Quality Festival 2025
ISBN 978-86-6335-121-9

Jelena Cekovic Djordjevic[1]
Aleksandra Simović
Dragana Savić
Tijana Prodanović
Suzana Živojinović
Milan Erić
Miladin Stefanović
Aleksandar Đorđević

Research paper
DOI – 10.24874/QF.25.043

QUALITY-DRIVEN MACHINE LEARNING FOR NEONATAL CARE: PREDICTING NECROTIZING ENTEROCOLITIS

Abstract: *Ensuring the quality and reliability of predictive models in neonatal healthcare is crucial for improving early disease detection and clinical decision-making. This study investigates the application of machine learning (ML) algorithms for predicting necrotizing enterocolitis (NEC) in neonatal populations, focusing on model selection, performance evaluation, and quality assessment. A dataset of 207 neonates, including 143 preterm and 64 term infants, was analyzed using six ML classification models: Logistic Regression (LR), Linear Discriminant Analysis (LDA), K-Nearest Neighbors (KNN), Classification and Regression Trees (CART), Naïve Bayes (NB), and Support Vector Machine (SVM). Model performance was assessed using accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUROC). This study underscores the potential of machine learning in neonatal care and suggests that a hybrid approach combining high-recall and high-precision models could optimize NEC detection. Future research should focus on ensemble learning techniques and clinical validation to further enhance predictive performance and practical implementation in neonatal intensive care units.*

Keywords: *machine learning, data-driven analyses, predictive modeling, quality assessment, reliability, healthcare analytics, neonatal care, necrotizing enterocolitis (NEC).*

1. Introduction

In recent years, the integration of data-driven approaches has become a crucial element in scientific research and healthcare advancements. The ability to analyze and extract meaningful insights from large datasets has led to significant improvements in various medical fields, including neonatology (Bashiri et al., 2003). Neonatology, which focuses on the care of newborns, particularly those who are premature or critically ill, generates an extensive amount of clinical data through continuous monitoring, laboratory tests, and imaging studies. The need for precise and timely decision-making in neonatal care has paved the way for the application of advanced computational techniques such as machine learning (ML) to improve diagnostic accuracy and treatment outcomes (Patel et al., 2024).

Machine learning algorithms, through predictive modeling, offer a data-driven approach to identifying high-risk conditions in neonates. One such condition is

[1] Corresponding author: Jelena Cekovic Djordjevic
Email: j.cekovic86@gmail.com

necrotizing enterocolitis (NEC) (Yee et al., 2012), a severe gastrointestinal disease primarily affecting premature infants. NEC is characterized by inflammation and necrosis of the intestinal wall, leading to life-threatening complications if not detected early (Ackermann et al., 2022). Despite advances in neonatal care, early detection remains a significant challenge due to the disease's multifactorial nature and nonspecific clinical presentation. Traditional diagnostic approaches rely on clinical symptoms, radiographic findings, and laboratory markers, but these methods often detect NEC at an advanced stage when treatment options are limited. Given these challenges, there is an increasing interest in leveraging machine learning models to enhance early NEC detection and improve patient outcomes.

The application of machine learning in neonatal healthcare has gained attention due to its potential to uncover hidden patterns in complex datasets. By analyzing large volumes of patient data, ML algorithms can identify subtle indicators of disease progression that may not be immediately apparent through traditional clinical assessments. Various ML techniques, including random forests, logistic regression, artificial neural networks, and support vector machines, have been explored in previous studies for NEC prediction (Beam et al., 2024). However, achieving a balance between predictive accuracy, reliability, and clinical applicability remains a key challenge.

A critical aspect of implementing machine learning in neonatal healthcare is ensuring the quality and reliability of predictive models. The effectiveness of an ML model depends not only on its ability to achieve high accuracy but also on its robustness, generalizability, and ability to provide interpretable results for clinical decision-making. In the context of NEC prediction, ensuring model reliability is essential, as false positives could lead to unnecessary interventions, while false negatives might result in delayed treatment and increased mortality risk. Therefore, quality assessment of ML models—through rigorous validation techniques, hyperparameter optimization, and performance benchmarking—is crucial for their successful application in neonatal care.

