

## Original Research



# Interpretable deep neural networks for advancing early neonatal birth weight prediction using multimodal maternal factors

Muhammad Mursil<sup>a,\*</sup>, Hatem A. Rashwan<sup>a</sup>, Adnan Khalid<sup>a</sup>, Pere Cavallé-Busquets<sup>b</sup>, Luis Santos-Calderon<sup>c</sup>, Michelle M. Murphy<sup>c</sup>, Domenec Puig<sup>a</sup>

<sup>a</sup> Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili, Tarragona, 43007, Tarragona, Spain

<sup>b</sup> Unit of Obstetrics & Gynaecology, University Hospital Sant Joan, Reus, IISPV, CIBEROBn ISCII, 43201, Tarragona, Spain

<sup>c</sup> Faculty of Medicine and Health Sciences, IISPV, Universitat Rovira i Virgili, Reus, CIBEROBn ISCIII, 43201, Tarragona, Spain

## ARTICLE INFO

## Keywords:

Explainable AI  
Machine learning  
Deep learning  
Maternal health  
Birth weight prediction  
Fetal development  
Maternal factors

## ABSTRACT

**Background:** Neonatal low birth weight (LBW) is a significant predictor of increased morbidity and mortality among newborns. Predominantly, traditional prediction methods depend heavily on ultrasonography, which does not consider risk factors affecting birth weight (BW).

**Objective:** This study introduces a robust deep neural network for a clinical decision-support system designed to early predict neonatal BW, using data available during early pregnancy, with enhanced precision. This innovative system incorporates a comprehensive array of maternal factors, placing particular emphasis on nutritional elements alongside physiological and lifestyle variables.

**Methods:** We employed and validated various traditional machine learning models as well as an interpretable deep learning model using the TabNet architecture, noted for its proficient handling of tabular data and high level of interpretability. The efficacy of these models was evaluated against extensive datasets that encompass a broad spectrum of maternal health indicators.

**Results:** The TabNet model exhibited outstanding predictive capabilities, achieving an accuracy of 96% and an area under the curve (AUC) of 0.96. Significantly, maternal vitamin B12 and folate status emerged as pivotal predictors of BW, emphasizing the crucial role of nutritional factors in influencing neonatal health outcomes.

**Conclusions:** Our results demonstrate the substantial benefits of integrating multimodal maternal factors into predictive models for neonatal BW, markedly enhancing the precision over traditional AI methods. The developed decision-support system not only has a possible application in prenatal care but also provides actionable insights that can be leveraged to mitigate the risks associated with LBW, thereby improving clinical decision-making processes and outcomes.

## 1. Introduction

Birth weight (BW) is a crucial indicator of a newborn's health and a predictor of future developmental outcomes [1]. Globally, neonatal low birth weight (LBW) is associated with significant morbidity and mortality and affects approximately 20% of babies born each year [2]. Infants born with LBW are at a higher risk of early mortality and long-term health complications, underscoring the need for accurate prediction and effective management strategies.

Traditionally, the prediction of neonatal BW has relied heavily on ultrasonography. However, while widely used, ultrasonography can be inaccurate due to its dependence on the operator's skill, the mother's physiological condition, limited access to ultrasound technology in low-resource areas and poor image quality [3–5]. Ultrasonographic data

alone have limitations for LBW assessment in AI models, potentially leading to suboptimal prenatal care and assessment of LBW risk.

Research has shown that maternal physiological, lifestyle and particularly nutritional factors play significant roles in determining neonatal BW. Current methods often do not consider the full range of maternal factors that can influence BW. These factors include maternal age, body mass index (BMI), socioeconomic status, smoking, physical activity, status in folate, vitamin B12, betaine, choline, anemia, previous pregnancy, adverse pregnancy history, and gestational age (GA) at birth [6–8]. Yet, these factors are often underutilized in predictive models. Furthermore, most existing models need a deep learning (DL) approach that could fuse these multimodal data sources to offer high predictive accuracy and interpretability [9]. TabNet is a DL

\* Corresponding author.

E-mail address: [muhammad.mursil@urv.cat](mailto:muhammad.mursil@urv.cat) (M. Mursil).

architecture designed for tabular data, which facilitates data fusion by leveraging sequential attention and feature masking mechanisms. It uses sequential attention to identify and focus on the most relevant features for decision-making. This approach is constructive in extracting the most relevant features from multimodal data, ensuring effective integration and interpretation of diverse data types. Additionally, TabNet uses a feature masking mechanism, which selectively allows the inclusion of certain features in the network at each decision step. This dynamic masking enhances both the model’s interpretability and performance by preventing irrelevant or redundant features from influencing predictions [10].

Consequently, this work addresses these gaps by developing and validating a deep neural network-based model, specifically utilizing the TabNet architecture [10], for the early and more accurate prediction. This research paper aims to answer two research questions: How can DL enhance the prediction of neonatal BW based on multimodal data? What is the contribution of maternal nutritional factors and other factors to these predictions?

Our work makes the following significant contributions:

1. We integrate a broad spectrum of maternal health indicators, ranging from traditional physiological and lifestyle factors to less commonly analyzed nutritional factors, within a sophisticated DL framework, TabNet. This integration allows for the development of a more comprehensive, precise, and interpretable tool for predicting neonatal BW.
2. Using the TabNet model, we apply feature importance and SHAP analysis, providing clear insights into which factors are most significantly associated with BW predictions. This level of interpretability enhances the utility of the predictive model, enabling healthcare providers to understand the underlying reasons for risk assessments.
3. Our approach allows for better-informed prenatal care and intervention strategies by providing actionable insights derived from a deep understanding of the integrated data. This is pivotal for improving prenatal care outcomes and ultimately enhancing neonatal health.

In conclusion, while existing literature successfully predicts BW using ultrasound and demographic (physiological and lifestyle) data by applying Machine Learning (ML) models, it overlooks maternal nutritional factors and advanced DL techniques like TabNet. Additionally, there is a lack of focus on model interpretability. This study addresses these gaps by incorporating these elements to enhance BW prediction and support clinical decision-making. The statement of significance of this work is as follows:

Statement of significance	
Problem or Issue	LBW significantly raises neonatal morbidity/mortality risks. Current models lack integration of multimodal maternal factors, particularly nutrition, introducing bias without ultrasound data.
What is already known	Ultrasound is the primary maternal assessment method, but no robust decision-support exists without it. LBW models rely heavily on physiological traits and neglect maternal nutrition.
What this paper adds	This study emphasizes multimodal maternal factors, especially nutritional factors, proposing a deep learning system for BW prediction in early pregnancy, without ultrasound. A clinically interpretable model is developed using diverse maternal data.

## 2. Related work

This section reviews the related work on BW prediction, focusing on two primary approaches: ultrasonography-based predictions and those relying on clinical data. By examining previous studies in both domains, we aim to highlight key advancements, identify gaps, and establish a foundation for further investigation.

### 2.1. BW prediction, by AI, using ultrasonography

Most studies have explored the potential of ultrasonography to predict neonatal BW using ultrasound images, video, and ultrasound characteristics (e.g., fetal biometry measurements) by applying ML and DL techniques to develop a classification and regression model.

Using deep belief network (DBN), Feng et al. [11] estimated fetal BW in China. The study utilized a dataset based on ultrasound parameters of 7875 singleton fetuses collected from the West China Second University Hospital between January 2016 and December 2017. The proposed model achieved a mean absolute percentage error (MAPE) of  $6.09 \pm 5.06\%$  and a mean absolute error (MAE) of  $198.55 \pm 158.63$  g, outperforming commonly used regression formulas. Plotka et al. [12] proposed BabyNet, an end-to-end method for direct fetal birth weight (FBW) prediction from spatiotemporal ultrasound video scans. BabyNet extends a 3D ResNet-based network with a Residual Transformer Module (RTM) to capture both spatial and temporal features, improving weight estimation accuracy. Evaluated on 225 fetal ultrasound videos from 75 patients, BabyNet outperformed existing methods, achieving an MAE of 254 g and an MAPE of 7.5%, comparable to human experts. The best results were obtained by combining BabyNet’s predictions with human estimates, reducing the MAE to 180 g and MAPE to 5.2%, outperforming both individual approaches. Szymon et al. [13] used 900 fetal ultrasound examinations and their GA to predict fetal weight using DL, bypassing traditional biometry measurements like head and femur circumference. The proposed model closely matched human expert estimates, with a MAPE of 3.75%, showing promise as a more efficient and reliable alternative to manual biometry. Li et al. [14] developed new BW prediction models for various GA stages in a Chinese population using 2D ultrasound measurements of over 19,000 fetuses. Different models were applied for different gestational stages: a Volume-based model (VM) for early stages (28–36 weeks), Multiple Linear Regression (MLR) for mid-gestation (37–39 weeks), and Fractional Polynomial Regression (FPR) for later stages (40–42 weeks). Their results showed significant improvements compared to their previously published models in accuracy, with an 85.1% prediction rate, and reduced systematic and random errors, with an error of 10% of the actual BW.

