

NeoVault: empowering neonatal research through a neonate data hub

*Original*

NeoVault: empowering neonatal research through a neonate data hub / Pigueiras-del-Real, J.; Ruiz-Zafra, A.; Benavente-Fernandez, I.; Lubian-Lopez, S. P.; Shah, S. A. H.; Shah, S. T. H.; Gontard, L. C.. - In: BMC PEDIATRICS. - ISSN 1471-2431. - ELETTRONICO. - 24:1(2024). [10.1186/s12887-024-05276-y]

*Availability:*

This version is available at: 11583/2995240 since: 2024-12-12T12:32:01Z

*Publisher:*

BioMed Central Ltd

*Published*

DOI:10.1186/s12887-024-05276-y

*Terms of use:*

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*

(Article begins on next page)

RESEARCH

Open Access



# NeoVault: empowering neonatal research through a neonate data hub

Janet Pigueiras-del-Real<sup>1\*</sup>, Angel Ruiz-Zafra<sup>2</sup>, Isabel Benavente-Fernández<sup>3,4,5</sup>, Simón P. Lubián-López<sup>3,4</sup>, Syed Adil Hussain Shah<sup>6</sup>, Syed Taimoor Hussain Shah<sup>7</sup> and Lionel C. Gontard<sup>1,8</sup>

## Abstract

**Background** Stability during early postnatal life in preterm infants is related to better outcomes. Although vital signs are monitored continuously in Neonatal Intensive Care Unites, this monitoring does not include all physiological parameters nor data such as movement patterns. Although there are scattered sources of data, there is no centralized data hub for neonates information.

**Results** We have created the first neonate data hub for easy and interactive access to upload or download postural, physiological, and medical data of neonates: NEOVAULT. NEOVAULT is a platform that provides access to information through two interfaces: 1) via a Web interface (designed for medical personnel, data scientists, researchers); and 2) via a RESTful API (Application Programming Interfaces) -designed for developers-, aiming to integrate access to information into third-party applications. The web access allows searching and filtering according to specific parameters, visualization of data through graphs and images, and generation of datasets in CSV format. Access through the RESTful API is described in OpenAPI, enabling access to information from any device, facilitating it in an interoperable format. Currently, it contains nearly 800,000 postural records and 3.000 physiological data entries. The physiological and postural data stored for each neonate in NEOVAULT are collected through the NRP (*Neonates Recording Platform*) tool, which allows for the automatic and reliable collection of data.

**Conclusion** NEOVAULT is an open platform for simple access to postural, physiological, and medical data of neonates that can be utilized by researchers, data scientists, medical personnel, and programmers. It enables integration into third-party applications and the generation of customized datasets.

**Keywords** Preterm, Data hub, Pose tracking, API, Physiological, Services, Neovault

\*Correspondence:

Janet Pigueiras-del-Real  
janet.pigueiras@uca.es

<sup>1</sup> Department of Condensed Matter Physics, University of Cádiz,  
Cádiz 11510, Spain

<sup>2</sup> Department of Software Engineering, University of Granada,  
Granada 18071, Spain

<sup>3</sup> Division of Neonatology, Department of Paediatrics, Puerta del Mar  
University Hospital, Cádiz 11019, Spain

<sup>4</sup> Research Unit, Biomedical Research and Innovation Institute of Cádiz  
(INIBICA), Cádiz 11019, Spain

<sup>5</sup> Area of Paediatrics, Department of Child and Mother Health  
and Radiology - Medical School, University of Cádiz, Cádiz 11003, Spain

<sup>6</sup> Department of Research and Development (R&D), GPI SpA,  
Trento 38123, Italy

<sup>7</sup> PolitoBIOMed Lab, Department of Mechanical and Aerospace  
Engineering, Politecnico di Torino, Turin 10129, Italy

<sup>8</sup> IMEYMAT, University of Cadiz, Cádiz 11510, Spain



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Background

Preterm infants are babies born before the 37<sup>th</sup> week of gestation. Due to their immature development they can face an increased risk of experiencing respiratory problems (e.g bronchopulmonary dysplasia), cardiac issues, infections, and neurodevelopmental disorders [1–4].

Due to their physical conditions, they are admitted to the Neonatal Intensive Care Unit (NICU), where the clinical evaluation of the baby's health and constant monitoring of vital signs are carried out. This health assessment is most often conducted through visual observation of the infants' behavioral traits (movements, facial expressions, crying), neuroimaging exploration, and cardio-respiratory and ECG monitoring by neonatologists [5, 6].

The medical information collected from the clinical analysis of preterm infants is often documented in notes or reports crafted by specialists. Additionally, in certain instances, is stored in private platforms or storage systems (such as in the case of medical images) accessible only to designated healthcare professionals, ensuring privacy of the patient's information.

In recent times, there has been a rising interest in exploring the collection and subsequent analysis of video and audio data as non-invasive methods for information gathering. This data, upon thorough examination, has the potential to improve the clinical monitoring of a neonate's health and the assessments conducted in a NICU [5, 7–11].

The integration of this medical information into databases (gathered by collection tools or generated in hospitals) significantly streamlines the process for medical staff when reviewing records. This technological advancement not only enhances efficiency but also serves as a valuable tool in gaining insights into the effectiveness of various treatments. By centralizing and organizing the data, the database provides medical staff with a comprehensive and accessible platform for thorough record analysis. This, in turn, contributes to a more informed understanding of the efficacy of different treatment modalities, fostering continuous improvement in patient care and outcomes [12–14].

Public medical databases that house information about preterm infants play an essential role in scientific research, healthcare enhancement, early detection of health issues, and the innovation of new therapies and medications. Beyond that, these databases serve as invaluable assets for the education of healthcare professionals, students, and scientists. They actively promote collaboration among researchers and healthcare professionals, propelling technological advancement by creating a conducive environment for the testing and development of cutting-edge data analysis tools and information technologies in the realm of healthcare [15].