Previous research on ML-based NEC prediction has demonstrated promising results. One of the earliest studies in this area, conducted by Mueller et al. (2009), applied artificial neural networks to identify significant risk factors for NEC, such as low birth weight and artificial ventilation requirements. More recent studies have explored alternative machine learning models, including decision trees, gradient boosting methods, and ensemble learning techniques, to improve prediction accuracy. However, while these models have shown high performance in controlled datasets, challenges related to model generalization and real-world implementation remain.

This study aims to build upon existing research by focusing on ML model selection for NEC prediction, with an emphasis on quality assessment and reliability evaluation (Beam et al., 2024). The primary objectives of this research include:

- Comparative evaluation of multiple machine learning algorithms to determine the most effective model for NEC prediction.
- Optimization of model hyperparameters to enhance predictive accuracy and reduce overfitting.
- Validation and quality assessment of the selected model using appropriate statistical and machine learning metrics.
- Analysis of feature importance and risk factor identification, ensuring that the model provides interpretable insights for clinical use.

By conducting a rigorous evaluation of ML techniques, this study seeks to provide a quality-driven framework for ML applications in neonatal care. Ensuring that predictive models meet high standards of reliability and interpretability is essential for their integration into clinical workflows.

A fundamental challenge in developing high-quality machine learning models for NEC prediction lies in the selection of relevant features and their impact on predictive accuracy. NEC is a complex disorder with multiple contributing factors, including gestational age, birth weight, maternal health conditions, and prenatal environmental influences. The ability to extract meaningful features from neonatal datasets plays a crucial role in enhancing model performance. Furthermore, considering the variability in clinical presentation, ML models must account for heterogeneity in neonatal patient populations, ensuring that predictions remain accurate across diverse clinical settings.

This study is based on data collected from the Neonatology Clinic in Kragujevac, Serbia, providing a real-world dataset for evaluating ML-based NEC prediction models. The research framework involves data preprocessing, feature selection, model training, hyperparameter tuning, and performance evaluation using cross-validation techniques.

2. Materials and Methods

This study was conducted at the Neonatology Center, Pediatric Clinic, and Obstetrics Clinic of the University Clinical Center Kragujevac (UCC Kragujevac). Data were prospectively collected from January 2018 to January 2021, followed by a retrospective case-control observational analysis. The study was approved by the Ethics Committee of the Faculty of Medical Sciences, University of Kragujevac and the University Clinical Center Kragujevac.

Exclusion criteria included fatal outcomes within 24 hours, gestational age <24 or >42 weeks, and the presence of major congenital anomalies or chromosomal abnormalities. Clinical and laboratory data were collected from electronic medical records, including fecal calprotectin (FCP) measurements, complete blood count (CBC), C-reactive protein (CRP), procalcitonin (PCT), glucose levels, and blood cultures. Fecal samples were analyzed using the Alegriá® ELISA test, while blood samples were processed with standard hematological and biochemical assays.

To develop a predictive model for NEC, six ML classification algorithms were applied: Logistic Regression (LR), Linear Discriminant Analysis (LDA), K-Nearest Neighbors (KNN), Classification and Regression Trees (CART), Naïve Bayes (NB), and Support Vector Machines (SVM). Given the imbalanced dataset, model performance was evaluated using accuracy, precision, recall, F1-score, and the area under the ROC curve (AUROC), as classification accuracy alone was insufficient to assess model effectiveness.

Incomplete prenatal records for neonates transferred from other hospitals, potential unmeasured confounders, and variability in fecal calprotectin measurements may have affected the results. Additionally, as only a single fecal sample was collected per patient, the dynamic nature of calprotectin levels over time could not be assessed. Finally, the imbalance in the dataset required the implementation of oversampling techniques to improve the reliability of ML training.

