Heliyon 11 (2025) e41516

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Functional gastrointestinal disorders predictors in neonates and toddlers: A machine learning approach to risk assessment

Flavia Indrio [a,*], Elio Masciari [b], Flavia Marchese [c], Matteo Rinaldi [d], Gianfranco Maffei [d], Ilaria Gangai [c], Assunta Grillo [c], Roberta De Benedetto [c], Enea Vincenzo Napolitano [b], Isadora Beghetti [e], Luigi Corvaglia [e,f], Antonio Di Mauro [g], Arianna Aceti [e]

[a] Department of Experimental Medicine School of Medicine University of Salento, Lecce, Italy
[b] Dipartimento di Ingegneria Elettrica e delle Tecnologie dell'Informazione, University Federico II, Naples, Italy
[c] Department of Medical and Surgical Science Pediatric Section, University of Foggia, 71100, Foggia, Italy
[d] Department of Neonatology and NICU, Ospedali Riuniti Foggia, Viale Pinto 1, 71122, Foggia, Italy
[e] Department of Medical and Surgical Sciences, University of Bologna, Bologna, Italy
[f] Neonatal Intensive Care Unit, IRCCS AOUBO, Bologna, Italy
[g] Pediatric Primary Care, National Pediatric Health Care System, ASL BA, BARI, Italy

ARTICLE INFO

Keywords:
Functional gastrointestinal disorders
Neonatal risk prediction
Artificial intelligence

ABSTRACT

Background: Functional Gastrointestinal Disorders (FGIDs) can pose a great burden on affected children, their families, and the healthcare system. Due to the lack of knowledge about the precise pathophysiology of FGIDs, a proper identification of children at risk to develop FGIDs has never been attempted. The research aims to identify early-life risk factors for FGIDs such as infantile colic, regurgitation, and functional constipation, within the first year of life.
Methods: This prospective observational cohort study enrolled both term and preterm infants from a tertiary care university hospital between January 1, 2020, and December 31, 2022. The study employed both traditional statistical methods and artificial intelligence (AI) techniques, specifically a random forest classification model, to identify key risk factors associated with the development of FGIDs. Based on these findings, an AI-based predictive model will be developed, along with a user-friendly, web-based interface designed for practical risk assessment.
Results: 6060 infants were enrolled. 8.1 % were born preterm. According to random forest classification model by AI, birth weight (BW), cord blood pH, and maternal age were the most relevant variables linked to development of FGIDs in the first year of life. Some discrepancies between potential risk factors identified through conventional statistics and AI were detected.
Conclusion: For the first time machine learning allowed to identify BW, cord blood pH and maternal age as important variable for risk prediction of FGIDs in the first year of life. This practical risk assessment tool would help clinicians to identify infants at risk of FGIDs who would benefit from a tailored preventive approach.

* Corresponding author.
E-mail addresses: flaviaindrio1@gmail.com (F. Indrio), elio.masciari@unina.it (E. Masciari), flavia.marchese@hotmail.it (F. Marchese), mrinaldi@ospedaliriunitifoggia.it (M. Rinaldi), gmaffei@ospedaliriunitifoggia.it (G. Maffei), gangai_ilaria@libero.it (I. Gangai), assuntagrillo88@gmail.com (A. Grillo), roberta_debenedetto.549343@unifg.it (R. De Benedetto), eneavincenzo.napolitano@unina.it (E.V. Napolitano), isadora.beghetti@unibo.it (I. Beghetti), luigi.corvaglia@unibo.it (L. Corvaglia), dimauroantonio@msn.com (A. Di Mauro), arianna.aceti2@unibo.it (A. Aceti).

https://doi.org/10.1016/j.heliyon.2024.e41516
Received 16 August 2024; Received in revised form 16 December 2024; Accepted 25 December 2024
Available online 27 December 2024

1. Introduction

1.1. Background

Functional Gastrointestinal Disorders (FGIDs) are characterized by chronic or recurrent gastrointestinal (GI) symptoms that cannot be attributed to biochemical or structural abnormalities of the GI tract [1]. Despite the lack of an organic cause, FGIDs can significantly impact the well-being of affected children, their families, and the healthcare system [2]. These disorders can manifest in both neonates and older children, with diagnoses and classifications made according to the Rome IV symptom-based criteria [3]. FGIDs in infants, such as infant colic, regurgitation, functional diarrhea, and functional constipation, are prevalent worldwide, although the reported prevalence rates vary widely. Studies suggest that approximately 50 % of infants experience FGIDs, including regurgitation, colic, and constipation, during their first year of life [4–8].

The pathophysiology of FGIDs remains poorly understood [9,10]. Various factors have been implicated, including genetic predispositions, psychosocial influences, abnormal intestinal motility, visceral hyperalgesia, gut inflammation, disrupted microbiota-gut-brain interactions, and early life stress or trauma [11]. Early life events, such as mode of delivery, feeding practices, gestational age (GA), and early antibiotic exposure, have been proposed as potential contributors to the risk of FGIDs in infants [12–15]. However, due to the incomplete understanding of FGIDs' precise pathophysiology, effective identification of at-risk children has not been systematically attempted.

Recently, artificial intelligence (AI) methods have emerged as transformative tools in biomedical research and healthcare, offering significant potential, particularly in pediatric healthcare, where early detection and intervention can substantially influence long-term health outcomes [16,17]. Furthermore, AI techniques can be employed to demonstrate the feasibility of novel data-driven approaches to analysis, moving beyond traditional rule-based paradigms in medicine.

1.2. Objectives

The primary objective of our study was to identify early-life risk factors associated with the development of FGIDs (infantile colic, regurgitation, functional constipation) during the first year of life, within a large, population-based cohort, highlighting potential differences between conventional statistics and AI-based methods. The secondary aim was to develop an AI-based predictive model and a practical risk assessment tool to assist clinicians in identifying infants at high risk for FGIDs who might benefit from targeted preventive strategies.