To the best of the authors' knowledge, there is a lack of publicly accessible databases specifically designed for the

study of neonates. This scarcity is primarily attributed to privacy and security considerations that place constraints on obtaining ethical approval [5]. Nevertheless, it is noteworthy to mention some data-sets and database that could be found:

1. *BabyPose*: it is a data-set, encompassing data on 12 limb-joint locations and depth images related to the movement of neonates [15, 16].
2. *MIAdataset*: it consists in the states vector, along with the corresponding timestamp, derived from depth measurements of a preterm infant. It contains a timeline of 16 different states in which the infant under examination was in. To obtain this dataset, you have to complete, sign and return a form that you can find in the web page where is located the information of the dataset. After that, you will receive the credentials to download it. Note that the data-set is available only for research purposes [17].
3. *Preterm Clinical Network (PCN)*: a web-based systematic method for collecting data concerning the care of women at risk of preterm birth. Notably, it incorporates a registry of children born to women at risk, who have undergone specialized preterm surveillance and may have received preterm interventions, whether born prematurely or not. However, it is essential to note that it is not specifically a database providing neonatal-specific information, and direct access to the data is not feasible; interested parties can obtain it by making a request to the corresponding author [12].
4. *MMSdataset*: the Multi-Modal Stimulations data-set contains preterm infants data including: gestational age, chronological age, corrected gestational age, sex, birth weight, birth length, birth occipitofrontal circumference (OFC), APGAR at 1-min, and APGAR at 5-min, pre-post intervention (5 days) changes of weight, length, INFANIB (Infant Neurological International Battery) and NIPS (Neonatal Infant Pain Scale) [18, 19].

The aforementioned databases and datasets exhibit certain constraints that merit consideration: 1) the issue of direct access arises, as some databases demand the initiation of a formal access request process. This can introduce delays and procedural hurdles in obtaining the required data; 2) the absence of a user-friendly website interface is notable. A streamlined and intuitive interface can significantly enhance the user experience, particularly in terms of exporting data, reformatting information, or seamlessly integrating new data into the existing system. This aspect becomes important in ensuring efficient utilization and accessibility for diverse users; 3) a noteworthy limitation pertains to the scope of information within

these databases, particularly concerning body pose reference points. The available data-sets offer insights into a limited set of reference points, potentially constraining comprehensive analyses or applications requiring a more extensive array of pose-related data.

Moreover, a crucial observation is the absence of databases containing detailed information on the physiological parameters of neonates. This gap in the available resources underscores a potential limitation in comprehensive research and analysis focused on understanding and addressing the unique healthcare needs of preterm infants. As such, tackling these constraints could significantly contribute to the advancement of a research and medical care in the neonatal domain.

To overcome these limitations, we have developed NEOVAULT the first neonate data hub. NEOVAULT provides public access to a comprehensive collection of perinatal and neonatal data. Besides it goes beyond traditional databases by encompassing vital physiological parameters and data of 33 reference points (landmarks), which are crucial for understanding preterm infant body pose. The different data sources currently considered in this first version of NEOVAULT are shown in Table 1.

At the forefront of artificial intelligence, NEOVAULT functions as a platform for training cutting-edge AI models, facilitating the automating early detection of health problems (such as neurological and motor [35–38]) and holding the potential to revolutionize the efficiency and accuracy of neonatal care. Noteworthy is NEOVAULT's user-friendly interface, enabling seamless navigation through complex queries, data mining, and visualization, positioning it as an indispensable tool in the realm of preterm infants' healthcare and research.

Furthermore, NEOVAULT includes a high-level API (*Application Programming Interface*) implemented through a RESTful style [39] to provide a set of web services than can be used by developers/programmers to access data automatically or integrate data access in third-party applications.

In summary, NEOVAULT is a proposal for promoting open science by the collection of large and rich standardized datasets in the field of neonatal care. The primary focus of NEOVAULT is to provide a web for compiling neonatal information of clinical interest, accessible for reading and writing by both medical personnel and researchers. A data hub like NEOVAULT can be useful not only for easily visualizing and searching for information but also for providing a compilation of different measurands of clinical interest that often are found scattered in different data-bases, that are rarely stored in clinical settings, or that until recently have not been recorded as they depend on the subjective observations done by medical staff. Numerous studies have shown that combining data from various sources can be

beneficial for evaluating the physical condition of neonates [40], predicting risks [41], or associating movement patterns with motor development or the early identification of neurological or motor anomalies [32]. NEOVAULT not only provides all this information, but its access is also interactive and user-friendly, allowing users to obtain organized and structured data ready for analysis, without the need to worry about data cleaning or curation.

## Construction and content

NEOVAULT is a platform for interactive, real-time access, from any device, to physiological, medical, and postural information of neonates. NEOVAULT is hosted on <https://conversational.ugr.es/neovault/>. NEOVAULT follows a client-server architecture implemented through a Service-Oriented Architecture (SOA) philosophy [42].

In the server-side (also known as *backend*) of NEOVAULT, a set of web services are implemented to provide functionality: registering new neonates, searching by filters, listing all neonates, obtaining physiological parameters, generating datasets, etc. These services are implemented in Python using the Flask framework<sup>1</sup>. Additionally, other libraries are used for data processing and other functionalities such as Pandas<sup>2</sup>, Numpy<sup>3</sup>, os, zipfile, and shutil. MySQL is used as the database management system, and its interaction with the services is done through the SQLAlchemy library<sup>4</sup>.

One of the functionalities implemented in the backend, in addition to information management, is the ability to generate an animated image (GIF) from a set of postural data. This is done for each package of postural data uploaded for the neonate, to visualize it illustratively later on. This functionality is implemented using the MediaPipe<sup>5</sup>, Matplotlib, and ImageMagick libraries.

On the other hand, NEOVAULT's front-end utilizes native HTML5, CSS, and Javascript technologies. Additional libraries are used for design (CSS), such as Bootstrap<sup>6</sup> and Fontawesome, and for functionality (Javascript), such as JQuery<sup>7</sup> and Highcharts<sup>8</sup>.

All the information in NEOVAULT is accessible through two different interfaces:

<sup>1</sup> <https://flask.palletsprojects.com/en/3.0.x/>

<sup>2</sup> <https://pandas.pydata.org/>

<sup>3</sup> <https://numpy.org/>

<sup>4</sup> <https://www.sqlalchemy.org/>

<sup>5</sup> [https://developers.google.com/mediapipe/solutions/vision/pose\\_landmarker](https://developers.google.com/mediapipe/solutions/vision/pose_landmarker)

