Computer aided prediction of sepsis-related mortality risk in neonatal intensive care units.

Sergio Muñoz Lezcano [1], Fernando López[2] and Alberto Corbi[3]

[1] PhD student of the Program in Computer Science, Universidad Internacional de La Rioja (UNIR), Avenida de La Paz, 137, 26006, Logroño (La Rioja) Spain. smunozle@gmail.com

[2] Associate Professor, Universidad Internacional de La Rioja (UNIR), Avenida de La Paz, 137, 26006, Logroño (La Rioja) Spain. fernando.lopez@unir.net

[3] Research Institute for Innovation & Technology in Education (UNIR iTED), Universidad Internacional de La Rioja (UNIR), Avenida de La Paz, 137, 26006, Logroño (La Rioja) Spain. alberto.corbi@unir.net

Abstract. Nowadays, sepsis is considered a global burden disease with an annual incidence of three million neonatal cases. Nevertheless, there are no homogeneous criteria for neonatal sepsis. Furthermore, adult sepsis scores don't work properly in neonatal Intensive Care Units (ICUs) settings due to the specific characteristics of the neonates' immune systems. This work describes and surveys a machine-learning computer-aided diagnosis approach for predicting sepsis mortality in neonatal ICUs. The survey is based on a retrospective cohort study in which each patient has an initial sepsis-related diagnosis in the first 24h after ICU admission. Our experiments are based on four different machine-learning techniques: decision trees, random forests, support vector machines and artificial neural networks. The predictive power was assessed using accuracy, sensitivity, and specificity. The importance of the variables was obtained automatically through data science techniques using R. The approach with the best performance was the random forest, which achieves an accuracy of 97% in the prediction of mortality.

Keywords: Sepsis, CAD, NICU, MIMIC-III, Artificial Intelligence.

1 Introduction

Computer-aided diagnosis (CAD) based on machine-learning (ML) can already be leveraged to predict sepsis mortality in patients admitted [1] in Intensive Care Units (ICUs). The same principle applies to Neonatal ICUs (NICUs) where early infant exitus are often caused by sepsis [2]. In this context, a CAD tool can assist clinicians in making medical decisions based on historical case results that the machine has learned.

The neonatal stage starts with the patient's date of birth and terminated 28 days after birth [3]. If the baby is born prematurely, the delivery of care in NICUs could continue beyond the first 28 days [4]. The provision of care in NICUs is known to be especially complex in healthcare organizations where the decisions and results can be influenced by factors such as technology, people, setting, logistics, skills, culture, and experience of professionals [5]. Because of these factors, healthcare professionals working at NICUs suffer a high level of stress that has a direct impact in the attention given to neonates [6]. The continuous fatigue experienced by doctors and nurses in these settings may reduce the sepsis diagnosis accuracy [7], treatment definition and patients' monitoring. The situation worsens taking into consideration that the sepsis scoring systems for neonates differ in small but fundamental details from adult sepsis [8].

Sepsis in neonates presents the highest disease incidence among all age groups of patients with three million of cases by year [9]. In 2018, almost 15% of neonatal deaths (375,000) were caused by sepsis [10]. In 2001 a task force of 19 critical care clinicians defined sepsis as a "life-threatening organ dysfunction caused by a dysregulated response to infection" [11] but there is not a homogeneous criteria for neonatal sepsis [12]. Based on this definition, clinicians have developed scores tools that help to get more accurate diagnosis in adults and pediatrics patients such as the Sequential Organ Failure Assessment Score (SOFA), quick SOFA (qSOFA) and pediatrics SOFA (pSOFA) [13][14]. Compared to adults, the immaturity of neonates' immunological systems [15] and the potential exposure to infections at the intrauterine level, should involve different diagnoses and risk assessment to reach better short- and long-term outcomes [12]. Since these patients require a specific tools, clinicians have designed neonatal SOFA (nSOFA) score to predict mortality in the in the NICUs because of Late-onset sepsis (LOS) [16].

Sepsis situations emerged because of the quality of the received medical care, state of the healthcare infrastructure (national, regional, and local), lack of prevention, and deficient management of resources. LOS occurs after the first weeks of life, and it is associated with contaminated medical equipment, environmental pathogens, or obstetric complications. Early-onset sepsis (EOS) [17] appears during the first days of life and it is associated with gynecologic complications and presents a quicker disease progression that may involve multi-organ failure. Attending to incidence and mortality for neonatal sepsis, the overall figures are higher for EOS than LOS [18]. Other factors, such as early delivery (25–32 weeks), low-birthweight (<1,500g), immaturity of the immune system and scarcity of blood volume (below 13.6 ml/kg) [19], also contribute to the child death rate. The complexity of this equation is further increase by common symptoms [2], Diversity of pathogens typically found at NICUs [20] and Diagnostic variables [13] used for sepsis prediction. The combination of those factors makes it difficult to find a formula to predict sepsis [21].