2. Methods

2.1. Data

This prospective, observational cohort study was conducted from January 1st, 2020, to December 31st, 2023, at the Obstetric and

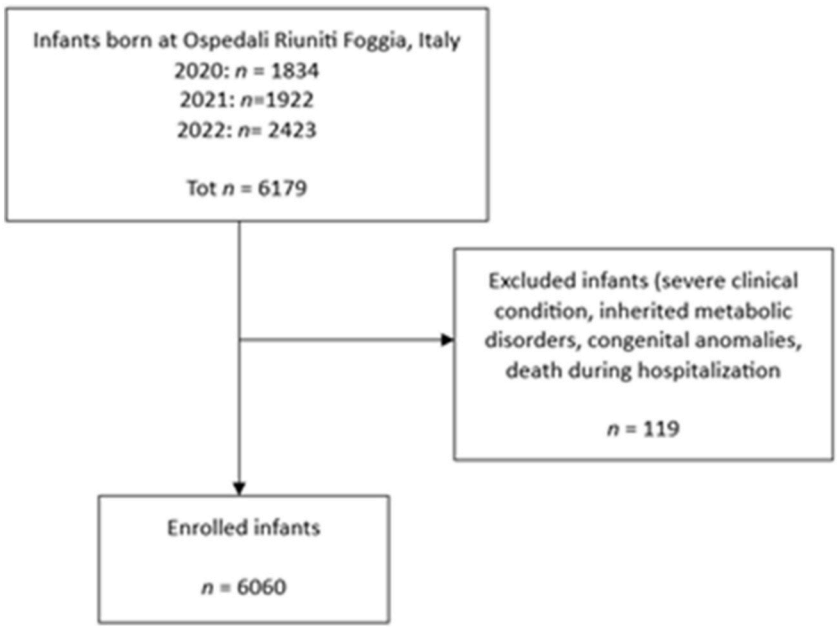

Fig. 1. Flow chart of the patients' enrollment.

Neonatal Unit of the "Ospedale Riuniti" in Foggia, Italy. The study adhered to institutional requirements for data protection, and written informed consent was obtained from the legal representatives of participating infants. The study protocol was approved by the Institutional Review Board and the study was conducted in accordance with the principles of the Declaration of Helsinki.

2.2. Participants

Term and preterm infants were eligible for inclusion in the study. Exclusion criteria included severe clinical conditions, major neonatal complications such as inherited metabolic disorders, congenital anomalies, and neonatal deaths during hospitalization. Newborns meeting the inclusion criteria were consecutively recruited within the first days of life. A flowchart detailing patient enrollment is presented in Fig. 1.

2.3. Data preparation

Data on potential perinatal risk factors for FGIDs, such as gestational age (GA), birth weight (BW), sex, mode of delivery, Apgar score, venous cord blood pH, maternal demographic characteristics, admission to the Neonatal Intensive Care Unit (NICU), antibiotic administration, and feeding practices at discharge, were collected from hospital records and follow-up visits. Feeding at discharge was classified as exclusive breastfeeding or non-exclusive breastfeeding (including exclusive formula feeding and mixed feeding). FGIDs were diagnosed and classified according to Rome IV criteria [3] and assessed by a trained pediatrician during follow-up visits in the first year of life. Additionally, information on family history of allergic disease and parental smoking was obtained from the parents at discharge and at subsequent visits at 3, 6, and 12 months of age.

2.4. Outcome

The primary outcome was to explore associations between FGID development (infantile colic, regurgitation, functional constipation) and perinatal/neonatal characteristics, utilizing both traditional statistical approaches and machine learning (ML) techniques. The secondary outcome was to develop an AI-based predictive model and a practical risk assessment tool to aid clinicians in identifying infants at risk for FGIDs.

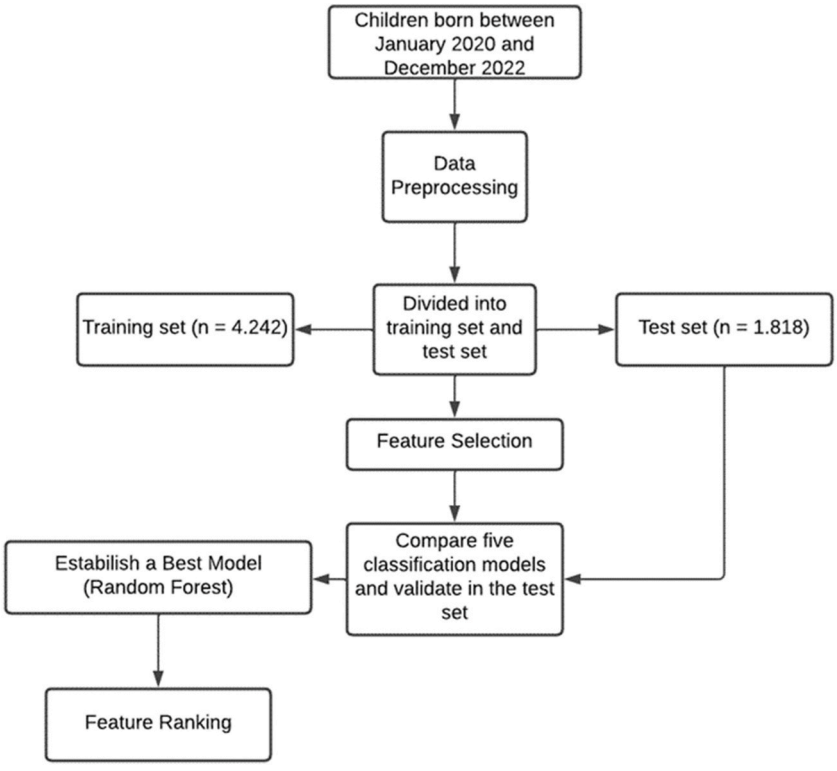

Fig. 2. Flow of machine learning model development Operations and Ranking of variables by importance.