### 2.2. BW prediction using demographic data

Recent studies utilized the potential of maternal physiological and lifestyle factors to predict neonatal BW without ultrasonography by applying ML models. Khan et al. [15] utilized a dataset comprising 821 instances from three hospitals in the UAE, with 88 features related to maternal physiological and lifestyle data. The study aimed to estimate the BW and classify LBW using 30 ML algorithms. The Random Forest (RF) algorithm provided the best results for BW estimation with a MAE of 294.53 g. For LBW classification, logistic regression achieved the highest accuracy of 90.24%, precision of 87.6%, recall of 90.2%, and an F1 score of 0.89. GA, hypertension, and diabetes are highlighted as the most contributing features to BW classification. Alabbad et al. [16] applied ML models to predict BW using two datasets: IEEE with 1801 records and King Fahd University Hospital (KFUH) with 639 records. Both datasets were based on physiological and lifestyle factors. The Extra Trees (ET) model achieved 98% accuracy on the KFUH dataset, while RF reached 96% accuracy on the IEEE dataset, using 12 selected features to enhance prediction performance. Ranjbar et al. [17] predict

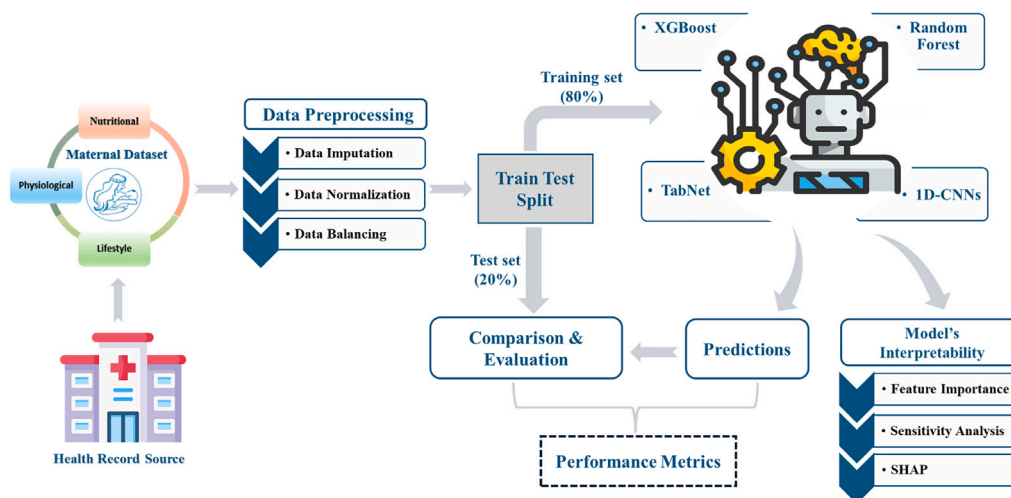


Fig. 1. Pipeline and workflow of the proposed approach.

neonatal LBW using the IMaNet dataset from 2022–2024 in Iran, focusing on physiological and lifestyle factors. The XGBoost model demonstrated strong performance, achieving 79% accuracy and an AUCROC of 0.79, with model validation performed using k-fold cross-validation. GA and a previous history of LBW were identified as the most significant predictors, according to the feature importance analysis. Lu et al. [3] used a dataset of 4212 clinical records from a hospital in Shenzhen, China, focusing on fetal weight prediction without using ultrasound. The authors employed an ensemble learning approach combining RF, XGBoost, and LightGBM algorithms. Results showed that the ensemble model achieved higher accuracy, reducing the mean relative error to 7% compared to traditional methods. Additionally, the Intersection-over-Union value exceeded 0.6, demonstrating the model's effectiveness in predicting fetal weight at various GAs.

### 3. Materials and methods

The proposed study utilized a range of ML and DL models to accurately classify neonatal BW and pinpoint the critical maternal factors that influence it. In addition, the study interprets the model's results to demonstrate how each factor impacts BW. Fig. 1 outlines the study's approach, which is built around four main components: (i) cohort construction, (ii) data preprocessing, (iii) application of ML and DL models, and (iv) model interpretation to support clinical decision-making.

#### 3.1. Data description and analysis

We employed two datasets: the Reus–Tarragona birth cohort, a private dataset, and the IEEE Child Birth Weight dataset, which is publicly available in [16].

##### 3.1.1. Reus-Tarragona cohort

The participants in this study were acquired from longitudinal research by the Unit of Preventive Medicine and Biostatistics, Faculty of Medicine and Health Sciences, University Rovira i Virgili, and the University Hospitals Sant Joan, Reus, and Joan XXIII, Tarragona. The obstetric team recorded the data based on maternal physiological, nutritional, and lifestyle factors at <12, 15, 24–27 and 34 gestational weeks (GWs). This study relies on maternal data available at 15 GWs to propose a clinical decision-support system for predicting neonatal BW in early pregnancy. Data from 15 GWs were used due to the advantage of being an early pregnancy time point with a fixed GA, minimizing variability and physiological effects on blood biomarkers. It also aligns with the typical timing of first prenatal check-ups, supporting early prevention strategies. Based on the eligibility criteria, 742 mother-child

dyads participated in this study. The study was open to women who provided before 12 weeks of pregnancy fasting blood sample and had a confirmed singleton pregnancy with a viable fetus. However, women with a history of chronic illness, diabetes, or surgeries that affected their levels of folate or cobalamin status or eating habits were not eligible to participate. Neonatal BW is measured shortly after birth and categorized into two groups, LBW (<2500 g) and normal birth weight (NBW) ( $\geq 2500$  g) [2], instead into three groups because our study focuses primarily on accurately predicting LBW due to its significant negative impact on child health and development (see Figure S2 in supplementary material). For this research, both NBW and high birth weight (over 4000 g) are treated as a single class. The upper part of Table 1 reports the statistical description of the maternal factors and labels obtained for this study from the Reus–Tarragona cohort.

##### 3.1.2. IEEE child birth weight dataset

The IEEE child birth weight dataset is a comprehensive collection of clinical data encompassing 1215 records and 13 distinct features, derived from a cohort of pregnant women in Assam, India [16]. This dataset is a vital component of our study, offering a rich source of variably interacting maternal factors that are critical for assessing neonatal outcomes.

The features within this dataset were systematically collected under the supervision of medical professionals, ensuring the integrity and clinical relevance of the data [16]. Each variable was selected based on its established association with neonatal health outcomes, reflecting rigorous adherence to global medical standards and practices. The process was meticulously overseen to prevent any random selection of features, thus maintaining the dataset's scientific validity [16].

The dataset includes a range of demographic, physiological, nutritional, lifestyle and medical features such as maternal age, height, blood group, and detailed blood pressure measurements taken at different stages of pregnancy. Other key variables include fetal sex, initial and final hemoglobin levels, and blood sugar status. These attributes were chosen for their known impacts on birth outcomes, particularly BW.

Table 1 in the subsequent section provides a detailed breakdown of these attributes, offering insights into the dataset's comprehensive nature and its alignment with clinical research requirements. The selection of these features is grounded in well-established medical guidelines, ensuring that each attribute included in the dataset is recognized as essential for robust health assessments. The data was compiled to include variables that provide a holistic view of the maternal health environment, thereby enhancing the predictive model's ability to assess factors contributing to LBW accurately. This careful compilation not

**Table 1**  
Statistical description of maternal factors and label for Reus–Tarragona and IEEE datasets (Mean ± Standard Deviation or n (%)).

