - Iran, Islamic Rep.* <https://data.worldbank.org/indicator/SH.DYN.NMRT?locations=IR>
- [4] Eurostat. (2026). *Population change - Demographic balance and crude rates at national level.* https://doi.org/10.2908/DEMO_GIND
- [5] Sepanlou, S. G., Aliabadi, H. R., Naghavi, M., & Malekzadeh, R. (2022). Neonate, infant, and child mortality by cause in provinces of Iran: An analysis for the global burden of disease study 2019. *Archives of Iranian Medicine*, 25(8), 484–495. <https://doi.org/10.34172/aim.2022.80>
- [6] Khajavi, A., Pishgar, F., Dehghani, M., & Naderimagham, S. (2017). Socioeconomic inequalities in neonatal and postneonatal mortality: Evidence from rural Iran, 1998–2013. *International Journal for Equity in Health*, 16(1), 83. <https://doi.org/10.1186/s12939-017-0570-6>
- [7] Heidarzadeh, M., Rahimi, R., Mahmoudi, M., Habibelahi, A., Ali Akbari khoei, R., AbdollahiAbed, L., ... , & Daemi, A. (2025). Impact of managerial and clinical interventions on trend of neonatal care and mortality in a regional hospital in Iran: A retrospective study. *Egyptian Pediatric Association Gazette*, 73(1), 79. <https://doi.org/10.1186/s43054-025-00423-1>
- [8] Gephart, S. M., Quinn, M. C., Tolentino, D. A., & Wyles, C. (2023). Neonatal intensive care workflow analysis informing NEC-Zero clinical decision support design. *CIN: Computers, Informatics, Nursing*, 41(2), 94–101. <https://doi.org/10.1097/CIN.0000000000000929>
- [9] Yasrebinia, S., & Rezaei, M. (2024). Systematic review article: Artificial intelligence for predicting neonatal mortality in post-pregnancy: A systematic review. *Significance*, 3(1), 81–87. <https://doi.org/10.5281/zenodo.20231109>
- [10] Batista, A. F. M., Diniz, C. S. G., Bonilha, E. A., Kawachi, I., & Chiavegatto Filho, A. D. P. (2021). Neonatal mortality prediction with routinely collected data: A machine learning approach. *BMC Pediatrics*, 21(1), 322. <https://doi.org/10.1186/s12887-021-02788-9>
- [11] Bozhbanbayeva, N., Abilbayeva, A., Tarabayeva, A., Isgandarov, I., Tolykbayev, T., Adilbekova, I., & Yelyubayeva, D. (2025). *DeepSurv: A machine learning model for predicting mortality in very-low-birth-weight infants treated in intensive care units.* <https://doi.org/10.21203/rs.3.rs-7126001/v1>
- [12] Yang, Y. H., Wang, T. T., Su, Y. H., Chu, W. Y., Lin, W. T., Chen, Y. J., ... , & Lin, Y. J. (2024). Predicting early mortality and severe intraventricular hemorrhage in very-low birth weight preterm infants: A nationwide, multicenter study using machine learning. *Scientific Reports*, 14(1), 10833. <https://doi.org/10.1038/s41598-024-61749-1>
- [13] Do, H. J., Moon, K. M., & Jin, H. S. (2022). Machine learning models for predicting mortality in 7472 very low birth weight infants using data from a nationwide neonatal network. *Diagnostics*, 12(3), 1–10. <https://doi.org/10.3390/diagnostics12030625>
- [14] Conte, L., Amodeo, I., De Nunzio, G., Raffaelli, G., Borzani, I., Persico, N., ... , & Cavallaro, G. (2025). A machine learning approach to predict mortality and neonatal persistent pulmonary hypertension in newborns with congenital diaphragmatic hernia. A retrospective observational cohort study. *European Journal of Pediatrics*, 184(4), 238. <https://doi.org/10.1007/s00431-025-06073-0>
- [15] Khatibi, T., Hanifi, E., Sepehri, M. M., & Allahqoli, L. (2021). Proposing a machine-learning based method to predict stillbirth before and during delivery and ranking the features: Nationwide retrospective cross-sectional study. *BMC Pregnancy and Childbirth*, 21(1), 202. <https://doi.org/10.1186/s12884-021-03658-z>

- [16] Shariat, N., Kargari, M., Alavi, M., Shariat, S., & Valiollahi, A. (2024). Neonatal mortality prediction in nicus: A machine learning approach. *2024 10th international conference on web research (ICWR)* (pp. 231–237). IEEE. <https://doi.org/10.1109/ICWR61162.2024.10533355>
- [17] Qeytasi, M., Sepehri, M. M., & Habibelahi, A. (2025). Quantifying neonatal transport deficiencies: An integrated framework for weighting core KPIs and ranking challenges in resource-limited settings. *European Journal of Pediatrics*, *184*(11), 686. <https://doi.org/10.1007/s00431-025-06523-9>
- [18] Yoneda, K., Seki, T., Kawazoe, Y., Ohe, K., Takahashi, N., & of Japan, N. R. N. (2024). Immediate postnatal prediction of death or bronchopulmonary dysplasia among very preterm and very low birth weight infants based on gradient boosting decision trees algorithm: A nationwide database study in Japan. *Plos One*, *19*(3), e0300817. <https://doi.org/10.1371/journal.pone.0300817>
- [19] Karimi, P., Mahmudi, L., Azami, M., & Badfar, G. (2019). Mortality in neonatal intensive care units in Iran: A systematic review and meta-analysis. *Iranian Journal of Neonatology*, *10*(3), 70–80. <https://doi.org/10.22038/ijn.2019.36647.1566>