M-TabNet: A Transformer-Based Multi-Encoder for Early Neonatal Birth Weight Prediction Using Multimodal Data

Muhammad Mursil, Hatem A. Rashwan, Luis Santos-Calderon, Pere Cavallé-Busquets, Michelle M. Murphy, and Domenec Puig

***Abstract*— Birth weight (BW) is a key indicator of neonatal health, and low birth weight (LBW) is linked to increased mortality and morbidity. Early prediction of BW facilitates timely prevention of impaired foetal growth. However, available techniques such as ultrasonography have limitations, including less accuracy when applied before 20 weeks of gestation and operator-dependent variability. Existing BW prediction models often neglect nutritional and genetic influences, and focus mainly on physiological and lifestyle factors. This study presents an attention-based transformer model with a multi-encoder architecture for early (< 12 weeks) BW prediction. Our model effectively integrates diverse maternal data, including physiological, lifestyle, nutritional, and genetic data, addressing limitations seen in previous attention-based models such as TabNet. The model achieves a Mean Absolute Error (MAE) of 122 grams and an R^2 value of 0.94, showing its high predictive accuracy and interoperability with our in-house private dataset. Independent validation confirms generalizability (MAE: 105 grams, R^2: 0.95) with the IEEE children dataset. To enhance clinical utility, predicted BW is classified into low and normal categories, achieving a sensitivity of 97.55% and a specificity of 94.48%, facilitating early risk stratification. Model interpretability is reinforced through feature importance and SHAP analysis, highlighting significant influences of maternal age, tobacco exposure, and vitamin B12 status, with genetic factors playing a secondary role. Our results emphasize the potential of advanced deep learning models to improve early BW prediction, offering a robust, interpretable, and personalized tool to identify pregnancies at risk and optimize neonatal outcomes.**

***Index Terms*— Deep learning, Multi-encoder network, Birth weight prediction, Maternal factors, Neonatal health**

I. Introduction

BIRTH weight (BW), defined as the first recorded weight of a neonate within the first hour after birth, is a critical indicator of neonatal health. It significantly influences survival rates, postnatal growth trajectories, and long-term developmental outcomes [1]. The World Health Organization (WHO) categorizes low birth weight (LBW) as less than 2,500 grams and macrosomia as exceeding 4,000 grams, and both conditions are associated with greater risk of adverse postnatal health outcomes [2]. Both are linked to increased susceptibility to infections, higher incidence of chronic diseases such as cardiovascular diseases and diabetes in later life [3] [4], and a mortality risk 20 times greater than that of neonates with normal birth weight (NBW) [5]. Macrosomia is often associated with complications during childbirth, including birth trauma and elevated rates of cesarean sections [6]. Therefore, accurate early prediction of BW is essential for proactive identification of potential health risks, allowing timely medical interventions to improve neonatal outcomes and to mitigate complications associated with abnormal BW.

Traditional BW prediction methods have been primarily based on ultrasound, a technique that, despite its widespread use, has notable limitations related to operator dependency, equipment quality, and accessibility, particularly in low-resource settings [7] [8]. Furthermore, ultrasound becomes more reliable after 24 weeks of gestation, with optimal accuracy between 28 and 32 weeks. Before 20 weeks, its accuracy tends to be lower, making it less feasible for effective maternal health intervention [9] [10]. Ultrasound alone overlooks the complex interplay of maternal physiological, nutritional, lifestyle, and genetic factors that significantly influence neonatal BW [11]–[15]. This gap highlights the pressing need for predictive models to integrate multimodal maternal data, providing earlier, more precise, and interpretable predictions.

In recent years, deep learning (DL) approaches, such as TabNet [22], specifically designed for tabular data, have demonstrated remarkable improvements in predictive accuracy across various clinical applications. TabNet employs an attention mechanism to identify the most salient features at each decision-making step, applying feature masking to ensure sparsity and reduce computational complexity. However, despite its advantages, TabNet has limitations in the effective handling of multimodal data. Its attention mechanism can inadvertently prioritize features from a single data type, potentially neglect-

Muhammad Mursil, Hatem A. Rashwan and Domenec Puig are from the Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili (URV), 43007, Tarragona, Spain. (Corresponding author: Muhammad Mursil, Domenec Puig). (e-mail: muhammad.mursil; hatem.abdellatif; domenec.puig@urv.cat).

Pere Cavallé-Busquets is from the Unit of Obstetrics & Gynaecology, University Hospital Sant Joan, 43204 Reus, IISPV, CIBERObn ISCIII, Spain. (e-mail: pere.cavalle@salutsantjoan.cat).