This study applies machine learning techniques to predict NEC, integrating rigorous validation, and quality assessment. The findings are expected to contribute to the early diagnosis of NEC, improve neonatal care, and provide AI-driven decision support for clinical applications.

3. Results

During the three-year study period, a total of 207 neonates were analyzed, including 143 preterm and 64 term infants (Table 1, Figure

1). The distribution of sex was similar between the groups (p=0.254), indicating no significant difference. Likewise, the age at sample collection did not show a significant variation (p=0.225). However, significant differences were observed in gestational age (GA), birth weight (BW), Apgar score (AS), mode of delivery, need for resuscitation, inotropic support, and mechanical ventilation, all of which were group-dependent (p<0.001).

Table 1. Descriptive Statistics Table

Clinical Characteristics (n=207)	Preterm Neonates (n=143)	Term Neonates (n=64)	p-value
Male (%)	65/143 (45.5%)	31/64 (48.43%)	p=0.254*
Female (%)	78/143 (54.5%)	33/64 (51.56%)	
GA (weeks, mean±SD)	32.44±2.34	39.1±1.09	p<0.001**
BW (g, mean±SD)	1822.34±416.10	3390±418.94	p<0.001**
Apgar Score at 5 min (mean±SD)	6.83±1.66	9.06±0.68	p<0.001**
Vaginal Delivery (%)	63/143 (44%)	40/64 (62.5%)	p=0.010*
Cesarean Section (%)	80/143 (56%)	24/64 (37.5%)	
Use of Inotropes (%)	12/131 (8.4%)	0/64 (0%)	p=0.010*
Resuscitation (%)	70/143 (48%)	0/64 (0%)	p<0.001**
Mechanical Ventilation (%)	57/143 (39.8%)	0/64 (0%)	p<0.001**
Days on Mechanical Ventilation (mean±SD)	3.47±5.83	0	p<0.001**
Mortality (%)	4/143 (2.79%)	0/64 (0%)	p=0.225*

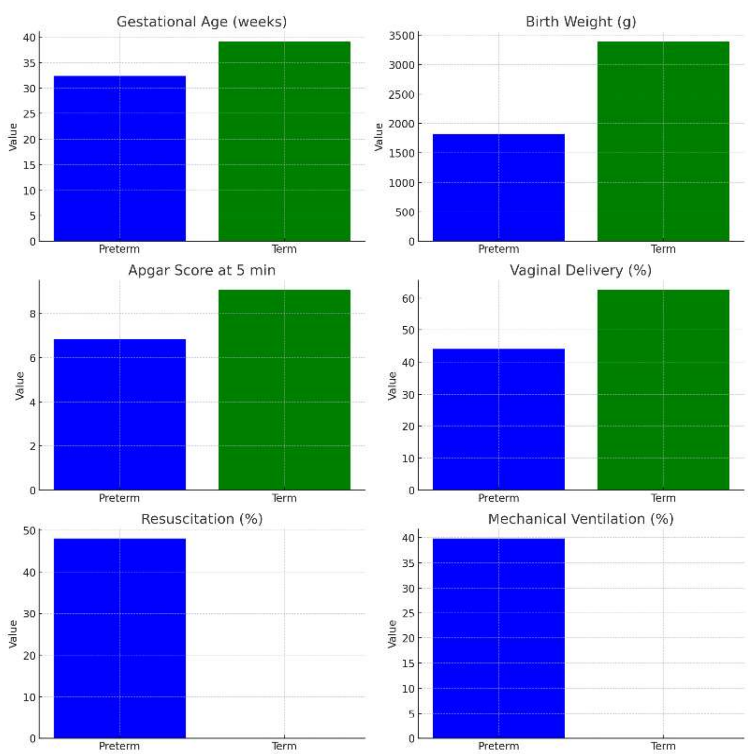

Figure 1. Clinical Differences Between Preterm and Term Neonates

Preterm infants had significantly lower mean gestational age (32.44±2.34 weeks) compared to term neonates (39.1±1.09 weeks, p<0.001). Similarly, birth weight was significantly lower in the preterm group (1822.34±416.10 g vs. 3390±418.94 g, p<0.001). The Apgar score at 5 minutes was also notably reduced in preterm neonates (6.83±1.66 vs. 9.06±0.68, p<0.001), reflecting a higher need for immediate postnatal support.