Reus-Tarragona dataset		
Feature	Statistical description	Imputed values
<b>Physiological factors</b>		
Maternal age	32 ± 4.64	30
Previous pregnancy	Yes: 397 (53.5%), No: 345 (46.4%)	19
Adverse pregnancy history	Yes: 306 (41.2%), No: 436 (58.7%)	18
Maternal BMI	24.28 ± 4.66 kg/m <sup>2</sup>	47
<b>Nutritional factors</b>		
Maternal plasma folate	31.33 ± 28.90 nmol/L	242
Maternal vitamin B12	340.03 ± 151.9 pmol/L	242
Maternal betaine	15.6 ± 3.83 μmol/L	242
Maternal choline	8.07 ± 1.73 μmol/L	242
Maternal anemia (hemoglobin <11 g/dL)	Yes: 12 (1.6%), No: 730 (98.4%)	45
<b>Lifestyle factors</b>		
Maternal smoking exposure	Yes: 273 (36.7%), No: 469 (63.3%)	41
Socioeconomic status	Low: 89 (12%), Middle: 341 (45.96%), High: 312 (42%)	36
Physical activity	Low (1): 456 (61.4%), 3–4 days: 279 (37.6%), Never: 7 (0.94%)	110
<b>Target variable</b>		
Neonatal BW	LBW: 47 (6.3%), NBW: 695 (93.7%)	0
IEEE dataset		
Feature	Statistical description	Imputed values
<b>Physiological factors</b>		
Maternal age	22 ± 4.28	146
Maternal height	142 ± 17.24 cm	270
Maternal blood group	A(+ve): 228 (221.2%), A(-ve): 30 (2.7%), B(+ve): 349 (35.5%), B(-ve): 18 (1.6%) AB(+ve): 134 (12.5%), AB(-ve): 5 (0.46%), O(+ve): 297 (27.7%), O(-ve): 11 (1.0%)	270
Previous pregnancy	0.6 ± 0.99	0
Fetal sex	Male: 559 (52.1%), Female: 512 (47.6%)	154
Initial systolic blood pressure	105.9 ± 12.33	147
Initial diastolic blood pressure	65.89 ± 7.7	147
Final systolic blood pressure	111.10 ± 13.11	149
Final diastolic blood pressure	70.61 ± 8.57	149
<b>Nutritional Factors</b>		
Initial hemoglobin level	10 ± 1.05 μmol/L	246
Final hemoglobin level	10.45 ± 0.96 μmol/L	246
Blood Sugar status	100.66 ± 11.48 mg/dL	431
<b>Lifestyle Factors</b>		
Socioeconomic status	Above Poverty Line: 305 (28.4%), Below Poverty Line: 767 (71.6%)	146
<b>Target variable</b>		
Neonatal BW	LBW: 142 (13.3%), NBW: 930 (86.7%)	143

only reinforces the dataset’s robustness but also its reliability and applicability in clinical research settings focused on prenatal and neonatal care.

### 3.2. Data preprocessing

Data preprocessing is essential for achieving strong predictive performance and understanding each maternal factor’s influence. We applied techniques such as data imputation for handling missing data, normalization, and balancing to prepare the dataset.

**Data imputation:** The Reus–Tarragona dataset exhibits a time-series structure, encompassing data collected at <12, 15, 24–27, and 34 GWs. We focused on maternal data from 15 GWs to facilitate early prediction of neonatal BW. Missing data points at this stage were addressed through linear imputation [18], utilizing historical maternal data from <12 GWs and future maternal data from 24–27 GWs. We used linear regression to determine the linear trend in the data and visualize data points with the fitted linear regression line, assessing the goodness of fit (see Figure S3). Additionally, model performance across different imputation techniques is compared (see Table S3). The mathematical representation of linear interpolation is as follows:

$$\hat{y} = y_1 + (x - x_1) \frac{(y_2 - y_1)}{(x_2 - x_1)}, \quad (1)$$

where  $x_1$  and  $y_1$  are the first coordinates,  $x_2$  and  $y_2$  are the second coordinates,  $x$  is the point to perform the interpolation and  $\hat{y}$  is the interpolated value.

The missing values in categorical features were filled out using RF-based imputation. RF imputation is a robust, non-parametric technique that leverages the predictive power of RF models to estimate missing values. It is particularly effective for categorical data because it captures non-linear relationships and interactions between variables without assuming specific distributions [19]. The IEEE dataset also contained missing values across several features. These missing data points were addressed using the same imputation strategies applied to the Reus–Tarragona cohort, ensuring methodological consistency. We have reported the number of imputed values for each feature in Table 1. While imputation helps maintain dataset usability, it is important to note that working with imputed values introduces assumptions about data distribution, which can influence the model’s behavior. This limitation is particularly relevant given the real-world incompleteness of electronic health records and should be considered when deploying models trained on such data.

**Data normalization:** Data normalization plays a crucial role in optimizing the performance of ML models by ensuring that features are on a comparable scale. In this study, we employed the widely used Min–Max scaler [20] for normalization. We transformed the data to a specific range, between 0 and 1, allowing each maternal factor to contribute equally to the learning process. Doing so enhances the model’s ability to generalize, ensure fair computation of feature importance, and reduce the risk of overfitting. The Min–Max scaling method

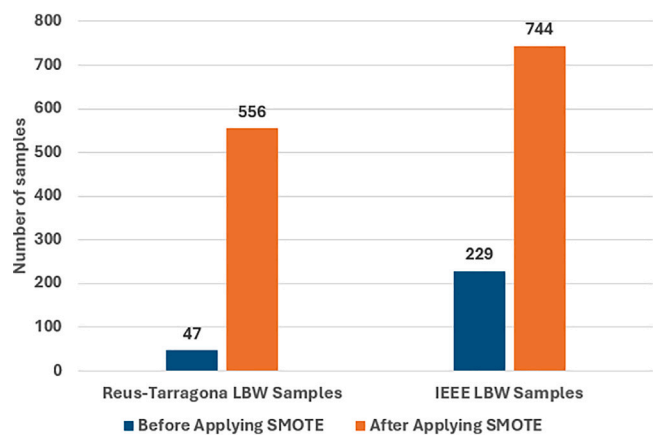


Fig. 2. Number of samples in the LBW and NBW classes before and after applying SMOTE to the training dataset.

is mathematically represented as follows:

$$X_{\text{scaled}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}, \quad (2)$$

where  $X$  represents the original data point,  $X_{\min}$  and  $X_{\max}$  are the minimum and maximum values in the dataset, respectively, and  $X_{\text{scaled}}$  is the scaled data point after normalization.

**SMOTE-based Data Augmentation:** Imbalanced data is a prevalent issue in medical datasets, often leading to ML models making biased predictions by favoring the majority class while neglecting the under-represented class [21]. To address this challenge, we applied different synthetic data generation methods that are common in the state of the art, such as, [22]. Among the various techniques evaluated, we have chosen the Synthetic Minority Over-sampling Technique (SMOTE), a robust data augmentation method [23] that demonstrated the best model performance. The performance evaluation of the SMOTE has been presented in Table S1 in the supplementary material. SMOTE synthetically generates samples in the minority class (LBW) by creating new data points that slightly differ from the original LBW data, thus helping to mitigate the data imbalance problem. Importantly, the synthetic data was generated exclusively for the training set, ensuring that no artificially created data was present in the test set. Fig. 2 illustrates the distribution of samples in the minority class (LBW) and the majority class (NBW) before and after applying SMOTE to the training dataset.

We thoroughly described the data preprocessing steps and emphasized the importance of applying the same preprocessing pipeline, including the same normalization technique, to all new and incoming data before making predictions to ensure the interoperability and robustness of our model across various systems. To validate this, we applied the model to two distinct datasets — the private Reus-Tarragona Birth cohort and the IEEE dataset — ensuring both were processed using identical normalization techniques. The model's sensitivity and adaptability to variations in data characteristics were then assessed, demonstrating its robustness and flexibility in handling diverse datasets.

### 3.3. Applied AI algorithms

In this study, we employed and evaluated a range of traditional ML and DL algorithms to classify neonatal BW into two categories: LBW and NBW. The ML algorithms utilized include RF, and XGBoost, along with DL algorithms, such as Soft-ordering 1-dimensional Convolutional Neural Networks (1D-CNNs), and TabNet. These methods were chosen due to their proven effectiveness and accuracy in handling similar classification tasks [24].

#### 3.3.1. Traditional ML methods

**Random Forest (RF):** RF is an ensemble learning algorithm that constructs multiple decision trees and aggregates their predictions to improve accuracy and generalization. This algorithm is highly suitable for medical applications because it handles complex and high-dimensional data [25]. RF was employed in this study because of its resistance to overfitting, a critical advantage when working with varied clinical datasets. Additionally, it provides valuable insights into feature importance, aiding in identifying key clinical variables, and is robust to noise, ensuring reliable performance even when the data contains anomalies or inconsistencies [26].