Luis Santos-Calderon and Michelle M. Murphy are from the Unit of Preventive Medicine & Biostatistics, Faculty of Medicine and Health Sciences, URV, CIBERObn ISCIII, 43201, Reus, Spain (e-mail: luisadolfo.santos; michelle.murphy@urv.cat).

TABLE I: SUMMARY OF DIFFERENT APPROACHES FOR BW PREDICTION AND HOW OUR WORK DIFFERS.

Approach	Key Studies	Method	Limitations	How Our Work Differs
Ultrasound-Based BW Prediction	Li et al. (2019) [16]	Regression models on 2D ultrasound	Accuracy declines before 20 weeks; operator-dependent	We predict BW <12 weeks gestation without ultrasound, using multimodal maternal data.
	Plotka et al. (2023) [17]	Deep learning (BabyNet) on ultrasound	Requires high-quality (HD) imaging and expertise	Our model does not depend on medical imaging, making it more accessible.
	Feng et al. (2019) [18]	Hybrid ML model (SVM + DBN)	Limited generalizability	We use a transformer-based multi-encoder, improving generalization across datasets.
Demographic & Clinical Data-Driven BW Prediction	Alabbad et al. (2024) [19]	ML models (ET, RF, SVR) on hospital datasets	Lacks interpretability and multimodal integration	We incorporate feature importance & SHAP analysis, improving interpretability.
	Khan et al. (2022) [20]	RF model for BW classification	Does not leverage nutritional/genetic data	We include nutritional and genetic factors for a more holistic prediction.
	Ranjbar et al. (2023) [21]	XGBoost on a large dataset	No feature selection insights	Our method ensures each modality contributes optimally using attention-based encoding.
Multimodal Predictive Models (Bridging the Gap)	TabNet (Arik & Pfister, 2021) [22]	Attention-based tabular learning	Struggles with multimodal data fusion	Our multi-encoder architecture improves handling of heterogeneous features.
	Camargo et al. (2023) [23]	Multiple kernel learning	Limited multimodal data integration and rely on HD imaging	Our model integrates multiple modalities simultaneously, enhancing cross-modal relationships for improved prediction.

ing the diversity of critical maternal information, thus reducing both predictive accuracy and interpretability.

To address these challenges, this study introduces a novel, transformer-based multi-encoder architecture to predict neonatal BW at an early gestational stage (<12 weeks). This approach effectively harnesses the rich, multimodal, and multidimensional nature of maternal data, encompassing physiological, nutritional, lifestyle, and genetic factors. Utilizing data from 730 mother-child dyads collected from two public hospitals in the Tarragona region, the proposed model achieves superior predictive performance, with a mean absolute error (MAE) of 122 grams and an r-squared (R^2) value of 0.94. In addition, we classify the predicted BW into low and normal BW categories with a sensitivity of 97. 55% and a specificity of 94.48%, demonstrating strong predictive performance in distinguishing neonatal BW categories. The model's interpretability is enhanced through feature importance, sensitivity analysis and Shapley Additive Explanations (SHAP), which elucidate the direction and magnitude of the impact of maternal factors on BW. This transformative approach represents a significant advancement in early BW prediction, offering a powerful tool for improving neonatal health outcomes through data-driven, personalized maternal care.

The primary contributions of this paper are as follows:

- **Novel Transformer-Based Multi-Encoder Model:** We propose a novel attentive transformer-based multi-encoder architecture for early BW prediction (<12 weeks of gestation), which effectively integrates multimodal maternal data (physiological, nutritional, lifestyle, and genetic). This approach overcomes the limitations of models such as TabNet by ensuring that each modality is appropriately represented and contributes to the prediction process.
- **Comprehensive Analysis of Underexplored Maternal Factors:** Our study provides a unique examination of the combined influence of four key maternal factors, which distinguishes it from existing models. Specifically, it explores the relative contributions of often underexplored factors, such as maternal nutrition and genotype, to neonatal BW prediction, offering a more holistic understanding of maternal influences on foetal development.
- **Clinically Interpretable Framework:** The proposed model ensures clinical applicability by incorporating SHAP analysis, feature importance, and sensitivity analysis. These methods provide transparent insights into how maternal factors influence BW predictions, thereby enabling the development of personalized care strategies and facilitating early intervention decisions.

II. Related work

Accurate prediction of neonatal BW has long been a focus of maternal-foetal research, with traditional approaches relying on ultrasonography and, more recently, clinical data-driven predictive models. Table I provides an overview of the key studies on BW prediction and highlights how our work differs.