2.5. Data analysis

Potential associations between each FGID and perinatal characteristics were first analyzed using conventional statistics and AI methods. Subsequently, an AI-based predictive model and practical risk assessment tools were developed for each FGID.

2.6. Conventional statistical analyses

Statistical analyses were performed using IBM SPSS Statistics 28.0 (IBM Corp., Armonk, NY, USA). Differences between infants developing or non-developing each FGID were evaluated using the independent sample *t*-test or the Mann-Whitney test for continuous variables, and the chi-squared test for categorical data. For each FGID, variables that proved to be significantly different among groups were included as independent variables into three different regression models (one for each FGID). A p value < 0.05 was considered as statistically significant.

2.7. Machine learning process

A ML process was implemented to analyze the dataset (Fig. 2). An accurate pre-processing was first performed to obtain a dataset suitable for ML analysis. Once a cleaned dataset was obtained, it was divided into a training set (70 %) and a test set (30 %) to ensure the representativeness of the data. The composition of the training and test sets was determined based on the balance of cases for each condition (colic, regurgitation and constipation) to maintain statistical representativeness and reliability of the results.

Feature selection identified the most important variables for outcome prediction. Finally, a classification model based on Random Forest was chosen due to its superior performance in handling non-linear data and its robustness to unbalanced data sets. This decision was further supported by the ease of interpreting the importance of features in Random Forest, which was in line with the practical goals of the study.

2.8. Data preprocessing

During data preparation, a rigorous data cleaning process was undertaken to ensure data integrity and usability. Characteristics with missing rates greater than 30 % were excluded. Missing values for continuous variables were replaced with the mean value, while missing values for discrete variables were replaced with zero. Instances with missing classification variables were removed from the dataset. Continuous variables were standardized using a standard scaler, and categorical variables were transformed using one-hot encoding.

2.9. Data selection

To ensure model robustness and validity, an initial correlation analysis was conducted. Following data preparation, variables were selected for inclusion in the model. Selected variables included BW, term/preterm birth, mode of delivery (vaginal vs. cesarean), NICU admission, sex, twin-birth, maternal age, parity, 5' Apgar score, feeding at discharge (exclusive vs. non-exclusive breastfeeding), maternal smoking, and paternal smoking.

2.10. Risk prediction model development

Several classification models, including logistic regression, support vector machine, decision tree classifier, extra tree classifier, and random forest classifier, were used to explore the relationships between variables when classifying the three target conditions: colic, constipation, and regurgitation [18]. Random Forest outperformed the others due to its ability to effectively manage tabular data, handle non-linear variables, and provide robust results with unbalanced datasets. Furthermore, the model's inherent method of calculating feature importance based on impurity reduction made the results easily interpretable. Following feature selection with Random Forest, the risk prediction model was implemented.

Logistic regression was chosen due to the binary nature of the target variables (presence or absence of a disorder) [19]. However, the unbalanced nature of the dataset posed a challenge, which was addressed using both oversampling and under-sampling techniques. Multiple models were tested with different configurations, ultimately identifying BW, maternal age, and pH as the key predictor variables. Depending on the configuration, additional variables such as sex, mode of delivery, or the 5' Apgar score were included as additional predictors.

The mathematical foundation of the models is represented by the logistic regression equation:

$$p(x) = 1/(1+e\hat{}(-(\beta\_0+\beta\_1\ x\_1+\beta\_2\ x\_2+\beta\_3\ x\_3)))$$

where x_1,x_2, 〚 and x〛 _3 represent BW, maternal age, and pH respectively, and β coefficients represent the intercept and slope coefficients for each variable. These coefficients quantify the relationship between each predictor and the likelihood of disorder occurrence.

3. Results

3.1. Participants

A total of 6060 infants were enrolled in the study (52.3 % male; 6.0 % twins). Of these, 488 infants (8.1 %) were born preterm. Most infants (60.8 %) were born through vaginal delivery. Mean Apgar score at 5 min was 8.9 (SD 0.4; range 3–10), and the mean venous cord blood pH was 7.32 (SD 0.08, range 6.86–7.55).

Within the study population, the incidence of colic was 27.3 %, regurgitation 18.7 %, and constipation 10.2 %. Specifically, the incidence of colic, regurgitation, and constipation was higher among preterm infants compared to term infants (38.1 % vs 25.8 %; 35.8 % vs 17.2 %; 21.6 % vs 9.2 %; $p < 0.001$). Demographic characteristics and clinical data for term and preterm subgroups are presented in Table 1, while clinical characteristics are reported in Table 2.

The univariate analysis identified several significant neonatal factors potentially associated with the development of colic, regurgitation and constipation.

Variables which proved to be significant at the univariate analysis were included as independent variables into three different regression models (one for each FGID), whose results are detailed below.

As for colic, potential independent risk factors identified by conventional statistics were being born from a nulliparous mother ($p = 0.012$), paternal smoking ($p < 0.001$), a low venous cord blood pH ($p < 0.001$), NICU admission at birth ($p < 0.001$), and the use of any formula feeding at hospital discharge ($p = 0.012$).

As for regurgitation, potential independent risk factors were being born from a nulliparous mother ($p = 0.008$), maternal smoking ($p < 0.001$), twin status ($p = 0.016$), a low venous cord blood pH ($p < 0.001$), NICU admission at birth ($p < 0.001$), and the use of any formula feeding at hospital discharge ($p < 0.001$).

As for constipation, potential independent risk factors were maternal smoking, a low venous cord blood pH, NICU admission at birth, and the use of any formula feeding at hospital discharge ($p < 0.001$ for all potential risk factors).