**Extreme Gradient Boosting (XGBoost):** XGBoost is a powerful ML algorithm that uses gradient boosting for regression, classification and ranking tasks. It builds decision trees sequentially, each improving on the previous by minimizing errors through gradient descent. XGBoost was chosen for this research due to its ability to reduce overfitting through regularization techniques, combined with its computational efficiency and parallel processing, making it an ideal choice for medical data analysis. Moreover, XGBoost's capacity to manage missing values and capture non-linear relationships ensures robust performance in healthcare applications, where data is often incomplete or noisy [27].

#### 3.3.2. Deep neural network-based methods

**Soft-ordering 1-dimensional Convolutional Neural Networks (1D-CNNs):** 1D-CNNs are designed to effectively model sequential or structured data by leveraging the power of convolutional layers to capture local dependencies and patterns within the input features. The “soft-ordering” aspect introduces flexibility in the order in which features are processed, allowing the network to handle scenarios where the strict order of input data may not always be optimal. This architecture employs multiple layers of convolutions, followed by pooling operations that progressively reduce the dimensionality while preserving essential information. The final layers use global pooling and fully connected neurons to generate predictions. Soft-ordering 1D-CNNs are particularly advantageous when working with complex data, as they balance the need for local feature extraction with the adaptability to variations in input sequences [28]. 1D-CNNs are highly beneficial in clinical decision-making, where patient data may include time series (e.g., vital signs), categorical (e.g., diagnosis codes), and numerical features (e.g., lab results). 1D-CNNs can capture clinically relevant patterns across these different data types. This multimodal approach enhances the network's ability to support accurate and robust predictions, ultimately contributing to more reliable decision-making in healthcare applications.

#### Attentive Interpretable Tabular Learning (TabNet):

TabNet is a deep learning model specifically designed for tabular data, utilizing attention mechanisms to focus selectively on relevant features during training. Unlike traditional models, TabNet combines interpretability with high performance by dynamically learning which features are most important at each decision step. One key aspect of TabNet's architecture is the implementation of feature masking, which allows the model to apply sparse attention to features, enabling it to ignore irrelevant data points and focus on those that contribute most significantly to the task. This helps reduce overfitting and enhances model efficiency. TabNet's sequential data processing further supports more efficient feature selection and generalization across diverse datasets. Moreover, TabNet's flexibility makes it highly suitable for multimodal data, allowing it to seamlessly integrate and process different data types within a single model framework. This versatility positions TabNet as a robust solution for handling complex datasets while maintaining transparency in its decision-making process [29]. Fig. 3 presents a comprehensive overview of the TabNet architecture.

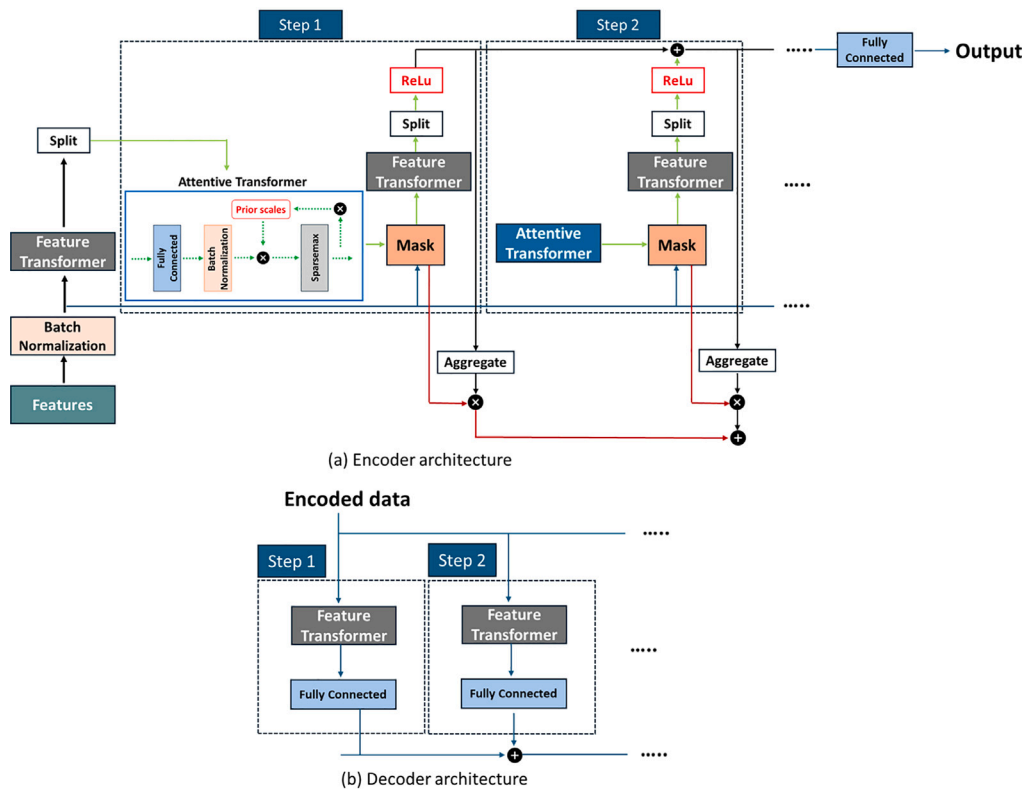


Fig. 3. The TabNet architecture consists of (a) an Encoder, where features undergo batch normalization, transformation through Feature and Attentive Transformers, masking, and aggregation in multiple steps; and (b) a Decoder, where encoded data passes through Feature Transformers and fully connected layers.

### 3.4. Model interpretability

In addition to developing a reliable model for predicting neonatal BW and assessing fetal health, it is equally important to understand the reasoning behind the model's decisions, including which factors influence the outcomes and how they impact predictions. Such insights can assist healthcare professionals in making more informed clinical decisions. We utilized feature importance, sensitivity analysis, and Shapley Additive Explanations (SHAP) analysis to provide this level of interpretability. These methods clearly understand the contribution, significance, and influence of various maternal factors involved in the clinical decision-making process.

### 3.5. Hyperparameter optimization

Hyperparameter optimization involves selecting the optimal set of hyperparameters that minimizes a model's loss function on a specific dataset, improving performance. We used a Tree-Structured Parzen Estimator (TPE) [30], a probabilistic model-based tuning method. TPE works by building two probability distributions: one for the good hyperparameter values and another for the less favorable ones. It evaluates new hyperparameters by comparing these distributions and selecting values likely to yield better results, making the tuning process more efficient than random or grid search. Table 2 presents the optimal values of the model's hyperparameters that yield the best performance.

## 4. Results and discussion

The experiments were conducted following the methodology outlined above. The dataset was divided into an 80:20 ratio, with 80% used for the training and validation sets and 20% reserved for the test set on unseen data. The model predictive analysis was performed on the Reus-Tarragona birth cohort, and the results are reported in Table 3. This dataset was used to evaluate the model's performance on the primary

Table 2

The optimal values and ranges of hyperparameters.

Classifier	Hyperparameter	Range	Optimal Value
Random Forest	Max_depth	[3, 15]	7
	Min_samples_leaf	[1, 10]	1
	Min_samples_split	[2, 20]	2
	N_estimators	[50, 200]	150
XGBoost	Learning_rate	[0.01, 0.3]	0.1
	N_estimators	[50, 200]	150
	Gamma	[0, 1]	0.3
	Booster	[Gbtree, Dart]	Gbtree
1D-CNNs	Hidden_neurons	[256, 2048]	1024
	epoch	[100, 300]	200
	batch_size	[16, 128]	32
	Learning_rate	[1e-5, 0.1]	0.001
	Kernel_size	[3, 7]	3
TabNet	N_steps	[3, 10]	5
	Batch_size	[32, 128]	64
	Virtual_Batch_size	[16, 64]	32
	Max_epochs	[100, 500]	200
	Optimizer_params	[ 1e-1, 1e-3]	lr=2e-2
	Mask_type	[sparsemax, entmax]	entmax
	Gamma_factor	[0.1, 1.0]	0.9

cohort. A separate analysis was conducted to test the generalization capability of the models using the IEEE Child BW dataset, with results summarized in Table 4. To ensure optimal performance, a five-fold cross-validation technique was applied, providing a robust evaluation of the ML model's accuracy and generalization.

### 4.1. Performance measures

This study focuses on predicting neonatal BW. It emphasizes the need for rigorous performance evaluation to develop a reliable support system when ultrasound imaging is unavailable for fetal health

**Table 3**  
Results comparison of models predicting neonatal BW using unimodal, bimodal, and multimodal maternal factors.