Regarding the mode of delivery, cesarean section was more frequently performed in preterm neonates (56%) compared to term neonates (37.5%, p=0.01). The need for cardiopulmonary resuscitation (CPR) was also significantly higher in the preterm group, with 48% requiring resuscitation, whereas none of the term neonates needed it (p<0.001). Similarly, inotropic support was required in 8.4% of preterm infants, while none of the term neonates needed inotropic agents (p=0.010).

A notable difference was observed in the need for mechanical ventilation (MV). Among preterm neonates, 39.8% required MV, while none of the term neonates needed respiratory support (p<0.001). The duration of mechanical ventilation also varied significantly, with preterm neonates requiring an average of 3.47±5.83 days of ventilatory support. Mortality was slightly higher in the preterm group (2.79% vs. 0% in term neonates), but this difference was not statistically significant (p=0.225), likely due to the limited number of cases.

These findings confirm that preterm birth is associated with a higher incidence of perinatal complications, increased need for medical interventions, and a greater likelihood of requiring intensive care support. The next section will explore the application of ML models in predicting the occurrence of necrotizing enterocolitis (NEC) based on these clinical parameters.

To evaluate the predictive potential of ML algorithms in detecting NEC, six different classification models were applied (Table 2): LR, LDA, KNN, CART, NB, and SVM. Each model was assessed based on key performance metrics, including accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (ROC AUC).

The results demonstrated high overall accuracy across all models, with NB achieving the highest mean accuracy (0.9881 ± 0.0206), closely followed by SVM (0.9878 ± 0.0122) and CART (0.9820 ± 0.0198). In terms of ROC AUC, several models, including LDA, NB, and SVM, achieved a perfect score of 1.0000, indicating their strong ability to differentiate between NEC and non-NEC cases.

Precision and recall values varied significantly between models. LDA and NB exhibited perfect recall (1.0000 ± 0.0000), ensuring that all NEC cases were correctly identified, whereas KNN and SVM showed lower recall (0.6250 ± 0.2165 and 0.7500 ± 0.2500, respectively). On the other hand, CART and SVM achieved the highest precision (1.0000 ± 0.0000), minimizing false positive predictions.

Table 2. Performance Metrics of Applied Machine Learning Models for NEC Prediction

Model	Accuracy (Mean ± Std)	Precision (Mean ± Std)	Recall (Mean ± Std)	F1 Score (Mean ± Std)	ROC AUC (Mean ± Std)
LR	0.9698 ± 0.0200	0.6250 ± 0.4146	0.5000 ± 0.3536	0.5417 ± 0.3608	0.9968 ± 0.0056
LDA	0.9759 ± 0.0168	0.7083 ± 0.1816	1.0000 ± 0.0000	0.8167 ± 0.1190	1.0000 ± 0.0000
KNN	0.9819 ± 0.0105	1.0000 ± 0.0000	0.6250 ± 0.2165	0.7500 ± 0.1443	0.9984 ± 0.0028
CART	0.9820 ± 0.0198	1.0000 ± 0.0000	0.8750 ± 0.2165	0.7917 ± 0.2165	0.8125 ± 0.2073
NB	0.9881 ± 0.0206	0.8750 ± 0.2165	1.0000 ± 0.0000	0.9167 ± 0.1443	1.0000 ± 0.0000

Model	Accuracy (Mean ± Std)	Precision (Mean ± Std)	Recall (Mean ± Std)	F1 Score (Mean ± Std)	ROC AUC (Mean ± Std)
SVM	0.9878 ± 0.0122	1.0000 ± 0.0000	0.7500 ± 0.2500	0.8333 ± 0.1667	1.0000 ± 0.0000

The F1-score, which balances precision and recall, was highest for NB (0.9167 ± 0.1443), suggesting that this model provides the most balanced classification performance. LDA also performed well (0.8167 ± 0.1190), while LR showed the weakest performance in this aspect (0.5417 ± 0.3608) due to lower recall.