A. Ultrasound-based BW prediction

Table I highlights various studies using ultrasound-based techniques for predicting BW. Li et al. [16] used 2D ultrasound data from 19,310 fetuses to develop regression models, but this approach may be affected by operator dependence and declining accuracy before 20 weeks of gestation. Plotka et al. [17] leveraged DL with ultrasound, requiring high-quality imaging expertise. Although effective, these methods still rely on ultrasound technology, which limits their accessibility and applicability. Feng et al. [18] proposed a hybrid machine learning (ML) model combining Support Vector Machine (SVM) and Deep Belief Networks (DBN) to predict BW based on ultrasound data. However, their model has limited generalizability, as it struggles to perform consistently across different datasets.

On the other hand, our model introduces a potential complementary tool to ultrasound by integrating multimodal maternal

data, allowing for the prediction of BW before 12 weeks of gestation. Unlike ultrasound-based methods, which require specialized medical imaging, our approach is more accessible and applicable to a broader population, especially in settings where ultrasound may not be available. Additionally, by using a transformer-based multi-encoder, we improve generalization across diverse datasets, addressing the generalizability limitations seen in recent key studies.

B. Demographic and Clinical Data-Driven BW Prediction

Previous reports primarily use demographic and clinical data to predict BW. Alabbad et al. [19] applied ML models, including Extra Trees (ET), Random Forest (RF), and Support Vector Regression (SVR), to hospital datasets but faced challenges related to multimodal integration and interpretability. Khan et al. [20] used the RF model for BW classification, yet their model did not consider genetic or nutritional data, limiting its predictive accuracy. Taeidi et al [24] developed and evaluated multiple ML models, including DL and RF, achieving an Area Under the Receiver Operating Characteristic Curve (AUROC) of 0.91 with DL as the most effective algorithm, although the study lacked emphasis on model interpretability and external generalizability. Ranjbar et al. [21] used XGBoost on large datasets, but their approach lacked feature selection insights, which could optimize prediction.

In contrast, our model not only integrates demographic and clinical data but also incorporates underexplored maternal factors such as nutrition and genetic information. By doing so, we provide a more comprehensive prediction of BW, addressing the limitations faced by the studies mentioned above. We also improve the interpretability of our model using SHAP analysis, sensitivity analysis, and feature importance techniques, which are not widely used in traditional models.

C. Bridging the Gap: Toward Multimodal Predictive Models

TabNet [22] is a popular attention-based model for tabular data, widely used in predictive medicine for tasks such as disease prediction and patient forecasting [25]. However, it struggles with multimodal data fusion, limiting its performance when integrating diverse data types [26]. Camargo et al. [23] use multiple kernel learning for BW prediction, but this approach processes each data modality separately, missing out the potentially complex relationships between them. Our transformer-based multi-encoder model improves multimodal integration by ensuring that each modality (e.g., nutritional, genetic, and lifestyle data) contributes optimally to BW prediction, overcoming the limitations of TabNet and recent studies. This ensures more accurate early predictions, accounting for the broader context of maternal health for personalized prenatal care.

III. Methodology

A. Mathematical Formulation of the Problem

Let $X_m = \{X_{\text{phys}}, X_{\text{nut}}, X_{\text{lifestyle}}, X_{\text{genetic}}\}$ represent the multimodal input data consisting of physiological, nutritional, lifestyle, and genetic factors, respectively. For each modality $m \in \{\text{phys}, \text{nut}, \text{lifestyle}, \text{genetic}\}$, we have $X_m \in \mathbb{R}^{n \times d_m}$, where n is the number of samples, and d_m is the dimensionality of the modality-specific data.

Each modality X_m is passed through a respective attentive transformer-based encoder E_m, where the encoder learns the important features from each category for BW prediction, $Z_m = E_m(X_m)$. The output of the attentive transformer $Z_m \in \mathbb{R}^{n \times d_e}$, where d_e is the dimensionality of the encoder output. Each encoder is then masked to focus on the relevant features, $\tilde{Z}_m = \text{Mask}(Z_m)$. The masked features for each modality $\tilde{Z}_m$ are passed through a feature transformer T_m to extract higher-level representations, $\hat{Z}_m = T_m(\tilde{Z}_m) \in \mathbb{R}^{n \times d_f}$, where d_f is the dimensionality of the higher-level representation.