Maternal factors	Classifiers	Accuracy (%)	Precision	Recall	F1-Score	AUC
Physiological	RF	89.93	0.9023	0.9023	0.9023	0.9069
	<b>XGBoost</b>	<b>90.37</b>	<b>0.9102</b>	<b>0.9016</b>	<b>0.9059</b>	<b>0.9045</b>
	1D-CNNs	82.52	0.8174	0.8319	0.8235	0.8257
	TabNet	85.94	0.8632	0.8598	0.8611	0.8626
Nutritional	<b>RF</b>	<b>86.09</b>	<b>0.8528</b>	<b>0.8777</b>	<b>0.8694</b>	<b>0.8761</b>
	XGBoost	85.41	0.8333	0.8884	0.8546	0.8612
	1D-CNNs	80.66	0.7812	0.8206	0.8145	0.8085
	TabNet	83.01	0.8428	0.8278	0.8419	0.8465
Lifestyle	RF	65.76	0.6574	0.6605	0.6589	0.6593
	XGBoost	64.72	0.6484	0.6416	0.6459	0.6491
	1D-CNNs	70.56	0.6981	0.7143	0.6971	0.7019
	<b>TabNet</b>	<b>69.42</b>	<b>0.7107</b>	<b>0.6768</b>	<b>0.7054</b>	<b>0.6949</b>
Physiological and Lifestyle	RF	92.09	0.9361	0.9077	0.9212	0.9231
	XGBoost	92.33	0.9320	0.9170	0.9242	0.9202
	1D-CNNs	90.07	0.8921	0.9101	0.9141	0.8932
	<b>TabNet</b>	<b>92.81</b>	<b>0.9351</b>	<b>0.9133</b>	<b>0.9312</b>	<b>0.9284</b>
All Factors	RF	96.16	0.9761	0.9488	0.9623	0.9686
	XGBoost	94.96	0.9563	0.9349	0.9504	0.9508
	1D-CNNs	93.17	0.9176	0.9318	0.9212	0.9322
	<b>TabNet</b>	<b>96.60</b>	<b>0.9801</b>	<b>0.9588</b>	<b>0.9650</b>	<b>0.9689</b>

assessment. We employed four classification models to achieve this and evaluated their predictive performance using five key metrics: accuracy, precision, recall, F1-score, and area under the curve (AUC). These metrics help ensure the reliability and effectiveness of the models in providing accurate predictions.

#### 4.2. Model prediction analysis

In our dataset, neonatal BW is the target variable, categorized into two classes: LBW (class 1) and NBW (class 0). We employed RF, 1D-CNNs, XGBoost, TabNet models for this classification task. BW prediction was analyzed in three distinct phases. The first phase involved an unimodal analysis, wherein each maternal factor — physiological, lifestyle, and nutritional — was examined independently and notably, including nutritional factors that represented a novel approach to predicting neonatal BW. Among the evaluated factors, physiological factors exhibited the highest predictive performance when utilizing the XGBoost model, yielding an accuracy of 90.37%, precision of 0.9103, recall of 0.9016, F1 score of 0.9059, and an AUC of 0.9045. In comparison, nutritional factors demonstrated competitive results, achieving an accuracy of 86.09% with the RF model. Conversely, lifestyle factors produced the lowest predictive performance, with TabNet achieving an accuracy of 69.42% (see at Table 3). These findings highlight the varying predictive capabilities of each maternal factor when modeled independently.

In the second phase, a bimodal analysis combined physiological and lifestyle factors, an approach frequently employed in recent studies. This integration significantly improved the predictive performance across models. TabNet demonstrated superior performance, achieving an accuracy of 92.81%, precision of 0.9351, recall of 0.9133, F1 score of 0.9312, and an AUC of 0.9384 while the 1D-CNNs yielded the lowest performance among the models, with an accuracy of 90.07% (see at Table 3). These results underscore the effectiveness of combining commonly used maternal factors to enhance model accuracy, particularly when using TabNet.

The third phase expanded the feature set by adding nutritional factors to the bimodal analysis, thereby creating a multimodal dataset. With this enriched dataset, TabNet outperformed 1D-CNN and XGBoost. TabNet also performed slightly better than Random Forest when trained on the Reus-Tarragona dataset, indicating its potential advantages in handling multimodal maternal data with higher interpretability, achieving impressive results with an accuracy of 96.60%, precision

of 0.9801, recall of 0.9588, F1-score of 0.9650, and an AUC of 0.9780. In comparison, 1D-CNNs demonstrated the lowest performance, with an accuracy of 93.17% (see Table 3). Including nutritional factors in the prediction analysis, they have significantly improved overall predictive performance, particularly TabNet's ability to leverage multimodal maternal features to predict neonatal BW. Table 3 presents the overall comparative performance of the models across unimodal and multimodal maternal factors, clearly demonstrating the impact of integrating multiple maternal data types on model accuracy and other evaluation metrics.

For proving statistical significance, the Tukey Honestly Significant Difference (HSD) test [31] was conducted to perform the pairwise comparison, indicating the statistical significance of the differences in the multimodal model's performance by comparing several performance metrics, including accuracy, precision, recall, F1-score, and AUC score. Fig. 4 statistical analysis confirm the enhancements observed with TabNet are statistically significant ( $p < 0.05$ ) as it performs significantly differently than the other multimodal models. While RF and TabNet demonstrate comparable performance with no significant differences, TabNet outperforms other models due to its ability to model complex data patterns and built-in interpretability. However, further validation on real-world clinical datasets is required to confirm its reliability and applicability in routine healthcare settings.

The results provide clear evidence that maternal physiological and lifestyle characteristics are crucial in predicting neonatal BW, as supported by prior research. However, our findings go further into insights and underscore the importance of integrating nutritional factors into predictive models for neonatal BW, as their inclusion notably boosts the model's accuracy and robustness. This improvement has significant implications for clinical practice, offering a more reliable and confident decision-support system without ultrasound imaging that could aid gynaecologists and healthcare providers in estimating fetal BW as early as 15 GWs. Incorporating multimodal maternal factors provides stability and reliability in the decision-support system by covering all possible relevant maternal factors. By identifying potential risks, such as LBW, early on, clinicians can take preventive measures to mitigate complications, ultimately improving maternal and neonatal health outcomes.

#### 4.3. Model generalization analysis

The generalization analysis was conducted to assess the robustness and reliability of the proposed method when applied to unseen maternal data from the IEEE dataset. The models were trained from scratch independently on the IEEE dataset, taking into account only the features available within it, using the same training setup as we did with the Reus-Tarragona dataset. Our method demonstrated exceptional and consistent predictive performance. Among the classifiers, TabNet achieved the highest predictive accuracy and performance metrics, as shown in Table 4, outperforming other models. In contrast, 1D-CNNs exhibited the lowest performance. The lower section of Table 4 compares our method's performance with state-of-the-art approaches, highlighting that our model surpassed the existing benchmark, further establishing its effectiveness and reliability in the BW prediction system.

Table 4 compellingly demonstrates the superior performance of our proposed TabNet model in predicting neonatal BW, achieving an accuracy of 98.11%, F1-score of 98.14%, and an AUC of 98.90% on the IEEE Child Birth Weight dataset. These results not only outperform the Random Forest, XGBoost, and 1D-CNNs models included in our study but also surpass the benchmarks set by recently published work, notably [16], who reported a maximum accuracy of 96.00%. Our contributions are significant, as they illustrate the effectiveness of employing an advanced, interpretable deep learning architecture for medical predictive modeling. By integrating TabNet, which utilizes an attention mechanism to focus on relevant features selectively, our

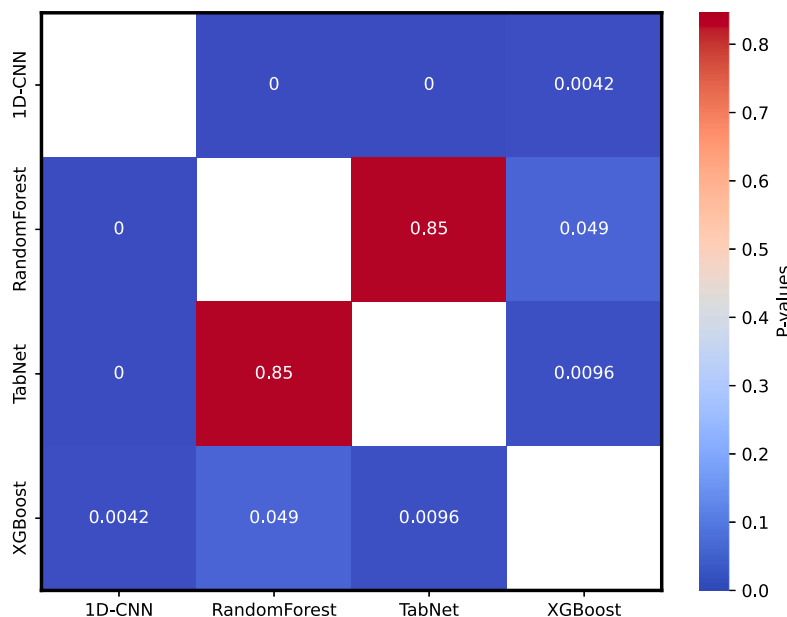


Fig. 4. Heatmap of Tukey HSD test p-values for pairwise model comparisons.