<sup>6</sup> <https://getbootstrap.com/>

<sup>7</sup> <https://jquery.com/>

<sup>8</sup> <https://www.highcharts.com/>

**Table 1** Data sources collected and available in NeoVAULT

Parameter	Description	Expected values	Justification for inclusion in NeoVAULT (medical relevance)
<b>Sex</b>	Biological sex at birth	Boy, Girl, Both	It is essential for improving the accuracy in risk assessment, personalizing treatments, and enhancing short and long-term outcomes, considering the biological differences that may influence the health and development of newborns [20, 21]
<b>Gestational Age</b>	Age of a baby during pregnancy	[25,36] Weeks	It is essential for assessing risk, planning interventions, predicting long-term development, and improving neonatal care protocols [22, 23]
<b>Birth Size</b>	Baby size, measuring from head to toe	[20,50] Centimeters	It is fundamental for assessing the health of the newborn and predicting short- and long-term risks. This variable allows for the identification of at-risk neonates, personalization of medical care, improved monitoring of development, and ultimately, optimization of both immediate and future health outcomes [24]
<b>Birth Head Circumference</b>	Head circumference of the baby	[15,45] Centimeters	This variable not only provides critical information about fetal growth but also helps predict short- and long-term risks, allowing for timely and appropriate interventions to improve the health outcomes of the newborn [25, 26]
<b>Birth Weight</b>	Newborn weight	[0-∞] Grams	It is a key indicator for assessing the health of the newborn and predicting short- and long-term risks [22]
<b>APGAR1 and APGAR5</b>	Quick assessment performed on a newborn at 1 minute and 5 minutes	[0,10]	The APGAR score is useful for the assessment and initial management of the newborn's health. Its use not only provides critical information about the newborn's condition at birth but also allows for monitoring their development and health over time, thereby contributing to improved outcomes in neonatal care [27–29]
<b>CRIB</b>	Clinical Risk Index for Babies	[0,23]	The CRIB score helps identify newborns at risk and provides valuable information to optimize care and improve long-term health outcomes [30, 31]
<b>Brain Damage</b>	Brain damage or neurological damage	Yes/No	Recording neonatal brain damage in a database is essential for tracking development, guiding personalized care, and improving long-term outcomes. It helps link brain injury with clinical factors, enabling better treatment and research into neurological development [32]
<b>Vital Signs</b>	SP02 (Oxygen Saturation) and Heart Rate (Times of heart beats per minute - BPM-)	[0,100] for SP02 and [220] for Heart Rate	The monitoring of vital signs is essential for the continuous assessment and proper care of newborns. This data not only facilitates the early detection of health issues but is also crucial for personalizing treatment, improving clinical outcomes, and contributing to the development of best practices in neonatal care [33, 34]
<b>Pose Estimation</b>	A total of 33 body points, each with x, y, and z values (spatial position)	[0,1] for x and y [−1,1] for z (normalized values)	The study of neonatal movement patterns (pose tracking), along with other clinical data (weight, gestational age, physiological parameters), can be used to assess motor development and facilitate the early identification of neurological or motor anomalies [32, 35]

1. *Web Interface.* Through the options on the website, users can list available data, filter the information according to medical parameters as well as by date, and generate custom datasets. These datasets are

stored in CSV files, so they can be easily exploited by doctors and data scientists through data processing software such as Excel, Jupyter, etc.



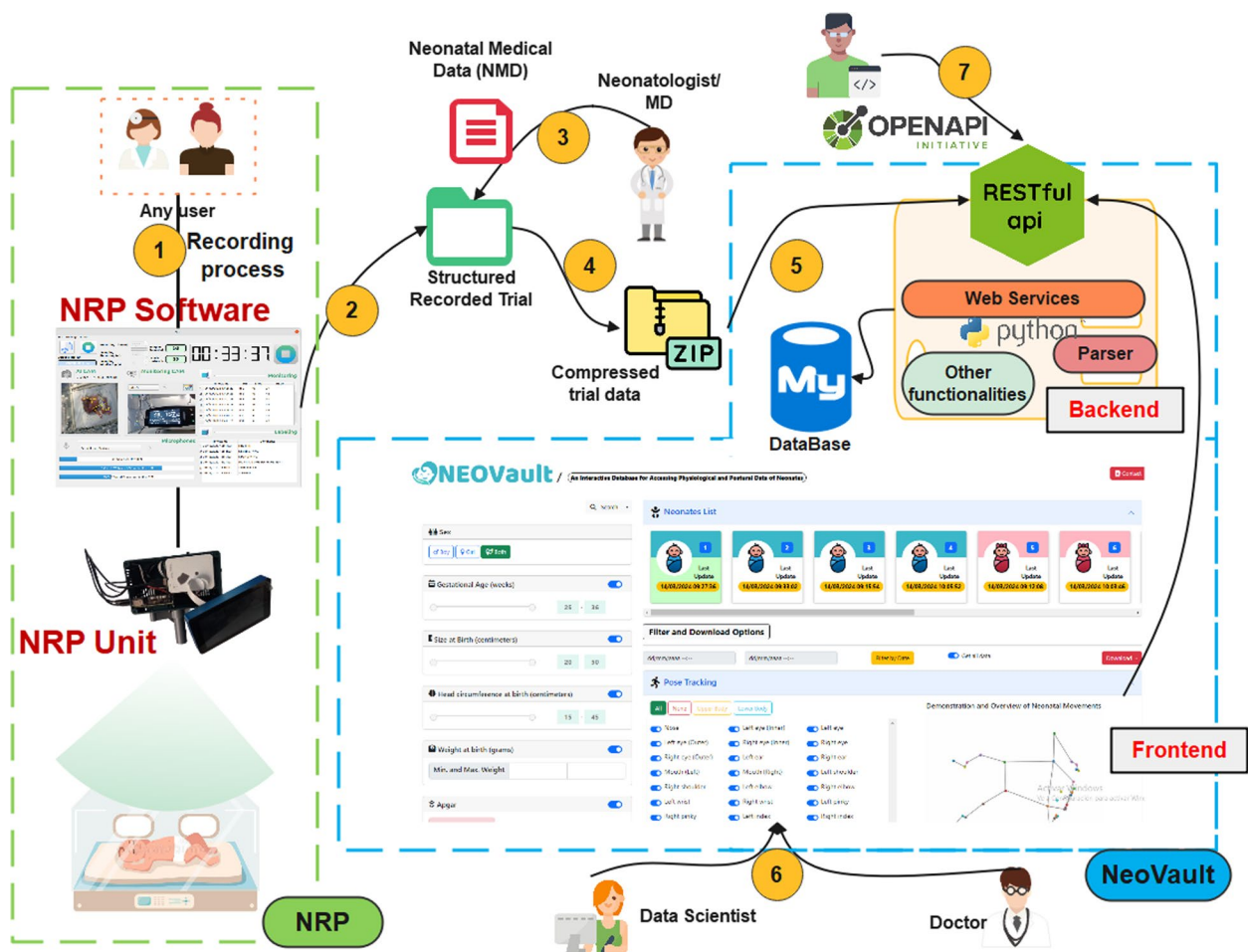
2. **API (Application Programming Interface).** A REST-full API is provided for programmers to access information automatically, aiming to integrate medical data in real-time into other types of systems. The API specification is described in OpenAPI, accessible from the same website through a dedicated web-page. All information returned by the services is represented in JSON format to ensure operability and enable access to the information from any device and platform.

NEOVault is a platform that has been created from scratch, aiming to expand its dataset over time. Currently, there are data registered for 11 neonates (8 boys and 3 girls) including the following medical indicators: sex, gestational age, birth size, head circumference at birth, birth weight, Apgar 1 and Apgar 5 tests, CRIB (*Clinical Risk Index for Babies*), and whether the neonate has brain injury or not assessed by clinicians through brain ultrasonography.