Furthermore, Figure 2 presents the comparative performance of the six applied machine learning models in NEC prediction.

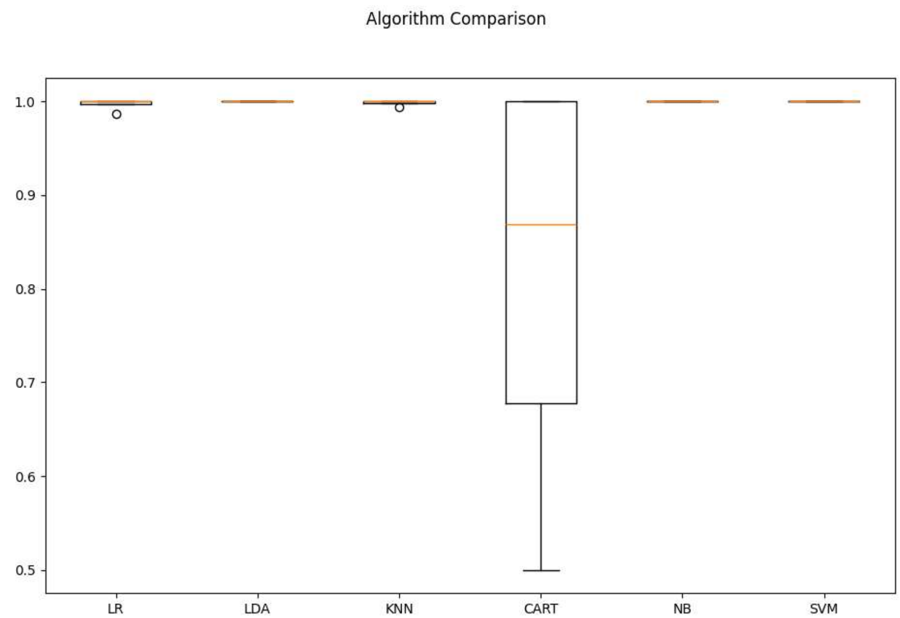

Figure 2. Comparison of Machine Learning Model Performance for NEC Prediction

The box plot highlights variations in model stability and predictive consistency across multiple runs. Notably, most models exhibit minimal variation, maintaining high performance with limited deviations. However, the CART model demonstrates a broader distribution, indicating greater variability in prediction accuracy. This suggests that while CART achieves competitive precision, its performance is less stable compared to other models such as LDA, NB, and SVM, which consistently maintain high accuracy with minimal deviation. Given the need for reliable early NEC detection, models with low variance and high recall, such as Naïve Bayes and LDA, may be preferable for clinical implementation.

These results indicate that Naïve Bayes and Support Vector Machine models provide the most reliable classification performance for NEC prediction, balancing high accuracy, recall, and precision. However, considering clinical applicability, models with high recall (such as LDA and NB) may be preferable, as they ensure early NEC detection while minimizing false negatives.

4. Discussion

The evaluation of ML models for NEC prediction in this study demonstrated high classification accuracy across all tested algorithms, with NB, SVM, and CART achieving the best overall performance. However, the importance of different metrics must be considered when assessing the real-world applicability of these models, particularly in clinical settings where early NEC detection is crucial to improve neonatal outcomes.

One key observation from this study is the trade-off between precision and recall among the models. While NB and LDA achieved 100% recall, meaning that all NEC cases were correctly identified, their precision values were lower, particularly in LDA (0.7083 ± 0.1816). This suggests that while these models excel at capturing NEC cases, they might produce more false positives, which could lead to unnecessary interventions. On the other hand, CART and SVM exhibited perfect precision (1.0000 ± 0.0000), meaning no false positives, but with slightly reduced recall (0.8750 ± 0.2165 and 0.7500 ± 0.2500, respectively), indicating a risk of missing some NEC cases.