Table 4

The BW prediction results of the proposed methods on IEEE child BW dataset and compared with recent published work.

Classifier	IEEE Dataset				
	Accuracy	F1-score	Precision	Recall	AUC
RF	0.9516	0.9530	0.9510	0.9515	0.9594
XGBoost	0.9462	0.9476	0.9389	0.9547	0.9476
1D-CNNs	0.9044	0.9089	0.8993	0.9137	0.9087
TabNet	0.9811	0.9814	0.9685	0.9846	0.9890
Study	Comparison with state-of-the-art work				
	Accuracy	F1-score	Precision	Recall	AUC
Alabbad et al. [16]	0.9600	0.9600	0.9800	0.9600	-
<b>Proposed method</b>	<b>0.9811</b>	<b>0.9814</b>	<b>0.9685</b>	<b>0.9846</b>	<b>0.9890</b>

approach offers a robust solution that enhances predictive accuracy and provides actionable insights into the impact of various maternal factors. These results affirm the potential of our methodology to be integrated into clinical decision-support systems, offering a more precise tool for healthcare providers in settings where traditional ultrasound might not be available.

#### 4.4. Model interpretation analysis

Using the Reus–Tarragona birth cohort, the best-performing model, TabNet, was analyzed for feature importance, sensitivity, and SHAP analysis to illustrate the maternal factors’ contribution and impact on neonatal BW. TabNet calculates feature importance by leveraging its sequential attention mechanism, which allows the model to focus on specific features at each decision step. This attention mechanism enables TabNet to assign varying weights to features depending on their relevance in different contexts. The importance of a feature is derived based on how frequently it is selected by the attention masks and the magnitude of its contribution to the final prediction.

Fig. 5(a) illustrates the significance and contribution of multimodal maternal factors influencing neonatal BW. Smoking, vitamin B12, and folate are the top three features with the highest contribution to BW prediction. In contrast, factors such as adverse pregnancy history and anemia exhibit minimal influence. The relatively low influence of anemia and adverse pregnancy in the model is likely because the Reus Tarragona Birth Cohort was carried out in the High-risk Obstetric Units

in 2 public University Hospitals, in a developed country. Preventive care to lower the risk of anemia and adverse pregnancy outcome is provided as part of the prenatal clinical guidelines that are universally applied in this region [32]. The importance of maternal iron status for the maintenance of the maternal–placental–fetal unit as well as postnatal maternal health and during lactation is well established and maternal hemoglobin status is systematically monitored and treatment initiated if anemia is detected. Furthermore, the Reus–Tarragona Birth Cohort protocol also included low dose iron supplementation from 12 GW for all mothers to protect maternal iron reserves against depletion in the face of any perinatal hemorrhage risk. Therefore, in the models, the risk of adverse outcome due to these factors is reduced due to the application of prenatal monitoring protocols. As shown in supplementary Figure S1, smoking, plasma vitamin B12, and folate consistently appeared as the most important features across all models, with slight variations in their relative rankings. A key finding from this analysis is the substantial contribution of nutritional factors, which surpasses that of physiological and lifestyle factors, the latter of which have been predominantly emphasized in recent studies. This demonstrates the critical role of nutritional factors in improving the accuracy of BW prediction models.

The sensitivity analysis was conducted by holding all maternal features at their mean values while varying a single input maternal feature to assess its influence on the BW variation generated by the TabNet model. Five representative input values were selected for each feature, including the maximum, median of the mean-to-maximum range, median of the minimum-to-mean range, and minimum. The baseline output was initially determined using the mean values for all input parameters, and the subsequent changes in BW (%) were calculated relative to this baseline output. In Fig. 5(b), the x-axis represents the average percentage change in BW values, calculated using a feature perturbation approach, while the y-axis represents the individual features. Predicted probabilities for LBW and NBW were translated into expected BW values based on representative population-level statistics, enabling the analysis of the relative impact of each feature on BW outcomes. Maternal age shows the greatest influence, with changes in this feature leading to an average variation of nearly 25% in BW. Similarly, Vitamin B12 and folate status each contribute significantly, with average changes of approximately 20% and 15% in BW, respectively. In contrast, the last two features, adverse pregnancy history and pregnancy history, have the least influence, associated with

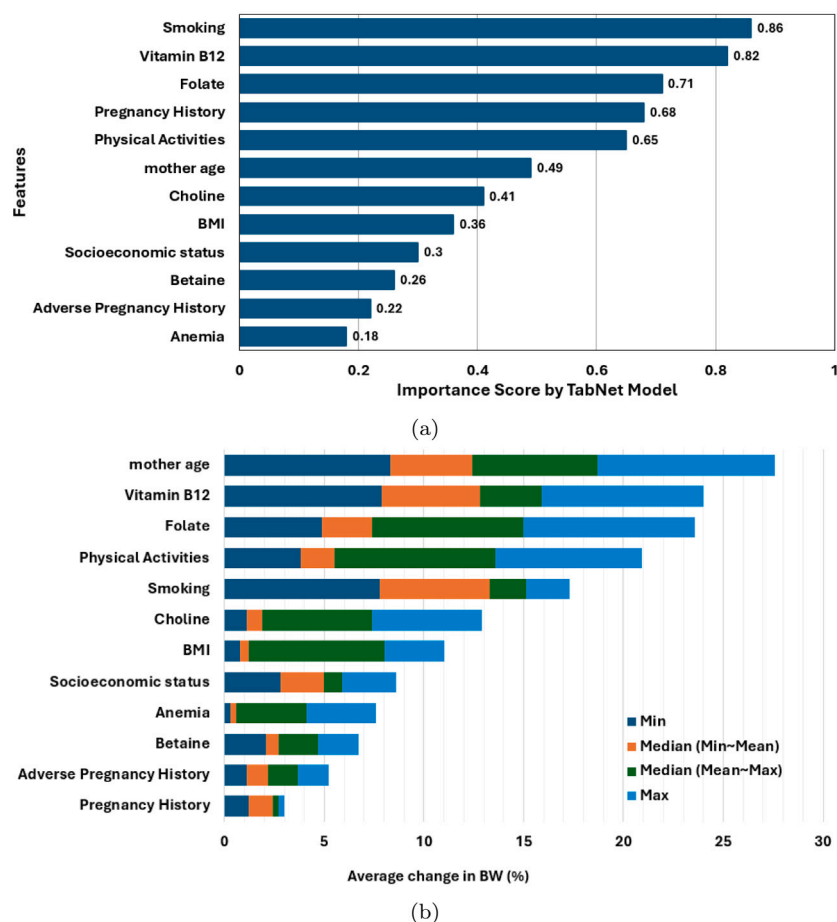


Fig. 5. (a) Feature importance score by TabNet for multimodal maternal factors and (b) Sensitivity analysis showing average changes in BW (%) due to varying each maternal feature individually while holding others at their mean.

only small average changes (approximately 3% or less) in BW. While their quantitative impact is low in our model due to preventive care, these factors are still routinely monitored by clinicians due to their potential association with other risk factors for LBW. These findings have significant clinical implications. They highlight the need to monitor maternal Vitamin B12 and folate status in routine prenatal care, as well as targeted care for at-risk maternal age groups. While factors like adverse pregnancy history and anemia have a limited impact in our analysis, they remain important in clinical settings, where individualized care is essential for comprehensive maternal health management.

We employed SHAP analysis to visualize the contribution and the direction of multimodal maternal factors' influence on neonatal BW in our most effective TabNet model. Fig. 6 illustrates a SHAP summary plot, displaying the SHAP values corresponding to each maternal feature. The y-axis lists the maternal features, while the x-axis represents the impact of each feature across various SHAP values, shown as colored dots. Red dots indicate higher feature values, while blue dots represent lower ones. Each dot corresponds to a SHAP value for a specific feature at a given data point, highlighting its effect on the model's predictions. Additionally, the ranking of feature contributions is evident from the order of the listed features.