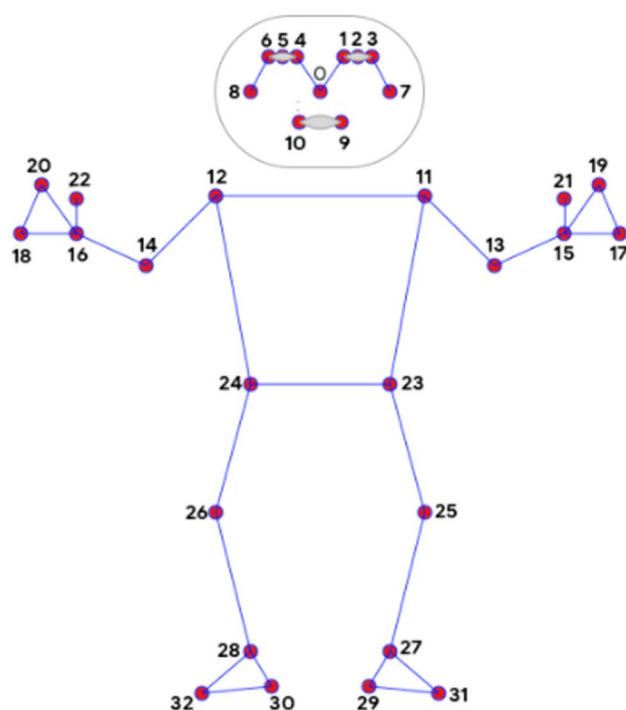
For each neonate, there are records at different time points regarding their physiological parameters (heart rate and oxygen saturation) and postural data. In total, as of March 2024, NEOVault has approximately 3000 physiological records (each containing heart rate and oxygen saturation) and close to 800.000 postural records, where each record stores information for 33 body points (nose, left elbow, various parts of the eyes, etc.) in three spatial coordinates ( $x, y, z$ ).

The Fig. 1 shows the architecture and data flow of NEOVault.

The data have been collected using the *Neonate Recording Platform (NRP)* [8], which was deployed at Puerta del Mar University Hospital (Cádiz, Spain) throughout the year 2023 (1). NRP allows for the scheduling of automatic clinical trials for data collection, creating a folder structure for each trial where study metadata (study identifier, neonate ID, start time, end time, etc.) are stored, along with physiological data captured in real-time through an artificial vision system, positional data captured in



**Fig. 1** NeoVault architecture and workflow



- |                    |                      |
|--------------------|----------------------|
| 0. nose            | 17. left_pinky       |
| 1. left_eye_inner  | 18. right_pinky      |
| 2. left_eye        | 19. left_index       |
| 3. left_eye_outer  | 20. right_index      |
| 4. right_eye_inner | 21. left_thumb       |
| 5. right_eye       | 22. right_thumb      |
| 6. right_eye_outer | 23. left_hip         |
| 7. left_ear        | 24. right_hip        |
| 8. right_ear       | 25. left_knee        |
| 9. mouth_left      | 26. right_knee       |
| 10. mouth_right    | 27. left_ankle       |
| 11. left_shoulder  | 28. right_ankle      |
| 12. right_shoulder | 29. left_heel        |
| 13. left_elbow     | 30. right_heel       |
| 14. right_elbow    | 31. left_foot_index  |
| 15. left_wrist     | 32. right_foot_index |
| 16. right_wrist    |                      |

**Fig. 2** Pose landmarker model - Image obtained from ([https://ai.google.dev/edge/mediapipe/solutions/vision/pose\\_landmarker](https://ai.google.dev/edge/mediapipe/solutions/vision/pose_landmarker)) (Property of Google)

real-time through an AI camera system, labeling data, and audio capture-related data (2).

NRP has two main cameras. The first camera is a regular camera (similar to a webcam) that focuses on the medical monitor connected to the neonate, where all vital signs (heart rate, SpO<sub>2</sub>, respiratory rate) are displayed. Through NRP, frames of the medical monitor's front are captured, and using an algorithm called *CardMed*, which is based on computer vision (e.g., deep learning classification, image cropping, Optical Character Recognition, OCR ...), these physiological parameters are extracted from each frame along with a timestamp. This algorithm is extensively described in [8], where it was shown to have a reliability of 91% for heart rate and 90% for SpO<sub>2</sub>.

On the other hand, the second camera is a three-lens depth and AI camera (Luxonis OAK-D model<sup>9</sup>), which loads the pre-trained Google Mediapipe® model<sup>10</sup> and allows the extraction of 33 body landmarks from the neonate in real time (>20 frames per second, fps). For each frame, all 33 landmarks are recorded in three dimensional spatial coordinates, that is, consisting of x, y, and z components (see Fig. 2). Additionally, when nurses or doctors access the incubator to handle the neonate, NRP only detects and stores the body of the neonate. On the

other hand, if an arm, hand, or object obscures the body, the software does not store partial body data—only complete data—to ensure its quality. For this reason, all data recorded in NEOVAULT correspond to full-body captures of the neonate.

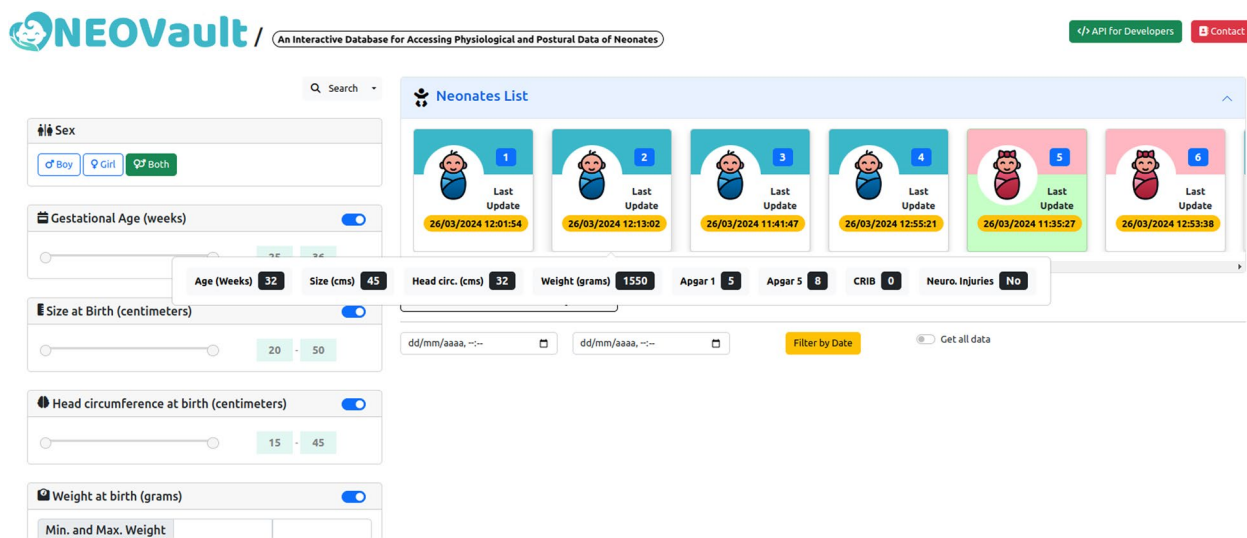
To this folder structure, a CSV file containing medical data known only to medical staff (for privacy reasons) is added (so far this process is done manually) (3). This entire information is compressed into a single file (4) that is used by NEOVAULT as input data. Currently, this information is uploaded through the RESTful API with security constraints (5), which is then transformed into database records via a parser implemented in Python.