It is important to note that statistical significance does not necessarily imply that these variables are essential for prediction. The predictive utility of these factors may depend on the context and can vary when using advanced methods like ML. ML approaches are expected to consider additional variables and complex interactions to improve prediction accuracy and reliability.

3.2. Machine learning for risk prediction

The correlation matrix revealed no collinear variables with the target variables, though some variables showed moderately strong positive or negative correlations with each other, such as maternal age with parity, and the 5' Apgar score with BW. Conversely, NICU admission was negatively correlated with BW and term birth (Fig. 3).

The statistical analysis was conducted primarily to validate the robustness and diversity of insights obtained through the application of AI. This complementary approach highlights the advantages of using AI for more comprehensive and nuanced predictions.

The Random Forest classifier results highlighted BW, cord blood pH, and maternal age as the most critical variables in classifying all three target conditions. These variables provide important insights into the health status of the infant and are influential in determining the likelihood of developing colic, constipation, and regurgitation (Fig. 4).

3.3. Risk prediction model

The models demonstrated comparable accuracy across different configurations; however, the model with the fewest variables was the most efficient. Consequently, BW, maternal age, and pH were selected as the core predictor variables.

Given the unbalanced nature of the dataset, a weighted average f1-score was employed for model evaluation. Among the models tested, the Random Forest classifier achieved the highest weighted average f1-score (0.74), precision (0.79), and recall (0.72). This model, therefore, outperformed other models and was implemented as the final risk prediction model.

3.4. Colic risk prediction model

The logistic regression model for predicting colic showed an intercept of −0.0068 and coefficients for birth weight (BW), maternal age, and cord blood pH of −0.0312, −0.02997, and −0.2147, respectively. The negative coefficients indicate that as these variables

Table 1
Demographic characteristics.

Variables	Term (n = 5572)	Preterm (n = 488)
Gestational age, days, mean (SD)	39 (1.1)	34.6 (2.1)
Birth wight, g, mean (SD)	3286 (421.7)	2368.4 (581)
Male, n (%)	2776 (49.8)	258 (51.8)
Twins, n (%)	291 (5.2)	50 (10.2)
Vaginal birth, n (%)	3124 (56.1)	96 (19.7)
Apgar Score 5′, median (IQR)	9 (0)	9 (1)
Cord blood pH, mean (SD)	7.32 (0.08)	7.32 (0.08)

Table 2
Clinical data.

Variables	Term (n = 5572)	Preterm (n = 488)
NICU admission, n (%)	216 (3.9)	303 (63.1)
Exclusive breastfeeding at discharge, n (%)	2695 (48.4)	173 (35.5)
Maternal age at delivery, years, mean (SD)	31.6 (5.7)	32.1 (6.5)
Multiparity, n (%)	2910 (52.2)	179 (36.7)
Maternal smoking, n (%)	739 (13.3)	99 (22.3)
Paternal smoking, n (%)	1345 (24.1)	141 (28.9)
Colic, n (%)	1437 (25.8)	186 (38.1)
Regurgitation, n (%)	939 (17.2)	171 (35.8)
Constipation, n (%)	502 (9.2)	103 (21.6)

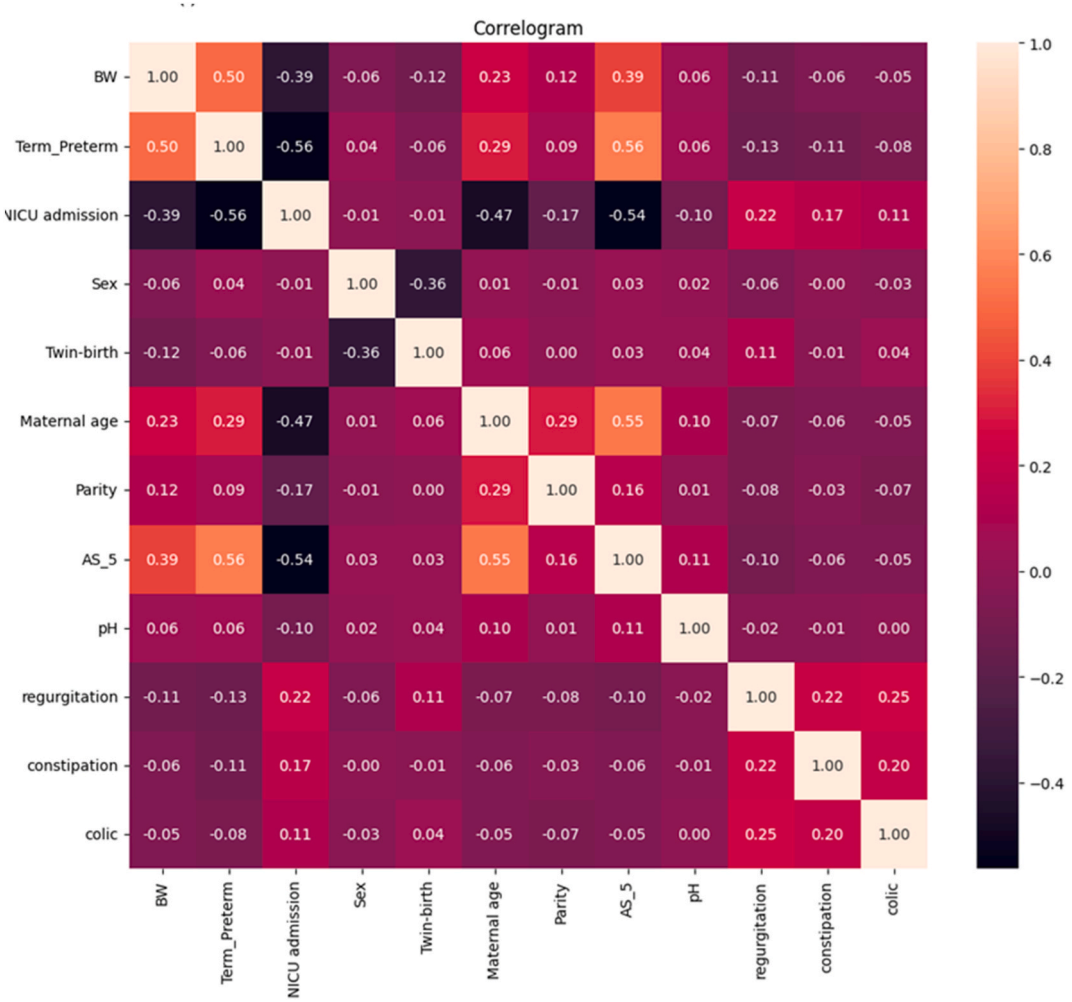

Fig. 3. Correlation Matrix showing the correlation between any pair of variables.