ML is especially effective in making predictions with non-linear factors. ML and data science techniques have already been applied to sepsis recognition through predictive models [22]. At this point, some alterations of vital signs such as heart rate (HR), respiratory rate (RR), and oxygen saturation from pulse oximetry (SpO2) have been identified as sepsis precursors in patients admitted to the NICU [23]. Scientists have also developed predictive algorithms by merging heterogeneous datasets acquired and documented at the NICUs. Other studies [24] recommend the inclusion of sociodemographic, obstetric, neonatal, and maternal infectious risk factors. Considering everything discussed so far, the aim of this study has been to survey the performance of ML classifiers to predict EOS sepsis mortality. We have surveyed decision trees (DT), random forests (RF), support vector machines (SVM), and artificial neural networks (ANN) focused on mortality prediction in very young patients admitted at the NICUs. The dataset Medical Information Mart for Intensive Care (MIMIC-III) described below was used as the main source of data for training/testing purposes of this research.

2 Patients and Methods

The clinical data used in this study were obtained from "Medical Information Mart for Intensive Care" (MIMIC-III) v.1.4 [25] developed by the Laboratory for Computational Physiology at Massachusetts Institute of Technology (MIT). The MIMIC-III database has been populated with information from "Beth Israel Deaconess Medical Center in Boston", United States (U.S) gathered from several sources such as the archives from critical care information systems, hospital electronic health record databases and the "U.S Social Security Administration Death Master File" [26]. In particular, MIMIC-III contains 53,423 hospital admissions for patients above 16 years of age admitted from 2001. MIMIC-III also includes 7,870 hospital admissions for neonates (from 2001 to 2008). This database is publicly and freely available and the information is anonymized to comply with the safeguarding patient privacy, following HIPAA deidentification standards [27]. The information related to dates has been also randomly altered for the patients in the database.

Bearing in mind the available information in MIMIC-III and the EOS [23] time frame characteristics, the initial cohort of patients grouped patients less than 12 months old who were admitted in the NICU. At the same time, the survey considered a range of time between 6 to 24 first hours of stay. With this premise, the study reaches a doble objective: prevent risk of bias from further clinical management and consider the inception of proinflammatory secretions [28]. The second criteria considered the available status/cause codified in the International Classification of Diseases or ICD-9-CM codes associated with septicemia of newborn being susceptible of sepsis diagnosis [29]. As a result of the selection criteria (based on structured query language (SQL) on top of PostgreSQL Database), a total of 247 neonates matched these conditions. A third filter was applied to this group of patients taking into consideration data related to markers of infection [30] collected from charted events (e.g., notes, laboratory tests, and fluid balance) compiled from the NICU clinical systems, Philips Care Value and iMDSoft MetaVision [25]. Vital signs measurements precursors of sepsis [2] were also extracted, including heart rate, admission weight and temperature.

The final filter considers all those patients with true positive results in pathogens associated with sepsis [31] considering the available information in MIMIC-III. In line with all the previous steps, the final cohort classifies 112 patients that was randomly split in two datasets for training (80%) and testing (20%) purposes looking for the sepsis mortality prediction. To reach this goal, the survey applies different ML techniques. In particular, we have surveyed the following classification algorithms: DT [32], RF [33], SVM [34], ANN [35].