Comparison with Previous Studies

Several previous studies have explored machine learning applications in NEC prediction, with varying results based on the dataset characteristics, feature selection, and algorithmic approaches.

Mueller et al. (2009) were among the first to apply ANNs to NEC prediction, identifying important risk factors such as low birth weight and mechanical ventilation requirements. Their model achieved moderate accuracy, but lacked interpretability, making it challenging for clinical implementation.

Doheny et al. (2018) focused on heart rate variability as a predictor for NEC, applying decision trees and ensemble learning methods. Their findings indicated that random forest performed well, but with lower recall compared to our study's LDA and NB models, which highlights the importance of balancing sensitivity and specificity in NEC detection.

A more recent study by Pantalone et al. (2022) used random forest models based on complete blood count (CBC) data, achieving high overall accuracy but with lower sensitivity in distinguishing between medical and surgical NEC cases. In contrast, our study demonstrates that Naïve Bayes and LDA models achieve higher recall, making them potentially more useful in clinical settings where early detection is a priority.

Cho et al. (2023) applied six supervised learning models to a dataset of 74 clinical features, finding that logistic regression and random forest achieved the best performance in terms of AUROC and sensitivity. However, in our study, LDA, NB, and SVM models achieved AUROC scores of 1.0000, outperforming many previously reported models.

While high accuracy and AUROC values suggest strong predictive capabilities, the choice of the best ML model depends on the clinical context. In NEC prediction, recall is often more critical than precision, as missing a true NEC case could lead to delayed diagnosis and higher morbidity and mortality rates. Therefore, LDA and Naïve Bayes, which demonstrated 100% recall, might be preferable for clinical implementation, despite their slightly lower precision.

However, if the goal is to minimize false positives and avoid unnecessary interventions, models like CART and SVM, which exhibited perfect precision, could be considered in combination with other clinical risk stratification methods. A hybrid approach, where high-recall models are used for screening and high-precision models for confirmation, may provide an optimal balance between sensitivity and specificity in NEC detection.

5. Conclusion

ML has the potential to revolutionize neonatal healthcare by providing data-driven insights for early disease detection and improving patient outcomes. However, ensuring high-quality predictive models requires a combination of robust algorithm selection, optimization techniques, and interpretability mechanisms. This study aims to contribute to this field by presenting a quality-driven approach to ML-based NEC prediction, with a focus on enhancing model reliability and practical applicability in clinical settings.

By addressing key challenges in predictive modeling, feature selection, and quality assessment, this research advances the use of machine learning in neonatal informatics, providing a foundation for future innovations in quality-driven healthcare analytics.

This study confirms that machine learning models, particularly Naïve Bayes, LDA, and SVM, can be effectively applied for NEC prediction with high accuracy. Compared to previous studies, our models demonstrate superior recall and AUROC performance, emphasizing their potential for early NEC detection and risk stratification. Future research should explore ensemble learning techniques and hybrid approaches to further improve predictive robustness and clinical applicability, ensuring that these models can be effectively integrated into neonatal care protocols.

The practical significance of this research lies in its potential to improve early NEC detection and intervention strategies in neonatal care. By developing a high-quality, reliable predictive model, this study contributes to reducing the incidence of severe NEC cases through proactive risk assessment and timely medical interventions.

Furthermore, this research aligns with broader themes in quality engineering, risk management, and healthcare analytics, making it relevant for discussions on quality-driven AI applications in medicine. By focusing on reliability, optimization, and explainability, this study addresses key challenges associated with the implementation of ML in clinical environments.

References:

Ackermann, K., Baker, J., Festa, M., McMullan, B., Westbrook, J., & Li, L. (2022). Computerized clinical decision support systems for the early detection of sepsis among pediatric, neonatal, and maternal inpatients: Scoping review. JMIR Medical Informatics, 10(5), e35061.