increase, the likelihood of colic decreases, suggesting they may protect against colic. However, the small coefficients for BW and maternal age suggest that their individual impact on colic risk may be limited, despite statistical significance. Clinicians should note that cord blood pH may be a more influential factor in assessing colic risk due to its higher coefficient.

3.5. Regurgitation risk prediction model

In the model predicting regurgitation, the intercept was −0.0313, with coefficients for BW at −0.1941, maternal age at −0.0890, and cord blood pH at −0.1891. The larger negative coefficients compared to the colic model suggest a stronger inverse relationship with regurgitation risk. BW and pH were the most significant predictors, indicating that higher values in these factors may reduce the likelihood of regurgitation.

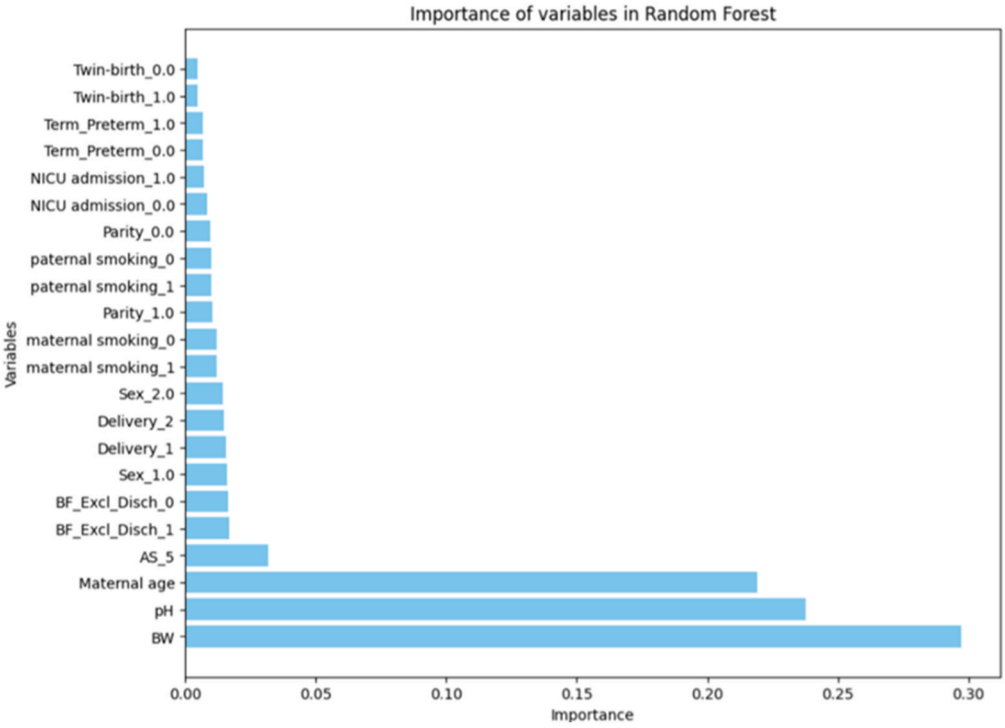

Fig. 4. Dichotomous variables state the presence of a given characteristic ('1' value) or its absence ('0' value). For Sex and Delivery, there is a different encoding: i.e., value 1 represents male sex and vaginal delivery, respectively, while value 2 represent female sex and caesarean delivery. Comparison of the importance of variables in the Machine Learning model.

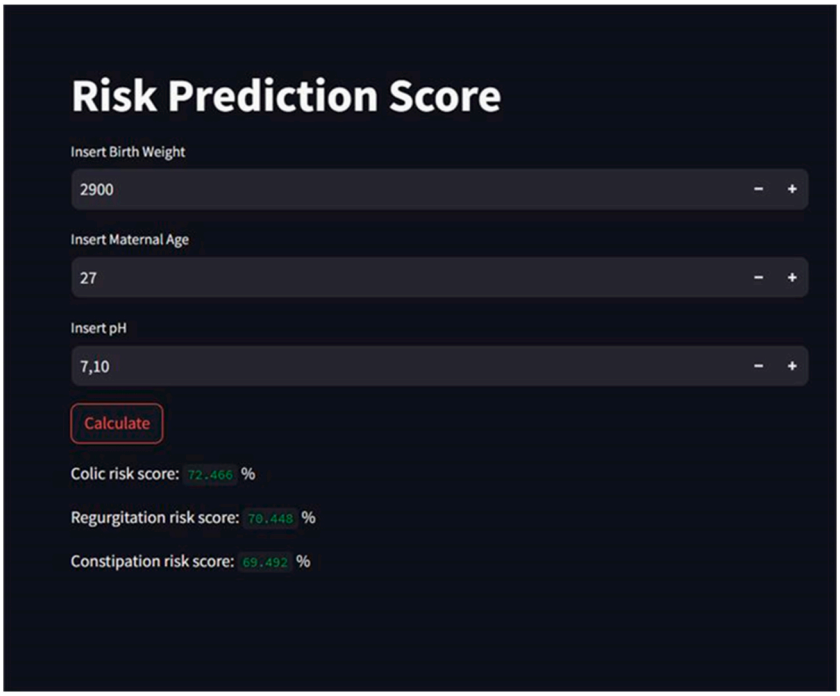

Fig. 5. Risk prediction score web interface.