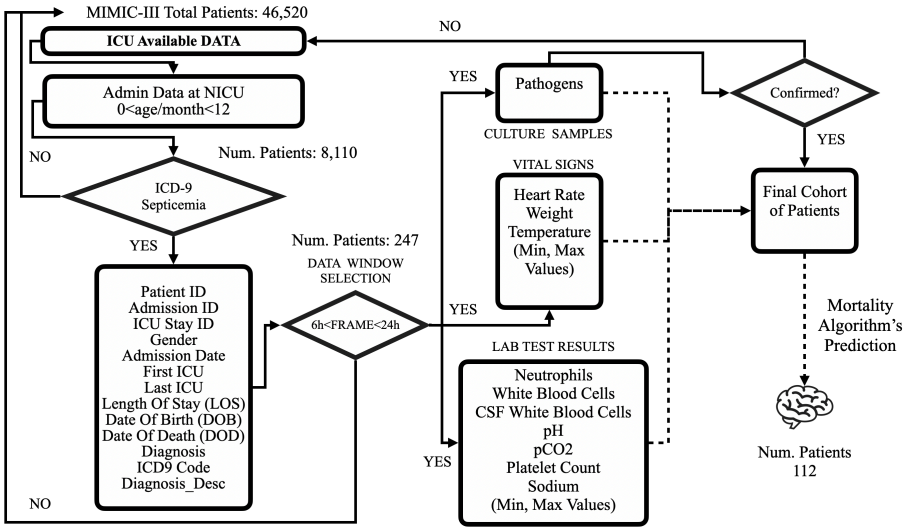

Fig. 1. CAD Sepsis Flow. Selection of variables regarding with MIMIC-III Database linked with the purpose of the analysis. First step, from the whole cohort of patients (46,520), the initial group of preliminary patients is filter base on the age of the patient (8,110). The number of patients is reduced filtering by diagnosis associated with ICD-9-CM septicemia codes. According with this premise, the number of patients is reduced to an intermediate cohort of 247. Vital signs, clinical and laboratory variables are added to the initial group (basically associates to demographic characteristics) considering the range of time between the six first hours in NICU and 24 hours. The final cohort of patients consider all those patients with a positive pathogen analysis.

3 Results

3.1 Decisions Trees

Decision Trees selects the most correlated variables. In this way, the most important variables are the number of days in the ICU (level 1), the minimum values from weight and maximum values from pH in terms of hydrogen ion (H+) concentration (level 2) and finally, the pathogen labeling in the third level of the tree's leaves. We pruned the tree to reduce the number of original branches (two levels) maintaining the length of stay (LoS) at NICU as the first layer in the tree. The minimum count of neutrophils was maintained as a second level of variables in the hierarchy of the tree while the rest of

variables were removed. This algorithm's conclusion is coherent with the risk of infection. The prune procedure did not show better results. The obtained accuracy (87%), sensitivity (89%) and specificity (66%) are surprisingly good.

3.2 Random Forest

In this case, the algorithm implementation has been carried out with 500 classification trees. As it is shown in Fig. 2, the mean decrease in Gini [36] coefficient identified, for each variable, how important was this coefficient attending to classifying data purposes. Moreover, Gini identified the contribution of each variable related to the whole random forest homogeneity, considering nodes and leaves of our model. In this way, the higher the value of the mean decreases in the Gini score, the higher the importance of the variable in our model.

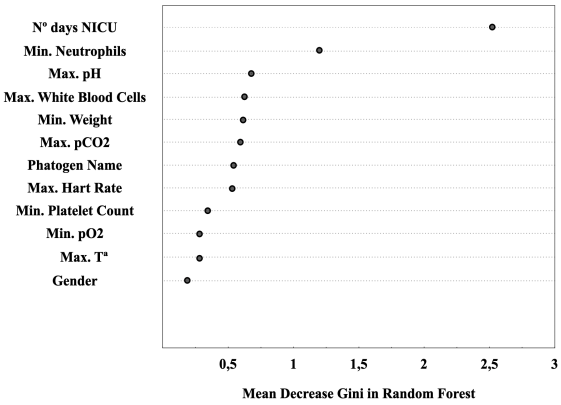

Fig. 2. Random Forest variable importance. The length of stay at NICU, the minimum count of neutrophils, maximum level of pH, maximum count of white blood cells, and minimum patient's weight are the most relevant variables in the analysis. This group of variables are the most correlated taking into consideration the correlation matrix and clinical literature that supports the suspected precursors of sepsis

The variables in the analysis, in descending order of relevance are LoS, the minimum count of neutrophils, the maximum level of pH, the maximum count of white blood cells, and the minimum patient's weight are the most relevant variables in the analysis. This group of variables are the most correlated taking into consideration the correlation matrix and clinical literature that supports the suspected precursors of sepsis. The implementation of this algorithm reached a high result of accuracy (96%), sensitivity (96%) and specificity (100%). The implementation was modified to increase the number of trees (1000) but results from this model suggests a possible over-fitting due the low volume of data. In this case, the implementation with 500 trees was considered the best option.