Bashiri, A., Zmora, E., Sheiner, E., Hershkovitz, R., Shoham-Vardi, I., & Mazor, M. (2003). Maternal hypertensive disorders are an independent risk factor for the development of necrotizing enterocolitis in very low birth weight infants. Fetal Diagnosis and Therapy, 18(6), 404–407. .

Beam, K., Sharma, P., Levy, P., & Beam, A. L. (2024). Artificial intelligence in the neonatal intensive care unit: The time is now. Journal of Perinatology, 44(1), 131–135.

Cho, H., Lee, E. H., Lee, K. S., & Heo, J. S. (2022). Machine learning-based risk factor analysis of necrotizing enterocolitis in very low birth weight infants. Scientific Reports, 12(1), 21407.

Mueller, M., Taylor, S. N., Wagner, C. L., & Almeida, J. S. (2009, June). Using an artificial neural network to predict necrotizing enterocolitis in premature infants. In 2009 International joint conference on neural networks (pp. 2172-2175). IEEE.

Patel, S. Y., Palma, J. P., Hoffman, J. M., & Lehmann, C. U. (2024). Neonatal informatics: Past, present, and future. *Journal of Perinatology, 44*(6), 773–776. doi: 10.1038/s41372-024-01924-4.

Yee, W. H., Soraisham, A. S., Shah, V. S., Aziz, K., Yoon, W., Lee, S. K., & Canadian Neonatal Network. (2012). Incidence and timing of presentation of necrotizing enterocolitis in preterm infants. Pediatrics, 129(2), e298–e304.

Jelena Ceković Đorđević
Center for Neonatology,
Pediatric Clinic, University
Clinical Center Kragujevac,
Kragujevac,
Serbia
j.cekovic86@gmail.com
ORCID 0000-0003-3312-6307

Aleksandra Simović
Department of Pediatrics, Faculty
of Medical Sciences, University of
Kragujevac,
34000 Kragujevac,
Serbia
Center for Neonatology, Pediatric
Clinic, University Clinical Center
Kragujevac,
Kragujevac,
Serbia
aleksandra.simovic@yahoo.com
ORCID 0000-0003-0055-1972

Dragana Savić
Department of Pediatrics,
Faculty of Medical Sciences,
University of Kragujevac,
34000 Kragujevac,
Serbia
Center for Neonatology,
Pediatric Clinic, University
Clinical Center Kragujevac,
Kragujevac,
Serbia
drsavicdragana@gmail.com
ORCID 0000-0001-9324-3279

Tijana Prodanović
Department of Pediatrics,
Faculty of Medical Sciences,
University of Kragujevac,
34000 Kragujevac,
Serbia
Center for Neonatology,
Pediatric Clinic, University
Clinical Center Kragujevac,
Kragujevac,
Serbia
tijanaprodanovic86@gmail.com
ORCID 0000-0002-7399-2480

Suzana Živojinović
Department of Pediatrics, Faculty
of Medical Sciences, University of
Kragujevac,
34000 Kragujevac,
Serbia
Center for Neonatology, Pediatric
Clinic, University Clinical Center
Kragujevac,
Kragujevac,
Serbia
zivojinovicsuzana@yahoo.com
ORCID 0000-0002-6844-2150

Milan Erić
Department for Production
Engineering, Faculty of
Engineering, University of
Kragujevac, Kragujevac,
Serbia
ericm@kg.ac.rs
ORCID 0000-0001-8691-4863

Miladin Stefanović
Department for Production
Engineering, Faculty of
Engineering, University of
Kragujevac,
Kragujevac,
Serbia
miladin@kg.ac.rs
ORCID 0000-0002-2681-0875

Aleksandar Đorđević
Department for Production
Engineering, Faculty of
Engineering, University of
Kragujevac,
Kragujevac,
Serbia
adjordjevic@kg.ac.rs
ORCID 0000-0003-2856-6578