Once the information is registered, it is easily and conveniently accessible through the web interface for both medical personnel or data scientists (6), as well as for programmers or developers using the API (7), aiming to integrate this information in real-time and automatically into other systems or solutions.

The Web interface is structured around four blocks: (a) filtering and searching for neonates; (b) list of available neonates along with medical information; defining a time frame for (c) postural information, where you can filter to indicate which parts of the body you want to include in the dataset; and finally, (d) physiological parameters (heart rate and oxygen saturation).

<sup>9</sup> <https://shop.luxonis.com/products/oak-d>

<sup>10</sup> [https://ai.google.dev/edge/mediapipe/solutions/vision/pose\\_landmarker](https://ai.google.dev/edge/mediapipe/solutions/vision/pose_landmarker)



**Fig. 3** Search by medical parameters and neonate list in web interface

The search block (a) allows obtaining all available neonates (without filters) or applying a filter according to the following parameters: 1) gender (boy, girl, both); 2) gestational age (value between 25 and 36 weeks); 3) birth size (value between 20 and 50 centimeters); 4) birth head circumference (value between 15 and 45 centimeters); 5) minimum and maximum birth weight (in grams); 6) APGAR 1 and APGAR 5 scores (values between 0 and 10); 7) CRIB (Clinical Risk Index for Babies) score (value between 0 and 23); and finally, whether the neonate has 8) brain damage or not. Additionally, in this block, each parameter can be enabled or disabled to consider it in the search or not.

Once the search criteria are applied (b), the available neonates are listed, displaying for each of them the gender (using blue color for male and pink for female, along with an illustrative icon), the neonate identifier (integer number), and the last data update date. Additionally, when hovering over each neonate, the values associated with the medical parameters indicated in the search block (a) are displayed in a pop-up. Figure 3 shows an illustrative example of these two blocks.

From a selected neonate, it is possible to search for postural parameters data within a date range (c). NEOVAULT provides data for 33 landmarks (see footnote 5), and the platform allows selecting the upper body, lower body, entire body, or a customized selection. As an illustrative example, the user can view a 2D representation showing a few seconds of movement of the selected neonate. The same search request also returns existing physiological parameters (heart rate, oxygen saturation) (d). In NEOVAULT, each data record has, and will continue to have, an associated timestamp (in

milliseconds), regardless of the type of data. This allows for the chronological recording of all data and its subsequent exploitation, as well as the ability to observe evolution over time. For this reason. Additionally, instead of specifying a date range, the web interface allows returning all existing data for the selected neonate. Figure 4 illustrates these two blocks.

In order to enable automatic, real-time data access and integration into third-party systems, NEOVAULT provides an API for programmers/developers. This API follows a RESTful philosophy, and its usage description can be found on the page accessible from the web interface or from its specification through OpenAPI. All information for interacting with the services is described using JSON to ensure access by any platform or software. The services offered by the API include 1) listing neonates according to medical parameters; 2) obtaining postural data; and 3) obtaining physiological data for a neonate within a time range. The specification of these RESTful API endpoints can be found clicking in the green button located in the top-right corner with the text “API for Developers”.

Although the most common way to access information is through one of these two channels (Web interface or RESTful API), the database can be found in NEOVAULT data repository<sup>11</sup>, which will be periodically updated to provide the same information as both access interfaces.

## Utility and discussion

In contrast to other databases dedicated to preterm infants, as highlighted in “Background” section), NEOVAULT stands out by offering a multitude of advantages. These differentiating factors contribute to its uniqueness