3.6. Constipation risk prediction model

The model for constipation had an intercept of −0.0201 and coefficients for BW, maternal age, and cord blood pH of −0.1397, −0.0179, and −0.1663, respectively. BW and pH were significant predictors, both negatively associated with constipation risk. The small negative coefficient for maternal age suggests a weaker association with constipation compared to BW and cord blood pH. The substantial negative relationship with pH, consistent across models, suggests that cord blood pH could be an important predictor of gastrointestinal health in neonates.

3.7. Development of a risk prediction tool

Based on the predictive model, a practical risk assessment tool was developed to assist clinicians in identifying infants at high risk of developing FGIDs. The tool's design is user-friendly, providing a simple interface where clinicians can input key variables (BW, maternal age, pH) to obtain a risk score for colic, regurgitation, or constipation (Fig. 5).

4. Discussion

Interpretation In our study, for the first time in the literature, we identified several risk factors for the development of FGIDs in infants and toddlers using two different methodological approaches: conventional statistics and ML converged in identifying some potential risk factors, while diverged for other variables. Additionally, we developed a ML predictive model (FRIP) for the early diagnosis of FGIDs in children. Healthcare practitioners can input patient data into designated fields, and the interface processes this information through logistic regression models. The output is a set of risk coefficients, each corresponding to one of the three conditions, representing the likelihood of a patient developing a particular condition. This assists healthcare providers in making informed decisions about diagnosis, prevention, and treatment strategies.

The prediction score results indicate the probability (as a percentage) of the occurrence of one of the three disorders. Notably, instead of providing exact values for each risk factor, the result is generated by an equation that integrates all the factors. Traditional statistics evaluate the significance of each observed variable in relation to a disease. In this study, we utilized AI tools for feature selection to identify the minimum number of variables required for accurate prediction, evaluating their combined influence on overall disease occurrence (e.g., colic). Specifically, we found that only three variables are needed to reliably predict the risk score, and these variables can thus be considered risk factors. According to our dataset, birth weight emerged as the main risk factor based on statistical analysis, while the trained AI model identified cord blood pH and maternal age as important variables for risk prediction.

Accurate identification of early life events is crucial for identifying children at risk of developing FGIDs, enabling timely intervention and improving the quality of life for both children and their families. However, the absence of a universally accepted gold standard often leads to overestimation or underestimation of potential early-life risk factors.

In our study population, the incidence of colic was 27.3 %, regurgitation 18.7 %, and constipation 10.2 %. In the preterm infant subgroup, the incidence was significantly higher at 38.1 %, 35.8 %, and 21.6 %, respectively, compared to term infants. The incidence of FGIDs has been reported with wide variability across different studies [4,5,20,21]. This variability can be attributed to differences in diagnostic criteria and poor population stratification.

For instance, a cross-sectional Brazilian study evaluating FGIDs in the first two years of life found no significant difference between preterm and term infants [13].

However, a separate study from Türkiye reported a higher prevalence of regurgitation, infantile colic, and dyschezia in preterm infants during the first 12 months compared to term infants [14].

The perinatal period, during which the brain-gut-microbiota axis matures, is critical as various determinants during this time can have significant long-term consequences. Preterm infants are at high risk of both gastrointestinal and brain maturation impairments due to intrinsic immaturity and their unique developmental environment, including impaired gut microbiota assembly [22,23]. These factors likely contribute to the development of FGIDs [24].

Our results indicate that low venous cord blood pH is a risk factor for FGIDs, confirming previous findings [25]. Umbilical cord pH serves as a marker of neonatal metabolic and oxygenation status at birth, with low pH levels (acidosis) indicating potential fetal distress. While the link between neonatal acidemia and neurological problems is well established, there is limited information on the impact of low venous cord blood pH on gastrointestinal diseases [26]. We speculate that this condition may disrupt gastrointestinal development, alter gut microbiota, and affect the enteric nervous system, increasing the risk of FGIDs. Hypoxia–ischemia and reperfusion in the brain are known to trigger harmful events leading to neuronal death, while in the gastrointestinal tract, hypoxia increases TLR4 expression in the intestinal mucosa, causing an imbalance between pro-inflammatory and anti-inflammatory signals, thereby increasing mucosal susceptibility to harmful stimuli [27,28]. These findings suggest that increased monitoring of infants with low venous cord blood pH at birth could be beneficial, potentially allowing for early detection of an increased risk of FGIDs during the first year of life.

Finally, maternal age further contributes to this risk, as younger mothers may face higher rates of pregnancy complications or stress, while older mothers are more likely to experience conditions such as gestational diabetes or hypertension, which can impact fetal development. Additionally, maternal microbiota changes associated with age may influence neonatal gut health during delivery, highlighting the multifactorial nature of FGID risk.

4.1. Limitations

This study has several limitations that should be acknowledged. First, only a limited set of key variables were selected for analysis, which may have excluded other important risk factors relevant to the development of Functional Gastrointestinal Disorders (FGID), such as antibiotic use and birth weight classification [12,29]. However, the primary objective of this study was to leverage AI techniques to identify a subset of variables that could provide an accurate prediction of a newborn's risk of developing FGID, with the understanding that additional variables may be explored in future research.

Furthermore, the study was conducted at a single hospital in Italy, which limits the generalizability of the results to other healthcare settings or populations. Additionally, the research was carried out during the peak of the SARS-CoV-2 pandemic, a period marked by significant disruptions in healthcare delivery and daily life. The inclusion of subjects during this time, and in the subsequent post-pandemic period, introduces a potential bias, as the exceptional circumstances of the pandemic may have influenced both parental behaviors and environmental factors that could, in turn, affect the development of FGID.