The output of the feature transformer of all modalities is fused into a joint representation, $Z_{\text{joint}} = \text{Fusion}(\hat{Z}_{\text{phys}}, \hat{Z}_{\text{nut}}, \hat{Z}_{\text{lifestyle}}, \hat{Z}_{\text{genetic}}) \in \mathbb{R}^{n \times d_{\text{joint}}}$, where d_{joint} is the dimensionality of the joint representation after fusion. Generally in this analysis, all the features are concatenated.

Finally, the joint representation Z_{joint} is passed through a decoder f to predict early neonatal BW, $\hat{y} = f(Z_{\text{joint}}) \in \mathbb{R}$. The predicted BW $\hat{y}$ is then classified as:

$$\hat{c} = \begin{cases} \text{LBW}, & \text{if } \hat{y} < 2500 \text{ grams} \\ \text{NBW}, & \text{otherwise} \end{cases}$$

B. Dataset

1) Reus–Tarragona Birth Cohort: This work is based on data from the Reus–Tarragona Birth Cohort, a prospective longitudinal study designed to track maternal and child health outcomes from early pregnancy through childhood. The cohort is from the Unit of Preventive Medicine and Biostatics at the Faculty of Medicine and Health Sciences, Universitat Rovira i Virgili (URV), in collaboration with the Departments of Obstetrics and Gynecology at Sant Joan University Hospital in Reus and Joan XXIII University Hospital in Tarragona. It is registered at ClinicalTrials.gov under the identifier NCT01778205. The clinical and research teams collected data on various maternal factors, including physiological, nutritional, lifestyle, and genetics, at key gestational time points: $<$ 12 GWs, 15 GWs, 24–27 GWs, and 34 GWs. For this analysis, we focus specifically on the data collected during the early pregnancy phase at $<$12 GWs (median 9 GW). The advantage of such early monitoring is the minimal influence of the physiological effects of pregnancy on blood biomarkers. Additionally, it aligns with the usual timing of the first prenatal check-up, making it an ideal period for implementing early prevention strategies. Pregnant women were recruited based on the following criteria:

Inclusion Criteria: Women with a confirmed singleton pregnancy and a viable fetus less than 12 weeks pregnant at their first prenatal blood sample collection.

Exclusion Criteria: Women were excluded if they were on medication known to affect folate or cobalamin status, had chronic diseases or had undergone surgical procedures affecting nutritional status.

TABLE II: COMPARATIVE STATISTICAL SUMMARY OF MATERNAL HEALTH INDICATORS IN THE REUS-TARRAGONA AND IEEE CHILDBIRTH DATASETS.

Reus-Tarragona Dataset		IEEE Childbirth Dataset	
Feature	**Statistical Description**	**Feature**	**Statistical Description**
Physiological Factors			
Maternal Age	32 ± 4.64	Maternal Age	22 ± 4.28
Previous Pregnancy	Yes: 391 (53.6%), No: 339 (46.4%)	Previous Pregnancy	0.6 ± 0.99
Gestational Weeks	9.1 ± 1.8 GWs	Maternal Height	142 ± 17.24 cm
Adverse Pregnancy History	Yes: 298 (40.8%), No: 432 (59.2%)	Foetal Sex	Male: 698 (51.7%), Female: 652 (48.3%)
Maternal BMI	24.28 ± 4.66 kg/m^2	Initial Systolic BP	105.9 ± 12.33
–	–	Initial Diastolic BP	65.89 ± 7.7
–	–	Final Systolic BP	111.10 ± 13.11
–	–	Final Diastolic BP	70.61 ± 8.57
–	–	Blood Group	A(+): 302 (22.4%), A(-): 64 (4.7%), B(+): 423 (31.3%), B(-): 48 (3.5%), AB(+): 168 (12.4%), AB(-): 39 (2.8%), O(+): 261 (19.3%), O(-): 45 (3.4%)
Nutritional Factors			
Maternal Plasma Folate	31.33 ± 28.90 nmol/L	Initial Hemoglobin	10 ± 1.05 μmol/L
Maternal Vitamin B12	340.03 ± 151.9 pmol/L	Final Hemoglobin	10.45 ± 0.96 μmol/L
Maternal Betaine	15.6 ± 3.83 μmol/L	Blood Sugar	100.66 ± 11.48 mg/dL
Maternal Choline	8.07 ± 1.73 μmol/L	–	–
Maternal Anaemia (Hb $<$11 g/dL)	Yes: 12 (1.6%), No: 718 (98.4%)	–	–
Lifestyle Factors			
Physical Activity	Low: 61.4%, Medium: 37.6%, High: 0.96%	Socioeconomic Status	Below Poverty Line: 67.2%, Above: 32.8%
Tobacco Exposure	No: 73.7%, Yes: 25.8%	–	–
Socioeconomic Status	Lower: 11.6%, Middle: 46.2%, Higher: 42.2%	–	–
Sun Exposure	Never: 29.6%, Sporadic: 54.0%, Regular: 16.4%	–	–
Genetic Factors			
MTHFR C677T	Wild: 33.9%, Hetero+Homo: 66.1%	–	–
MTRR A66G	Wild: 28.9%, Hetero+Homo: 71.1%	–	–
MTHFD1 105TC	Wild: 27.3%, Hetero+Homo: 72.6%	–	–
NOS7 T786C	Wild: 30.4%, Hetero+Homo: 69.6%	–	–
MTR A2756G	Wild: 64.4%, Hetero+Homo: 35.6%	–	–
Target Variable			
Neonatal BW	3230 ± 470 g	Neonatal BW	2700 ± 430 g