<sup>11</sup> <https://github.com/bihut/neovault-database>



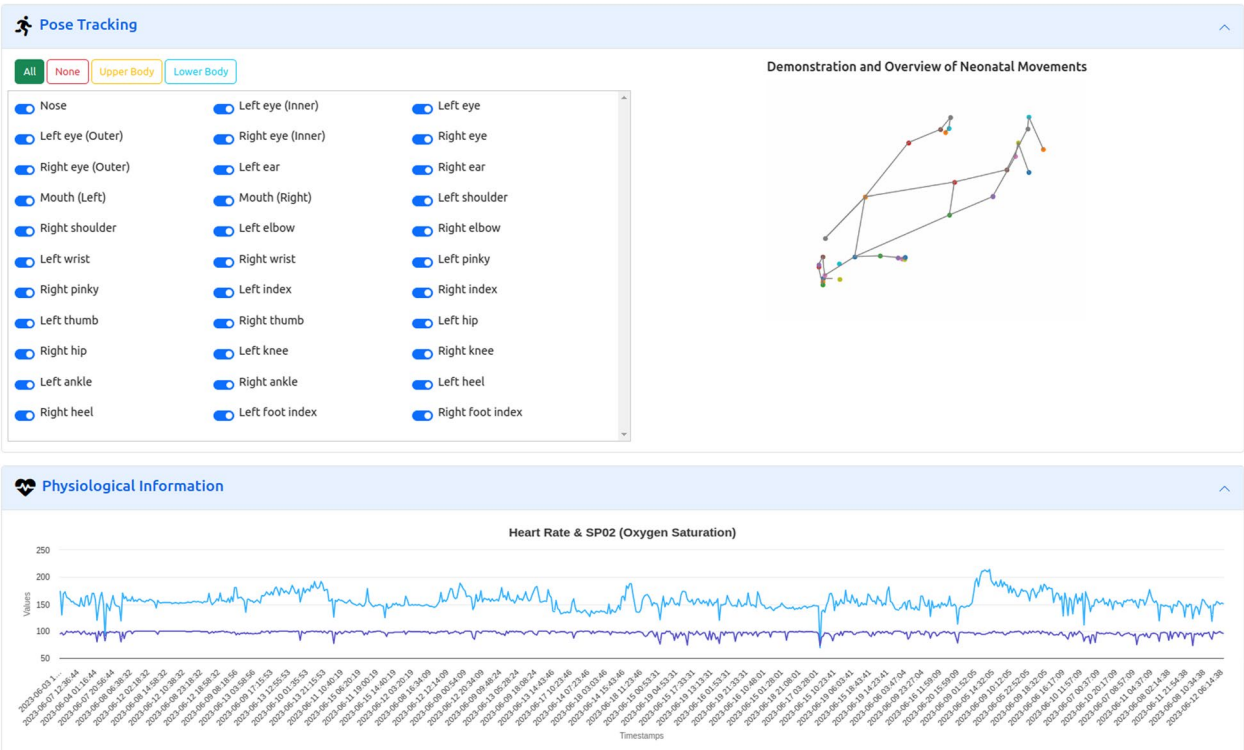


Fig. 4 Postural and physiological data

and enhanced utility in the domain of neonatal care and research. A summary of the comparison between NEOVAULT and other databases is included in Table 2.

One of the main contributions of NEOVAULT is its operation as a publicly accessible database, that houses perinatal and neonatal data, encompassing crucial physiological parameters such as oxygen saturation and heart rate. Additionally, it meticulously provides spatial data of 33 landmarks defining the body pose of preterm infants,

offering an understanding of their intricate neuromotor development. The platform goes beyond mere data storage by providing an useful 2D visualization tool of neonates body movements.

Moreover, NEOVAULT emerges as an invaluable resource for healthcare professionals, researchers, and AI scientists. In the realm of neonatal care, NEOVAULT could play a pivotal role in early diagnosis thanks to the data that can be found in the database, empowering

Table 2 Comparison of preterm infants data-base

Features	BabyPose	MIAdataset	Preterm Clinical Network (PCN)	MMSdataset	NeoVault
Intuitive Navigation	Yes	Yes	No	Yes	Yes
Responsive Design	Yes	Yes	Yes	Yes	Yes
3D visualization tool	No	No	No	No	Yes
Data Type	12 limb-joint locations and depth images related to the movement of neonates	A timeline of 16 different states vector of a preterm infant	Data concerning the care of women at risk of preterm birth and registry of children born to women at risk(born prematurely or not)	Perinatal data-neonatal data	Perinatal data-neonatal data; physiological parameters and 33 reference points of the preterm body
User-friendly interface	No	No	Yes	No	Yes
Accessibility	Public	Restricted	Restricted	Public	Public
Support for Complex Queries	No	No	Yes	No	Yes
Collaborative environment	No	No	Yes	No	Yes
Cost	Free	Free	Free	Free	Free
Ease Import/Export	Yes	No	No	Yes	Yes

healthcare practitioners to identify developmental deficits in the preterm infant's stages. This early detection is crucial for the initiation of therapeutic interventions, fostering optimal outcomes for the neonates.

Although, in the AI era, NEOVAULT stands at the forefront, offering a robust foundation for training AI models. The data housed within the database becomes a training ground for cutting-edge AI, enabling the automation of early detection processes for neonatal medical issues. This intersection of healthcare and AI holds the promise of revolutionizing the efficiency and accuracy of neonatal care.

Beyond its substantive data holdings, NEOVAULT distinguishes itself with a user-friendly interface. This interface is not merely a gateway for data retrieval, also provides seamless data visualization, integration, and formatting, enhancing the overall user experience and making NEOVAULT a versatile and indispensable database in the landscape of preterm infants healthcare and research. On the other hand, the possibility of accessing NEOVAULT data through its RESTful API not only enables another way for developers or programmers to access the data, but also the opportunity to access that data in real-time and integrate such access into other applications or systems, such as hospital systems, private software, medical repositories for data scientists, etc.

The current version of NeoVault is stable, but it is true that there are several shortcomings which are being considered for future work and will be implemented in the near future. These planned improvements include:

1. *User management system.* Currently, information is uploaded manually without any record of which users (neonatologists, doctors) contribute that information. The aim is to add a user management system where each medical personnel can manage their neonates through a username/password access method. However, the user/password access method, while necessary, is not sufficient. Additional personnel verification mechanisms will be implemented to ensure that any user wishing to register to upload information is indeed healthcare or research personnel. For example, institutional email will be verified, and professional credentials or those from their institution will be requested, among other methods, to validate their identity and role.
2. *Improve the system for adding data.* Currently, adding new information is done through the RESTful API using a username and password to prevent unauthorized users from uploading false or unreliable data in the current version. With the user management mentioned in (1), each identified user will be able to upload data more easily and with more credibility.

This means opening up the capability for the rest of the community to contribute new data. NEOVAULT accepts data structured according to the format used by *NRP*. In future versions, there are plans to publish the specification so that any researcher or medical personnel can adapt their data to the format accepted by NeoVault and upload it easily.

3. *Automatic data dump from NRP.* Although the option to upload information manually (2) will be available, the *NRP* software (used as data collector) [8] will be modified so that the data collected is automatically saved in NEOVAULT, eliminating the need for manual processing.
4. *Adding new data sources.* In addition to physiological and postural parameters, *NRP* records audio and labels made by medical personnel that can be included as data sources in NeoVault. Additionally, external data sources can be added to NEOVAULT, such as blood analysis results or medical test outcomes (e.g., heel prick test) and clinical data that may be useful from a medical standpoint, such as post-menstrual age, whether the birth was by cesarean section, premature apnea, systolic murmur or even maternal characteristics like genetic factors, age, pre-existing medical conditions like diabetes, hypertension, heart disease, etc.. These data will be added based on medical criteria and the usefulness they may have as fundamental variables for data exploitation. The different improvements will be gradually included in NEOVAULT.
5. *Automatic tagging of events.* An improvement of the recording of the data will be that of the implementation of algorithms for the automatic annotation of the videos. For example, an algorithm could be applied to the video source during recording to classify events happening in the field-of-view and to add tags such as if the baby is present or not, if the baby is being manipulated or not by the medical staff, or if the position of the baby is not in supine.
6. *Ensuring data quality.* Although all information stored in and uploaded to NEOVAULT is consistent, with no empty values, null values, etc., there remains the possibility—either due to error or ill intent from users in the future—that the data, despite following a correct structure, may not be accurate or reliable. This means that the data could be fabricated, generated (for example, through generative AI), erroneous, or duplicated. Therefore, in the future, work will focus on algorithms or techniques that allow for the analysis of movement patterns before uploading the data to determine if they are coherent and real, thus minimizing the risk of erroneous data contaminating the rest of the reliable dataset.