3.3 Support Vector Machines

Support Vector Machines were implemented with four personalized kernel approaches: polynomial, lineal (vanilla), hyperbolic and radial. Taking into consideration these assertions, even when performance reached by the hyperbolic and radial configuration present good results as shows the Table I, the radial kernel was chosen to optimize the algorithm's performance (accuracy, 92%; sensitivity, 100% and specificity, 0%). Since the performance of the SVM algorithm depends on their hyperparameter settings, the kernel was customized according to Cost (C) and Sigma hyperparameter values [37]. This customization prevents generalization issues that are related to the theoretical predictions goals that don't commonly take place in real-world cases. In this way, the hyperparameters C Cost (5.18e+05) and Sigma (7.2e-08) were customized for this purpose but the implementation did not reach better results (accuracy, 87%).

3.4 Artificial Neural Network

Artificial Neural Networks were the final approach in this study. To tackle the problem with ANN, the expected result (mortality prediction) was compared with different implementations and network complexities. The initial and most simple architecture is based on a logistic activation function. This simple approach shows good results in terms of accuracy (83%) but, to increase the optimization of the architecture, it was modified with different and more complex networks. The best results were reached with an architecture of three layers and four neurons per each. Considering this architecture, it was necessary to set the ANN hyper-parameters that can significantly reduce improves the performance. The learning rate (LR) is a customizable hyperparameter that manages the speed of the model's adaptation to the problem. The LR usually has a small positive value in the range between 0.0 and 1.0. In this configuration, the learning rate parameter was modified with an exponential approach (0.1, 0.01, 0.001, 0.0001). Taking into consideration these changes, the accuracy was improved to 88% as the Table I shows. From this point of view, the accuracy has reached an optimal level of performance.

This research has assessed the feasibility of a CAD approach for sepsis-related mortality prediction. In our experiments, the RF classifier achieved the maximum performance: 97% of accuracy, 95% of specificity and 100% of sensitivity. Note that sensitivity (100%) indicates how well the classifier predicts sepsis, while specificity (95%) indicates how well the classifier predicts healing. Besides the performance obtained with RF, the accuracy reported by the rest of the models is very similar: between 88% and 96% as it shown in Table 1. These percentages can be considered as reliable results considering that a good clinical predictor should have a predictive power greater than 80-85% [37][38]. The results confirm that LOS, blood neutrophil concentrations, pH, white blood cells, pCO2 and weight are the most significative variables to predict mortality.

Table 1. Algorithms results for sepsis mortality prediction

	Algorithm in test scenario						
	Trees		Random Fores500 Trees	SVM			
	No Pruned	Pruned		Polynomial	Linear	Hyperbolic	Radial
Accuracy	0.87	0.87	0.96	0.88	0.88	0.92	0.92
Precision	0.96	0.96	1	0.92	0.92	0.92	0.92
Sensitivity	0.89	0.89	0.96	0.95	0.95	1	1
Specificity	0.66	0.66	1	0	0	0	0

4 Conclusion

Since neonates have a low volume of blood, it is difficult to get a high number of laboratory test results. For this reason, we have also considered other vital signs such as blood pressure and temperature even when these variables have proven not to be so relevant according with the algorithm's correlations. Quite interestingly, the pathogen cultures do not show strong evidence of correlation. This would entail that the type of microorganism behind each sepsis episode would not have seen a sharp impact on the patient's outcome, bearing in mind the algorithms' results. This fact may have an important consequence in scheduling, prioritization and decision making at the NICUs since the cultures usually take several hours to complete. According to our results, this step, although aprioristically crucial, can be adjourned in favor of other life-saving actions.

Even when it is difficult to establish comparations between studies due to dataset differences, cohort differences, and measurement of outcome metrics, our survey results are aligned with other studies focus on sepsis prediction in terms of specificity and sensibility applying different AI techniques but none of them focus mortality prediction:

- The dataset [38] was set from the CHOP NICU registry in Verona Hospital, Italy. The data was extracted from the Electronic Health Record (EHR) considering 618 patients that met the criteria of inclusion/exclusion (0.98 of sensitivity/0.71 of specificity).
- The dataset [39] was collected during 18 months from medical records from the Instituto Nacional de Perinatología Isidro Espinosa de los Reyes (INPerIER), Mexico. The survey considered 236 neonates hospitalized in the Neonatal Intensive Care Unit (NICU) that met the criteria of inclusion/exclusion (93.3% sensitivity/80% specificity).