Finally, long-term health outcomes were not tracked, preventing us from assessing the impact of the predictive tool on the long-term health of the neonates. These limitations underscore the need for further studies to validate these findings, explore other potential risk factors, and assess the long-term effectiveness and applicability of the tool.

Usability of the model in the context of current care.

Our study provides new insights into early-life risk factors for FGIDs, underscoring the importance of perinatal and neonatal factors in predicting these disorders. By integrating AI methods, we developed a robust risk prediction model and a practical assessment tool to assist clinicians in identifying at-risk infants, enabling early interventions to mitigate the impact of FGIDs.

CRediT authorship contribution statement

Flavia Indrio: Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Investigation, Data curation, Conceptualization. **Elio Masciari:** Writing – review & editing, Writing – original draft, Investigation, Formal analysis, Data curation, Conceptualization. **Flavia Marchese:** Writing – original draft, Investigation. **Matteo Rinaldi:** Investigation. **Gianfranco Maffei:** Investigation. **Ilaria Gangai:** Investigation. **Assunta Grillo:** Investigation. **Roberta De Benedetto:** Investigation. **Enea Vincenzo Napolitano:** Investigation. **Isadora Beghetti:** Investigation. **Luigi Corvaglia:** Writing – original draft, Investigation, Data curation. **Antonio Di Mauro:** Writing – review & editing. **Arianna Aceti:** Writing – review & editing, Writing – original draft, Supervision, Formal analysis, Data curation, Conceptualization.

Informed consent statement

Patient consent was obtained from all the participants.

Institutional Review Board statement

The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of Ethical Committee Regione Puglia Protocol Number 3465/2021.

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, we used ChatGPT service in order to improve language and readability. After using this service, we reviewed and edited the content as needed and take full responsibility for the content of the publication.

Funding

This work was carried out as part of the project "AN APP TO SHED LIGHT ONTHE WINDOW OF OPPORTUNITY OFTHE FIRST 1000 DAYS OF LIFE", funded under the PRIN: Projects of Relevant National Interest – 2022 Call (Prot. 202247M2AP), with financial support from the European Union – NextGenerationEU.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e41516.

References

[1] J.S. Hyams, C. Di Lorenzo, M. Saps, R.J. Shulman, A. Staiano, M. van Tilburg, Functional disorders: children and adolescents, Gastroenterology 150 (6) (2016) 1456–1468.e2, https://doi.org/10.1053/j.gastro.2016.02.015.

[2] Y. Vandenplas, B. Hauser, S. Salvatore, Functional gastrointestinal disorders in infancy: impact on the health of the infant and family, Pediatr Gastroenterol Hepatol Nutr 22 (3) (2019) 207–216, https://doi.org/10.5223/pghn.2019.22.3.207. PMID: 31110953; PMCID: PMC6506429.

[3] M.A. Benninga, C. Faure, P.E. Hyman, I. St James Roberts, N.L. Schechter, S. Nurko, Childhood functional gastrointestinal disorders: neonate/toddler, Gastroenterology 150 (6) (2016) 1443–1455.e2, https://doi.org/10.1053/j.gastro.2016.02.016. PMID: 27144631.

[4] Y. Vandenplas, A. Abkari, M. Bellaiche, M. Benninga, J.P. Chouraqui, F. Çokura, et al., Prevalence and health outcomes of functional gastrointestinal symptoms in infants from birth to 12 Months of age, J. Pediatr. Gastroenterol. Nutr. 61 (5) (2015) 531–537, https://doi.org/10.1097/MPG.0000000000000949. Erratum in: J Pediatr Gastroenterol Nutr 2016;62(3):516. PMID: 26308317; PMCID: PMC4631121.

[5] L.A. Lestari, A.N. Rizal, W. Damayanti, Y. Wibowo, C. Ming, Y. Vandenplas, Prevalence and risk factors of functional gastrointestinal disorders in infants in Indonesia, Pediatr Gastroenterol Hepatol Nutr 26 (1) (2023) 58–69, https://doi.org/10.5223/pghn.2023.26.1.58. PMID: 36816434; PMCID: PMC9911174.

[6] N.F. Steutel, J. Zeevenhooven, E. Scarpato, Y. Vandenplas, M.M. Tabbers, A. Staiano, et al., Prevalence of functional gastrointestinal disorders in European infants and toddlers, J. Pediatr. 221 (2020) 107–114, https://doi.org/10.1016/j.jpeds.2020.02.076.

[7] A. Chogle, C.A. Velasco-Benitez, I.J. Koppen, J.E. Moreno, C.R. Ramírez Hernández, M. Saps, A population-based study on the epidemiology of functional gastrointestinal disorders in young children, J. Pediatr. 179 (2016) 139–143, https://doi.org/10.1016/j.jpeds.2016.08.095, e1.

[8] M.A. van Tilburg, P.E. Hyman, L. Walker, A. Rouster, O.S. Palsson, S.M. Kim, et al., Prevalence of functional gastrointestinal disorders in infants and toddlers, J. Pediatr. 166 (3) (2015) 684–689, https://doi.org/10.1016/j.jpeds.2014.11.039.

[9] G. Holtmann, A. Shah, M. Morrison, Pathophysiology of functional gastrointestinal disorders: a holistic overview, Dig. Dis. 35 (Suppl 1) (2017) 5–13, https://doi.org/10.1159/000485409.

[10] I.J.N. Koppen, M.A. Benninga, M.M.J. Singendonk, Motility disorders in infants, Early Hum. Dev. 114 (2017) 1–6, https://doi.org/10.1016/j.earlhumdev.2017.09.005.

[11] R. Shamir, I. St James-Roberts, C. Di Lorenzo, A.J. Burns, N. Thapar, F. Indrio, et al., Infant crying, colic, and gastrointestinal discomfort in early childhood: a review of the evidence and most plausible mechanisms, J. Pediatr. Gastroenterol. Nutr. 57 (Suppl 1) (2013), https://doi.org/10.1097/MPG.0b013e3182a154ff.