## Conclusions

NEOVault stands as a data hub for contributing to advancement in neonatal healthcare research. It is a platform created to become a key resource, offering researchers, medical professionals, students, and data scientists access to meticulously organized datasets of preterm infants. These datasets would not only facilitate investigations into diagnostic improvements and treatment strategies but also empower studies focused on early disease detection, providing valuable insights for enhancing neonatal care protocols.

At its core, NEOVault emerges as a publicly accessible database, offering a comprehensive repository of movement data, physiological parameters, and neonatal and perinatal records for preterm infants who underwent hospitalization in a NICU. Its user-friendly interface allows users to define and filter queries, visualize datasets, and seamlessly export results in various file formats. The platform goes beyond conventional data storage by providing an advanced 3D visualization tool, enabling users to explore and analyze neonatal movements dynamically.

Future initiatives involve leveraging the database for cutting-edge studies, particularly in the realm of early detection of neurological issues, such as those managed by the PARENT project<sup>12</sup>. Additionally, NEOVault envisions establishing connections with more extensive datasets encompassing larger populations of preterm infants, promising to broaden its scope and impact in the field of neonatal healthcare research.

## Abbreviations

AI	Artificial Intelligence
APGAR	Appearance - Pulse - Grimace - Activity - Respiration
API	Application Programming Interface
CRIB	Clinical Risk Index For Babies
CSS	Cascading Style Sheets
CSV	Comma-separated values
ECG	Electrocardiogram
GIF	Graphics Interchange Format
HTML	Hypertext Markup Language
JSON	Javascript Object Notation
My	co-founder Michael Widenius's daughter name
NICU	Neonatal Intensive Care Unit
NIPS	Neonatal Infant Pain Scale
NRP	Neonate Recording Platform
OFC	Occipitofrontal Circumference
PARENT	PrematuRe Newborn Motor and Cognitive Impairments
PCN	Preterm Clinical Network
REST	Representational State Transfer
SOA	Service-Oriented Architecture
SQL	Structured Query Language

## Acknowledgements

The authors would like to thank all the healthcare staff at Puerta del Mar University Hospital (Cádiz, Spain) for their collaboration during data collection, as well as the parents who granted permission for the collection of data about their baby.

## Authors' contributions

JPR conceived the idea of NeoVault, implemented the web interface (front-end), and conducted data collection at the hospital. ARZ designed the architecture of NeoVault and implemented the back-end (services and RESTful API). IBF and SPL planned the studies with neonates at the hospital and processed the documentation with the ethics committee. SAHS and SATS assisted in data cleaning, curation, and management. LCG is the project organizer and, along with JPR, prepared the paper which was then read and approved by all authors. The author(s) read and approved the final manuscript.

## Funding

Funding for open access publishing: Universidad de Cádiz/CBUA This work is part of the "PrematuRe nEwborn motor and cogNitive impairment: Early diagnosis PARENT project". The PARENT project has received funding from the European Union's Horizon 2020 research and innovation programme under the Maria Skłodowska-Curie Innovative Training Network 202, Grant Agreement N° 956394. In addition, this work also received inputs from the Spanish R&D projects funded by MCIN/AEI/10.13039/501100011033 GOMINOLA (Grant PID2020-118112RB-C21 and Grant PID2020-118112RB-C22).

## Data availability

The data used is available through the web interface of the NEOVault platform (<https://conversational.ugr.es/neovault>) as well as through the RESTful API <https://conversational.ugr.es/neovault/api/v1>. The raw dataset is available in the repository <https://github.com/bihut/neovault-database>.

## Declarations

### Ethics approval and consent to participate

The study was approved by the Puerta del Mar University Hospital (Cadiz, Spain) (reference number 125.22) and after obtaining the informed consent of the parents. All research data are de-identified and securely stored.

### Consent for publication

Not applicable.

### Competing interests

The authors declare no competing interests.

Received: 17 April 2024 Accepted: 22 November 2024

Published online: 30 November 2024

## References

1. Barfield WD. Public health implications of very preterm birth. *Clin Perinatol*. 2018;45(3):565–77.
2. Alvarez-Garcia A, Fornieles-Deu A, Costas-Moragas C, Botet-Mussons F. Maturational changes associated with neonatal stress in preterm infants hospitalised in the NICU. *J Reprod Infant Psychol*. 2014;32(4):412–22.
3. Soleimani F, Azari N, Ghiasvand H, Shahrokhi A, Rahmani N, Fatollahierad S. Do NICU developmental care improve cognitive and motor outcomes for preterm infants? A systematic review and meta-analysis. *BMC Pediatr*. 2020;20:1–16.
4. Kolb B, Harker A, Gibb R. Principles of plasticity in the developing brain. *Dev Med Child Neurol*. 2017;59(12):1218–23.
5. Leo M, Bernava GM, Carcagni P, Distant C. Video-Based Automatic Baby Motion Analysis for Early Neurological Disorder Diagnosis: State of the Art and Future Directions. *Sensors*. 2022;22(3):866.
6. Zhao T, Griffith T, Zhang Y, Li H, Hussain N, Lester B, et al. Early-life factors associated with neurobehavioral outcomes in preterm infants during NICU hospitalization. *Pediatr Res*. 2022;92(6):1695–704.
7. Olmi B, Frassinetti L, Lanata A, Manfredi C. Automatic Detection of Epileptic Seizures in Neonatal Intensive Care Units Through EEG, ECG and Video Recordings: A Survey. *IEEE Access*. 2021;9:138174–91.
8. Pigueiras-del Real J, Gontard LC, Benavente-Fernández I, Lubián-López SP, Gallero-Rebollo E, Ruiz-Zafra A. NRP: A multi-source, heterogeneous, automatic data collection system for infants in neonatal intensive care units. *IEEE J Biomed Health Inform*. 2023;28(2):678–89.