Fig. 6 highlights that maternal age is the most significant factor in predicting BW, with higher maternal age being a potential contributor to LBW in neonates. Vitamin B12 is the second most influential feature, and elevated maternal levels of this vitamin were associated with an increased risk of LBW, a finding corroborated by our cohort's clinical data. Conversely, average or slightly reduced maternal vitamin B12 status was linked to healthy fetal development, supporting the hypothesis of our medical team that optimal B12 transfer to the fetus promotes

better outcomes. Following vitamin B12, folate, and smoking emerged as major contributors. High maternal folate status positively impacted BW, while smoking, marked by high SHAP values, indicated maternal smoking throughout pregnancy, which could result in LBW neonates.

The study's analyses of feature importance, sensitivity, and SHAP values consistently identified vitamin B12, folate, smoking, maternal age, and pregnancy history as the most critical and highly correlated maternal factors, with minor variations in ranking across different analyses. Nutritional factors, particularly vitamin B12 and folate, showed a strong and stable contribution to BW, ranking second and third across all analyses. This stability underscores their reliability and predictive power, making them valuable tools for clinicians and gynaecologists during prenatal assessments. This analysis not only demonstrates our model's ability to pinpoint critical prenatal influences but also enhances the interpretability of predictive models in neonatal care. By illustrating how specific maternal factors affect birth outcomes, our study offers valuable insights for healthcare providers, facilitating more informed prenatal interventions and advancing neonatal health research.

#### 4.5. Clinical applications and limitations

Our study presents a predictive model for neonatal BW, designed for integration into web platforms and home-based prenatal care systems. This innovation would allow healthcare professionals to remotely monitor maternal and fetal health, enhancing prenatal care where ultrasound resources are limited.

Although fetal sex is a key determinant of BW, using the model we employed, this information would not be available at 15 GW. Therefore, we did not include it in the model. Nevertheless, the predictive

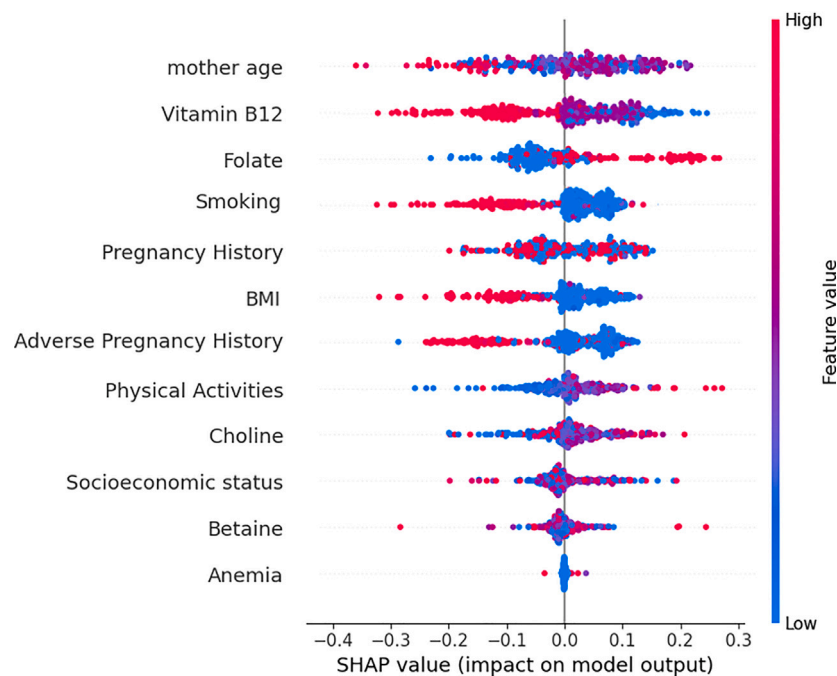


Fig. 6. SHAP summary plot of feature impact on model output and features are ordered by their importance. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

performance of our model was excellent. One limitation of our study is the potential selection bias introduced by the Reus–Tarragona dataset’s eligibility criteria, which may result in the underrepresentation of certain population groups and clinical profiles. While this could affect the model’s generalizability to broader real-world settings, we addressed this concern by validating the model on the independent IIEE Child Birth Weight dataset. This dataset, drawn from a different geographical region and clinical context, allowed us to demonstrate the model’s strong and consistent predictive performance, supporting its potential applicability across diverse populations. Another limitation of our study is the potential bias introduced by the controlled data collection protocols and demographic characteristics of the Reus–Tarragona and IIEE datasets. These datasets were collected in specific clinical environments under defined conditions, which may not fully replicate the variability encountered in real-world settings. Differences in healthcare access, population health status, and data recording practices could affect the model’s performance when deployed broadly. Therefore, while the results are promising, further validation across diverse clinical settings and populations is necessary to confirm the model’s generalizability and practical utility. Furthermore, the IIEE dataset is publicly available; however, the original study does not explicitly report detailed inclusion or exclusion criteria. Additionally, it does not specify whether maternal data were collected at specific gestational ages. To align with our study’s focus on early pregnancy prediction, we carefully selected only those features from the IIEE dataset that are typically available during the first trimester. While this approach allows consistency in input data across cohorts, the lack of clear eligibility criteria remains a limitation that may affect the comparability and generalizability of the results.

However, the applicability of our model is constrained by the study’s relatively small, region-specific cohort. Additionally, by not incorporating genetic data associated with BW [33], our model may miss critical factors that affect prenatal development and birth outcomes. Addressing these limitations will enhance the model’s accuracy and expand its clinical utility.

## 5. Conclusion

Birth weight (BW) is a crucial indicator of neonatal health, with low BW increasing risks of morbidity and mortality. Despite ultrasonography’s role in monitoring fetal health, its limitations in accuracy and accessibility necessitate more reliable predictive systems.

This work addresses these gaps by integrating nutritional, physiological, and lifestyle factors into a neural network-based decision-support system for BW prediction, validated across the Reus–Tarragona cohort and the IIEE childbirth dataset. The TabNet model outperformed conventional methods, improving predictive accuracy significantly when considering maternal vitamin B12 and plasma folate levels alongside other data. This comprehensive approach enhances prenatal care and offers actionable insights for managing low BW, especially valuable in regions with limited ultrasound availability.

Future work will focus on expanding the model to larger and more genetically diverse cohorts to enhance its generalizability and utility globally. Integrating genetic data could refine predictive accuracy further. Pilot testing the model in real-world clinical settings will help evaluate its practical implementation and identify potential improvements. Additionally, incorporating real-time data collection could enable dynamic, personalized prenatal assessments, improving maternal engagement and prenatal care outcomes.

## CRedit authorship contribution statement

**Muhammad Mursil:** Writing – review & editing, Writing – original draft, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation. **Hatem A. Rashwan:** Writing – review & editing, Validation, Methodology, Investigation, Conceptualization. **Adnan Khalid:** Writing – original draft, Methodology, Formal analysis. **Pere Cavallé-Busquets:** Writing – review & editing, Validation, Data curation, Conceptualization. **Luis Santos-Calderon:** Validation, Resources, Data curation. **Michelle M. Murphy:** Writing – review & editing, Validation, Supervision, Investigation, Funding acquisition, Data curation, Conceptualization. **Domènec Puig:** Writing – review & editing, Writing – original draft, Validation, Supervision, Investigation, Funding acquisition, Conceptualization.

## Institutional board and informed consent statement

The research was sanctioned by Sant Joan, Reus, and Joan XXIII, Tarragona, affiliated with Universitat Rovira i Virgili. All participants signed the informed consent form, and both hospitals' Ethics Committees approved the study under the Declaration of Helsinki.

## 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.

## Acknowledgments

Supported by grants from Interministerial Science and Technology Committee (SAF2005-05096); Carlos III Health Institute, National Scientific Research, Development and Technological Innovation Programme Health Investigation Resources (co-financed by European Regional Development Fund, ; to MM, PI10/00335, PI13/02500, PI16/00506, PI19/00844); European Union Horizon 2020 Research and Innovation program (to MM, EPIBRAIN project, funded by the Joint Programming Initiative "A Healthy Diet for a Healthy Life" (ERA HDHL); JFA2 Nutrition and the Epigenome, Horizon2020 grant agreement number 696300, with funding provided by Spanish State Agency for Investigation PCI2018-093098/AEI); Pere Virgili Health Research Institute (to MM, IISPV-2010/21); Biomedical Research Networking Center for the Pathophysiology of Obesity (CIBERobn); AGAUR, Generalitat de Catalunya (Support Research Groups: 2009–1237, 2014–332) and Italfarmaco S.A., Spain.

## Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.jbi.2025.104838>.

## Data availability

The datasets generated during and/or analyzed during the current study are not publicly available because participant consent covers data exploration in response to hypothesis testing within a defined field and with the compromise that this will be vetted by the Principal Investigators, Drs. Murphy and Cavallé-Busquets. They are willing to provide the data to interested parties on reasonable request and agreement that it will be exploited under the terms of participant consent and following further approval by the Ethics Committees if required.

## References

- [1] M. Gueye, A. Sow, D. Boiro, Y.M. Ibrahim, A.C. Bathily, B. Amane, A. Sylla, P.M. Faye, O. Ndiaye, Height and weight development of low-birth-weight infants at 9 months, *Arch. Pédiatr.* 30 (2) (2023) 100–103.
- [2] W.G.N. Targets, Low birth weight policy brief, *World Heal. Organ.* 7 (2014).
- [3] Y. Lu, X. Fu, F. Chen, K.K. Wong, Prediction of fetal weight at varying gestational age in the absence of ultrasound examination using ensemble learning, *Artif. Intell. Med.* 102 (2020) 101748.
- [4] A. Sotiriadis, A. Odibo, Systematic error and cognitive bias in obstetric ultrasound, *Ultrasound Obs. Gynecol.* 53 (4) (2019) 431–435.
- [5] H. Sun, A. Wu, M. Lu, S. Cao, Liability, risks, and recommendations for ultrasound use in the diagnosis of obstetric diseases, *Heliyon* (2023).
- [6] L. Skåren, B. Davies, Å. Bjørnerem, The effect of maternal and paternal height and weight on antenatal, perinatal and postnatal morphology in sex-stratified analyses, *Acta Obs. Gynecol. Scand.* 99 (1) (2020) 127–136.
- [7] J. da Mota Santana, V.A. de Oliveira Queiroz, M. Pereira, E.S. Paixão, S.M. Brito, D.B. Dos Santos, A.M. Oliveira, Associations between maternal dietary patterns and infant birth weight in the nisami cohort: A structural equation modeling analysis, *Nutrients* 13 (11) (2021) 4054.
- [8] R. Obeid, E. Derbyshire, C. Schön, Association between maternal choline, fetal brain development, and child neurocognition: systematic review and meta-analysis of human studies, *Adv. Nutr.* 13 (6) (2022) 2445–2457.
- [9] F. Behrad, M.S. Abadeh, An overview of deep learning methods for multimodal medical data mining, *Expert Syst. Appl.* 200 (2022) 117006.
- [10] S.Ö. Arik, T. Pfister, Tabnet: Attentive interpretable tabular learning, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35, 2021, pp. 6679–6687.
- [11] M. Feng, L. Wan, Z. Li, L. Qing, X. Qi, Fetal weight estimation via ultrasound using machine learning, *IEEE Access* 7 (2019) 87783–87791, <http://dx.doi.org/10.1109/ACCESS.2019.2925803>.
- [12] S. Plotka, M.K. Grzeszczyk, R. Brawura-Biskupski-Samaha, P. Gutaj, M. Lipa, T. Trzcíński, A. Sitek, BabyNet: residual transformer module for birth weight prediction on fetal ultrasound video, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2022, pp. 350–359.
- [13] S.S. Plotka, M.K. Grzeszczyk, et al., Deep learning for estimation of fetal weight throughout the pregnancy from fetal abdominal ultrasound, *Am. J. Obs. Gynecol. MFM* 5 (12) (2023) 101182.
- [14] C. Li, Y. Peng, B. Zhang, W. Ji, L. Li, J. Gong, W. Xia, Y. Li, S. Jin, R. Song, et al., Birth weight prediction models for the different gestational age stages in a Chinese population, *Sci. Rep.* 9 (1) (2019) 10834.
- [15] W. Khan, N. Zaki, M.M. Masud, A. Ahmad, L. Ali, M. Ali, L.A. Ahmed, Infant birth weight estimation and low birth weight classification in United Arab Emirates using machine learning algorithms, *Sci. Rep.* 12 (1) (2022) 12110.
- [16] D.A. Alabbad, S.Y. Ajibi, R.B. Alotaibi, N.K. Alsqr, R.A. Alqahtani, N.M. Felemban, A. Rahman, S.S. Aljameel, M.I.B. Ahmed, M.M. Youldash, Birthweight range prediction and classification: A machine learning-based sustainable approach, *Mach. Learn. Knowl. Extr.* 6 (2) (2024) 770–788.
- [17] A. Ranjbar, F. Montazeri, M.V. Farashah, V. Mehrnough, F. Darsareh, N. Roozbeh, Machine learning-based approach for predicting low birth weight, *BMC Pregnancy Childbirth* 23 (1) (2023) 803.
- [18] J. Park, J. Müller, B. Arora, B. Faybishenko, G. Pastorello, C. Varadharajan, R. Sahu, D. Agarwal, Long-term missing value imputation for time series data using deep neural networks, *Neural Comput. Appl.* 35 (12) (2023) 9071–9091.
- [19] F. Tang, H. Ishwaran, Random forest missing data algorithms, *Stat. Anal. Data Min.: ASA Data Sci. J.* 10 (6) (2017) 363–377.
- [20] I.M. Pires, F. Hussain, N.M. Garcia, P. Lameski, E. Zdravevski, Homogeneous data normalization and deep learning: A case study in human activity classification, *Futur. Internet* 12 (11) (2020) 194.
- [21] Z. Wu, K. Guo, E. Luo, T. Wang, S. Wang, Y. Yang, X. Zhu, R. Ding, Medical long-tailed learning for imbalanced data: bibliometric analysis, *Comput. Methods Programs Biomed.* (2024) 108106.
- [22] T. Wongvorachan, S. He, O. Bulut, A comparison of undersampling, oversampling, and SMOTE methods for dealing with imbalanced classification in educational data mining, *Inf. 14* (1) (2023) 54.
- [23] D. Elreedy, A.F. Atiya, F. Kamalov, A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning, *Mach. Learn.* 113 (7) (2024) 4903–4923.
- [24] Q. An, S. Rahman, J. Zhou, J.J. Kang, A comprehensive review on machine learning in healthcare industry: classification, restrictions, opportunities and challenges, *Sensors* 23 (9) (2023) 4178.
- [25] V. Jackins, S. Vimal, M. Kaliappan, M.Y. Lee, AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes, *J. Supercomput.* 77 (5) (2021) 5198–5219.
- [26] I. Ahmad, M. Basher, M.J. Iqbal, A. Rahim, Performance comparison of support vector machine, random forest, and extreme learning machine for intrusion detection, *IEEE Access* 6 (2018) 33789–33795.
- [27] S. Gündoğdu, Efficient prediction of early-stage diabetes using XGBoost classifier with random forest feature selection technique, *Multimedia Tools Appl.* 82 (22) (2023) 34163–34181.
- [28] F. Bano, S.H. Serbaya, A. Rizwan, M. Shabaz, F. Hasan, H.S. Khalifa, An artificial neural network and Levenberg-Marquardt training algorithm-based mathematical model for performance prediction, *Appl. Math. Sci. Eng.* (2024) 2375529.
- [29] Y. Wang, C. Jin, L. Ma, X. Liu, A robust TabNet-based multi-classification algorithm for infrared spectral data of Chinese herbal medicine with high-dimensional small samples, *J. Pharm. Biomed. Anal.* 242 (2024) 116031.
- [30] S. Watanabe, Tree-structured parzen estimator: Understanding its algorithm components and their roles for better empirical performance, 2023, arXiv preprint arXiv:2304.11127.
- [31] H. Abdi, L.J. Williams, Tukey's honestly significant difference (HSD) test, *Encycl. Res. Des.* 3 (1) (2010) 1–5.
- [32] W.H. Organization, et al., WHO Recommendations on Antenatal Care for a Positive Pregnancy Experience, World Health Organization, 2016.
- [33] D. Alset, D.V. Kubyskhina, E.V. Butenko, I.O. Pokudina, T.P. Shkurat, Association of C677T and A1298C genetic polymorphisms in MTHFR gene with fetal growth restriction, small for gestational age and low birth weight: A meta-analysis, *Hum. Genet* 37 (2023) 201190.