The field of neonatology is one of the areas of intensive care where the amount of data is being boosted thanks to the exponential innovation in medical technology. Future works could consider these new data sources (e.g., advanced medical imaging, IoT sensors in cradles, data from sources such as pumps, electrocardiograms, etc.) as part

of the datasets to improve the reached results. Additional advancements could be attained with the addition of new promising biomarkers (e.g., PCR, ILK6 / IK8, CRP or Procalcitonin levels) [39].

For now, and as stated at the beginning of this text, CAD systems, and machine learning represent a very powerful approach to achieving an acceptable compromise between the speed and accuracy related to the diagnosis, prognosis, and applied treatments. This is especially true in frantic healthcare areas, such as the NICUs. All those variabilities considering not only demographics and laboratory variables but also genes and novel biomarkers promote an emerging approach to precision diagnostics in terms of prevention. Fulfilling this goal is necessary for a huge amount of computational resources for being able to reach an individual level from a population point of view. AI applied to risk stratification, assessment and early-markers identification will play a key role in treatments and sepsis disease complications, combining AI-based with human experts' knowledge, experience, and skills. Replacement of healthcare professionals is not feasible nowadays.

In the future, CAD, AI-based ML, and robotics will be benefited from the clinical knowledge of today but it will demand a strong collaboration for AI engineers, clinical researchers, and practitioners. In any case, even in the future, it will be necessary to deal with limitations that need to be considered as a general challenge in AI and ML, the need for more samples to fit the models properly. With not enough samples the model induces bias reducing the robustness. So, as part of future work, we will consider the use of additional datasets with more predictors for reaching better results.

4.1 Acknowledgment

The authors thank Miguel Ángel Armengol de la Hoz, Head of the Big Data Department in Regional Ministry of Health of Southern Spain, Luis Arruza, neonatologist MD, PhD at Hospital Clínico San Carlos, and Esther Aleo Luján, Pediatrician, MD, PhD at Hospital Clínico San Carlos for their advice on statistics, neonatology, and pediatrician.

References

1. Fleuren, L., et al.: Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. Intensive Care Med.46(3), 383–400 (2020).
2. Shane, A., Sánchez, P., Stoll, B.: Neonatal sepsis. Lancet 390(10104), 1770–1780 (2017).
3. Pathirana, J., et al.: Neonatal death: Case definition & guidelines for data collection, analysis, and presentation of immunization safety data. Vaccine 34(49), 6027–6037 (2016).
4. Doubova, S. V., et al.: Evaluating the quality of the processes of care and clinical outcomes of premature newborns admitted to neonatal intensive care units in Mexico. Int. J. Qual. Heal. Care 30(8), 608–617 (2018).
5. Bondurant, P. G., Nielsen-Farrell, J., Armstrong, L.: The Journey to High Reliability in the NICU. J. Perinat. Neonatal Nurs 29(2), 170–178 (2015).
6. Fogaça, M., Carvalho, W., Cítero, Nogueira-Martins, L.: Factors that cause stress in physicians and nurses working in a pediatric and neonatal intensive care unit: bibliographic review. Rev. Bras. Ter. intensiva, 20(3), 261–6 (2008).