[12] S. Salvatore, M.E. Baldassarre, A. Di Mauro, N. Laforgia, S. Tafuri, F.P. Bianchi, et al., Neonatal antibiotics and prematurity are associated with an increased risk of functional gastrointestinal disorders in the first year of life, J. Pediatr. 212 (2019) 44–51, https://doi.org/10.1016/j.jpeds.2019.04.061.

[13] M.M.B.B. Gondim, A.L. Goulart, M.B. Morais, Prematurity and functional gastrointestinal disorders in infancy: a cross-sectional study, Sao Paulo Med. J. 140 (4) (2022) 540–546, https://doi.org/10.1590/1516-3180.2021.0622.R1.29102021.

[14] D. Bi, H. Jiang, K. Yang, T. Guan, L. Hou, G. Shu, Neonatal risk factors for functional gastrointestinal disorders in preterm infants in the first year of life, Turk. J. Pediatr. 65 (6) (2023) 919–930, https://doi.org/10.24953/turkjped.2022.1089.

[15] K. Kamphorst, E. Van Daele, A.M. Vlieger, J.G. Daams, J. Knol, R.M. van Elburg, Early life antibiotics and childhood gastrointestinal disorders: a systematic review, BMJ Paediatr Open 5 (1) (2021), https://doi.org/10.1136/bmjpo-2021-001028.

[16] H. Liang, B.Y. Tsui, H. Ni, C.C.S. Valentim, S.L. Baxter, G. Liu, et al., Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence, Nat Med 25 (3) (2019) 433–438, https://doi.org/10.1038/s41591-018-0335-9.

[17] H. Lonsdale, A. Jalali, L. Ahumada, C. Matava, Machine learning and artificial intelligence in pediatric research: current state, future prospects, and examples in perioperative and critical care, J. Pediatr. 221S (2020), https://doi.org/10.1016/j.jpeds.2020.02.039.

[18] K.J. Archer, R.V. Kimes, Empirical characterization of random forest variable importance measures, Comput. Stat. Data Anal. 52 (4) (2008) 2249–2260, https://doi.org/10.1016/j.csda.2007.08.015.

[19] M.E. Shipe, S.A. Deppen, F. Farjah, E.L. Grogan, Developing prediction models for clinical use using logistic regression: an overview, J. Thorac. Dis. 11 (Suppl 4) (2019), https://doi.org/10.21037/jtd.2019.01.25.

[20] O.F. Beser, F. Cullu Cokugras, G. Dogan, O. Akgun, M. Elevli, P. Yilmazbas, et al., The frequency of and factors affecting functional gastrointestinal disorders in infants that presented to tertiary care hospitals, Eur. J. Pediatr. 180 (2021) 2443–2452, https://doi.org/10.1007/s00431-021-04027-w.

[21] T.G. Dinan, J.F. Cryan, The microbiome-gut-brain Axis in health and disease, Gastroenterol Clin North Am 46 (1) (2017) 77–89, https://doi.org/10.1016/j.gtc.2016.09.007.

[22] I. Beghetti, M. Barone, P. Brigidi, A. Sansavini, L. Corvaglia, A. Aceti, et al., Early-life gut microbiota and neurodevelopment in preterm infants: a narrative review, Front. Nutr. 10 (2023) 1241303, https://doi.org/10.3389/fnut.2023.1241303.

[23] C. de Weerth, S. Fuentes, P. Puylaert, W.M. de Vos, Intestinal microbiota of infants with colic: development and specific signatures, Pediatrics 131 (2) (2013), https://doi.org/10.1542/peds.2012-1449.

[24] C.A. Velasco-Benitez, C.H. Axelrod, S. Gutierrez, M. Saps, The relationship between prematurity, method of delivery, and functional gastrointestinal disorders in children, J. Pediatr. Gastroenterol. Nutr. 70 (2) (2020), https://doi.org/10.1097/MPG.0000000000002543.

[25] F. Indrio, F. Marchese, M. Rinaldi, G. Maffei, V. Dargenio, R. Cinquepalmi, et al., Is acidemia at birth a risk factor for functional gastrointestinal disorders? Eur. J. Pediatr. 181 (10) (2022) 3625–3633, https://doi.org/10.1007/s00431-022-04565-x. PMID: 35927380; PMCID: PMC9508039.

[26] G.L. Malin, R.K. Morris, K.S. Khan, Strength of association between umbilical cord pH and perinatal and long term outcomes: systematic review and meta-analysis, BMJ 340 (2010), https://doi.org/10.1136/bmj.c1471.

[27] S.C. Gribar, R.J. Anand, C.P. Sodhi, D.J. Hackam, The role of epithelial Toll-like receptor signaling in the pathogenesis of intestinal inflammation, J. Leukoc. Biol. 83 (3) (2008) 493–498, https://doi.org/10.1189/jlb.0607358.

[28] D.I. Matara, A. Pouliakis, T. Xanthos, R. Sokou, G. Kafalidis, Z. Iliodromiti, et al., Microbial translocation and perinatal asphyxia/hypoxia: a systematic review, Diagnostics 12 (1) (2022) 214, https://doi.org/10.3390/diagnostics12010214.

[29] M.E. Baldassarre, A. Di Mauro, S. Salvatore, S. Tafuri, F.P. Bianchi, E. Dattoli, L. Morando, L. Pensabene, F. Meneghin, D. Dilillo, V. Mancini, V. Talarico, F. Tandoi, G. Zuccotti, M. Agosti, N. Laforgia, Birth weight and the development of functional gastrointestinal disorders in infants, Pediatr Gastroenterol Hepatol Nutr 23 (4) (2020 Jul) 366–376, https://doi.org/10.5223/pghn.2020.23.4.366.