<sup>12</sup> <https://parenth2020.com/>

9. Ruiz-Zafra A, Precioso D, Salvador B, Lubián-López SP, Jiménez J, Benavente-Fernández I, et al. NeoCam: An edge-cloud platform for non-invasive real-time monitoring in neonatal intensive care units. *IEEE J Biomed Health Inform.* 2023;27(6):2614–24.
10. Pigueiras-del-Real J, Gontard LC, Lubián-López SP, Benavente-Fernández I, Ruiz-Zafra A. Towards an AI driven early detection of brain injuries in neonates through non-contact audio and video recording. In *DETERMINED*. 2022. p. 122–32.
11. Cabon S, Porée F, Simon A, Rosenc O, Pladys P, Carraut G. Video and audio processing in paediatrics: A review. *Physiol Meas.* 2019;40(2):02TR02.
12. Carter J, Tribe RM, Sandall J, Shennan AH. The Preterm Clinical Network (PCN) Database: a web-based systematic method of collecting data on the care of women at risk of preterm birth. *BMC Pregnancy Childbirth.* 2018;18(1):1–9.
13. Ismail L, Materwala H, Karduck AP, Adem A. Requirements of health data management systems for biomedical care and research: scoping review. *J Med Internet Res.* 2020;22(7):e17508.
14. Samadbeik M, Fatehi F, Braunstein M, Barry B, Sareman M, Kalhor F, et al. Education and Training on Electronic Medical Records (EMRs) for health care professionals and students: A Scoping Review. *Int J Med Inform.* 2020;142:104238.
15. Migliorelli L, Moccia S, Pietrini R, Carnielli VP, Frontoni E. The babyPose dataset Data Brief. 2020;33:106329. <https://doi.org/10.1016/j.dib.2020.106329>.
16. Migliorelli L. BabyPose Dataset. <https://zenodo.org/record/3891404>. Accessed Nov 2024.
17. VRAI - Vision R, Marche AID. MIA (MOTION INFANT ANALYSIS) DATASET. <https://vrai.dii.univpm.it/mia-dataset>. Accessed Nov 2024.
18. Aranha VP. Multi modal stimulations to modify the neuromotor responses of hospitalized preterm infants [PhD dissertation]. Haryana: Maharishi Markandeshwar; 2019.
19. Aranha Vencita Priyanka BAK. Chahal Aksh: Multi modal stimulations. <https://data.mendeley.com/datasets/k4j689jfwj/1>. Accessed Nov 2024.
20. Bhat V, Bhandari V. Sex specificity in neonatal diseases. In: *Principles of Gender-Specific Medicine*. Elsevier; 2023. pp. 841–867.
21. Garfinkle J, Yoon EW, Alvaro R, Nwaesei C, Claveau M, Lee SK, et al. Trends in sex-specific differences in outcomes in extreme preterms: progress or natural barriers? *Arch Dis Child-Fetal Neonatal Ed.* 2020;105(2):158–63.
22. Crilly CJ, Haneuse S, Litt JS. Predicting the outcomes of preterm neonates beyond the neonatal intensive care unit: what are we missing? *Pediatr Res.* 2021;89(3):426–45.
23. Belachew A, Tewabe T. Neonatal sepsis and its association with birth weight and gestational age among admitted neonates in Ethiopia: systematic review and meta-analysis. *BMC Pediatr.* 2020;20:1–7.
24. Cheung Y, Yip P, Karlberg J. Size at birth and neonatal and postneonatal mortality. *Acta Paediatr.* 2002;91(4):447–52.
25. J P NA, Mitsuda N, Eitoku M, Yamasaki K, Maeda N, Fujieda M, et al. Influence of chest/head circumference ratio at birth on obstetric and neonatal outcomes: the Japan environment and children's study. *Am J Hum Biol.* 2023;35(6):e23875.
26. Lee KA, Hayes BC. Head size and growth in the very preterm infant: a literature review. *Res Rep Neonatol.* 2015;5:1–7.
27. Andegiorghis AK, Andemariam M, Temesghen S, Ogbai L, Ogbe Z, Zeng L. Neonatal mortality and associated factors in the specialized neonatal care unit Asmara. Eritrea *BMC Public Health.* 2020;20:1–9.
28. Desalew A, Sintayehu Y, Teferi N, Amare F, Geda B, Worku T, et al. Cause and predictors of neonatal mortality among neonates admitted to neonatal intensive care units of public hospitals in eastern Ethiopia: a facility-based prospective follow-up study. *BMC Pediatr.* 2020;20:1–11.
29. Thavarajah H, Flatley C, Kumar S. The relationship between the five minute Apgar score, mode of birth and neonatal outcomes. *J Matern-Fetal Neonatal Med.* 2018;31(10):1335–41.
30. Veir CA, Afiune SMRP, Portal DC, Miguel PDP, Saidah TK. Evaluation of neonatal mortality risk in the crib score application. Editorial Board. 2021;211:15–8.
31. Motlagh AJ, Asgary R, Kabir K. Evaluation of Clinical Risk Index for Babies to Predict Mortality and Morbidity in Neonates Admitted to Neonatal Intensive Care Unit. *Electron J Gen Med.* 2020;17(5):em232. <https://www.ejgm.co.uk/article/evaluation-of-clinical-risk-index-for-babies-to-predict-mortality-and-morbidity-in-neonates-admitted-7902>.
32. Wu Q, Xu G, Wei F, Chen L, Zhang S. Rgb-d videos-based early prediction of infant cerebral palsy via general movements complexity. *IEEE Access.* 2021;9:42314–24.
33. Mithal LB, Yogev R, Palac HL, Kaminsky D, Gur I, Mestan KK. Vital signs analysis algorithm detects inflammatory response in premature infants with late onset sepsis and necrotizing enterocolitis. *Early Hum Dev.* 2018;117:83–9.
34. Sullivan BA, Fairchild KD. Vital signs as physiologic markers of neonatal sepsis. *Pediatr Res.* 2022;91(2):273–82.
35. Shin HI, Shin HI, Bang MS, Kim DK, Shin SH, Kim EK, et al. Deep learning-based quantitative analyses of spontaneous movements and their association with early neurological development in preterm infants. *Sci Rep.* 2022;12(1):3138.
36. Zlatanovic D, Čolović H, Živković V, Stanković A, Kostić M, Vučić J, et al. The importance of assessing general motor activity in premature infants for predicting neurological outcomes. *Folia Neuropathol.* 2022;60(1):427–35.
37. Shin HI, Park MW, Lee WH. Spontaneous movements as a prognostic tool of neurodevelopmental outcomes in preterm infants: A narrative review. *Clin Exp Pediatr.* 2023;66(1):458–64.
38. Park MW, Shin HI, Bang MS, et al. Reduction in limb-movement complexity at term-equivalent age is associated with motor developmental delay in very-preterm or very-low-birth-weight infants. *Sci Rep.* 2024;14:8432.
39. Richardson L, Ruby S. RESTful web services. O'Reilly Media, Inc.; 2008.
40. Kermani F, Sheikhtaheri A, Zarkesh MR, Tahmasebian S. Risk factors for neonatal mortality in Neonatal Intensive Care Units (NICUs): a systematic literature review and comparison with scoring systems. *J Pediatr Neonatal Individualized Med.* 2020;9(2):e090226–e090226.
41. Mangold C, Zoretic S, Thallapureddy K, Moreira A, Chorath K, Moreira A. Machine learning models for predicting neonatal mortality: a systematic review. *Neonatology.* 2021;118(4):394–405.
42. Laskey KB, Laskey K. Service oriented architecture. *Wiley Interdiscip Rev Comput Stat.* 2009;1(1):101–5.

# Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.