7. Shafer, G., Suresh, G.: Diagnostic errors in the neonatal intensive care unit: A case series. Neonatal Intensive Care 32(1), 47–50 (2019).
8. Gil, J., Cebrián, M., Bello G., Diaz-Alersi, R.: Apache Ii. Intensive Care Med. 13(2), 143 (1987).
9. Fleischmann-Struzek, C., Goldfarb, D., Schlattmann, P., Schlapbach, L., Reinhart, K., Kasson, K.: The global burden of pediatric and neonatal sepsis: a systematic review. Lancet Respir. Med. 6(3), 223–230 (2018).
10. World Health Organization, Global report on the epidemiology and burden of sepsis: current evidence, identifying gaps and future directions, https://apps.who.int/iris/handle/10665/334216, last accessed 2022/08/20.
11. Singer, M., et al.: The third international consensus definitions for sepsis and septic shock (sepsis-3). JAMA - J. Am. Med. Assoc. 315(8), 801–810 (2016).
12. McGovern, M., et al.: Challenges in developing a consensus definition of neonatal sepsis. Pediatric Research 88(1), 14–26 (2020).
13. Napolitano, L.: Sepsis 2018: Definitions and Guideline Changes. Surg. Infect. (Larchmt). 19(2), 117–125 (2018).
14. Matics, T., Sanchez-Pinto, L.: Adaptation and validation of a pediatric sequential organ failure assessment score and evaluation of the Sepsis-3 definitions in critically ill children. JAMA Pediatr. 171(10), 1–9 (2017).
15. Zhang, X., Zhivaki, D., Lo-Man, R.: Unique aspects of the perinatal immune system. Nat. Rev. Immunol. 17(8), 495–507 (2017).
16. Wynn, L., Polin, R.: A neonatal sequential organ failure assessment score predicts mortality to late-onset sepsis in preterm very low birth weight infants. Pediatr. Res. 88(1), 85–90 (2020).
17. Simonsen, K., Anderson-Berry, A., Delair, S., Davies, H.: Early-onset neonatal sepsis. Clin. Microbiol. Rev. 27(1), 21–47 (2014).
18. Fleischmann, C., et al.: Global incidence and mortality of neonatal sepsis: A systematic review and meta-analysis. Arch. Dis. Child. 106(8), 45–752 (2021).
19. Howie, S.: Blood sample volumes in child health research: review of safe limits. Bull. World Health Organ. 89(1), 46–53 (2011).
20. Jiang, J., et al.: Neonatal sepsis in the neonatal intensive care unit: Characteristics of early versus late onset. J. Microbiol. Immunol. Infect. 37(5), 301–306 (2004).
21. Chiesa, C., Panero, A., Osborn, J., Simonetti, A., Pacifico, L.: Diagnosis of Neonatal Sepsis: A Clinical and Laboratory Challenge. Clin. Chem. 50(2), 279–287 (2004).
22. Komorowski, M., Celi, L., Badawi, O., Gordon, A., Faisal, A.: The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. Nat. Med. 24 (11), 1716–1720 (2018)
23. Fairchild, K., et al.: Vital signs and their cross-correlation in sepsis and NEC: A study of 1,065 very-low-birth-weight infants in two NICUs. Pediatr. Res. 81(2), 315–321 (2017).
24. López-Martínez, F., Núñez-Valdez, E., Lorduy L., García-Díaz, V.: A neural network approach to predict early neonatal sepsis. Comput. Electr. Eng. 76, 79–388 (2019).
25. Johnson, A., et al.: MIMIC-III, a freely accessible critical care database. Sci. Data 3, 1–9 (2016).
26. Hill, M., Rosenwaike, I.: The Social Security Administration's Death Master File: The Completeness of Death Reporting at Older Ages. Soc. Secur. Bull. 64(1), 45–49 (2002).
27. Frey, B.: Health Insurance Portability and Accountability Act. SAGE Encycl. Educ. Res. Meas. Eval. 9–18 (2018).
28. Reis, J., et al.: Neonatal sepsis and inflammatory mediators. Mediators Inflamm. 2014 (2014)

29. Balamuth, F., et al.: Identifying Pediatric Severe Sepsis and Septic Shock: Accuracy of Diagnosis Codes. 167(6), 1295–1300 (2016).
30. Ng, P.: Diagnostic markers of infection in neonates. Arch. Dis. Child. Fetal Neonatal Ed. 89(3), 229–235 (2004).
31. Stoll, B., et al.: Early onset neonatal sepsis: The burden of group B streptococcal and E. coli disease continues. Pediatrics 127(5), 817–826 (2011).
32. Gehrke, J.: Classification and Regression Trees. Encycl. Data Warehouse. Min., pp. 246–280 (2011).
33. Pavlov, Y.: Random forests. Random For. 1–122 (2019).
34. Vapnik, V.: An overview of statistical learning theory. IEEE Trans. Neural Networks, 10(5), 988–999 (1999).
35. Ching, T., et al.: Opportunities and obstacles for deep learning in biology and medicine. J R Soc Interface. 5(141) (2018).
36. Han, H., Guo, X., Yu, H.: Variable selection using Mean Decrease Accuracy and Mean Decrease Gini based on Random Forest. Proc. IEEE Int. Conf. Softw. Eng. Serv. Sci. ICSESS 0, 219–224 (2016).
37. Weerts, H., Mueller, A., Vanschoren, J.: Importance of Tuning Hyperparameters of Machine Learning Algorithms. arXiv (2020).
38. Masino, A., et al.: Machine learning models for early sepsis recognition in the neonatal intensive care unit using readily available electronic health record data. PloS One 14(2), e0212665 (2019)
39. Helguera, C., et al.: Neonatal sepsis diagnosis decision-making based on artificial neural networks. Frontiers in pediatrics 8,525 (2020)
40. Iroh P., Bendel, C.: Diagnostics for neonatal sepsis: Current approaches and future directions. Pediatr. Res. 82(4), 574–583 (2017).