Out of 831 pregnant women who initially consented to participate, some pregnancies did not result in a live birth due to factors such as miscarriage, stillbirth, pregnancy termination due to foetal abnormalities, changing hospital or loss to follow-up. A total of 730 mother-infant pairs with complete data on neonatal BW were included in the final analysis. The dataset stems from a cohort of participants recruited from the 2 major public hospitals in Tarragona Province. This supports the broader clinical applicability of our findings, thereby reinforcing its external validity. This ensures that our findings are not limited to a specific demographic but are more widely applicable to diverse populations. See Table II for more details.

2) IEEE Child Birth Weight Dataset: We use the IEEE Child Birth Weight Dataset [27], a meticulously curated collection of maternal and neonatal health parameters, to ensure the proposed model's robustness and generalizability. This dataset is useful for epidemiological studies, predictive modelling, and healthcare interventions related to BW and its factors. Compiled from a well-monitored birth cohort study in Assam, India, the dataset integrates diverse demographic, physiological, and nutritional indicators, offering a holistic view of the maternal health landscape. Data collected under the stringent supervision of medical professionals ensures high reliability, making it a valuable asset for research in perinatal healthcare.

The dataset, as shown in Table II, spans 1,350 pregnant participants and encompasses a spectrum of maternal attributes and district-level variables that exhibit substantial correlations with neonatal BW. These parameters provide a structured framework for investigating prenatal risk factors and optimizing the identification of useful parameters to consider in the research of early-life health outcomes. The dataset's depth and granularity make it particularly suited for advanced analytical techniques, such as ML-based prediction models, statistical inferences, and policy-driven healthcare enhancements.

C. Data Preprocessing

To ensure the dataset was properly curated prior to ML model development and to improve model accuracy and consistency, we performed several data preprocessing steps. We first identified and removed outliers in the Reus-Tarragona and IEEE datasets, such as extreme BWs (e.g., 890 g, 4470 g, 600 g, 1000 g, and 4500 g), using the Interquartile Range (IQR) method. Missing values, which are inevitable in longitudinal pregnancy studies, were handled using different approaches. For the time-series Reus-Tarragona dataset, missing values were imputed using linear interpolation, while categorical missing data were filled using Random Forest imputation. This same Random Forest imputation method was applied to the IEEE childbirth dataset to ensure consistency. We did not use any formal feature selection techniques, as the features selected for this study were primarily based on their availability during early pregnancy ($<$12 GWs), the clinical recommendations provided by healthcare professionals, and their known relevance to neonatal BW. We applied MinMax scaling to standardize the datasets, transforming all values into a range from 0 and 1, ensuring that each feature contributed equally to the model's performance, and preventing any one variable from disproportionately influencing the results. Additionally, to address class imbalance, we applied the Synthetic Minority Over-Sampling with the Gaussian Noise (SMOGN) [28] technique only to the training set. SMOGN helps generate synthetic samples in the underrepresented regions of the datasets, improving the model's ability to learn from these minority data points. Supplementary Figure S1 illustrates the BW distribution before and after applying SMOGN.

Supplementary Table S2 summarizes the predictive model analysis without SMOGN. Comparison of the results from the models with and without SMOGN highlights the impact of SMOGN on the overall predictive performance and calibration of the model. This preprocessing approach ensured a balanced, normalized, and robust dataset for predictive analysis.

D. Transformer-based Multi-Encoder:

The Transformer-based multi-encoder proposed in this study is designed to overcome the limitations of TabNet, particularly in processing multimodal data within the domain of prenatal health. Inspired by TabNet's [22] strength in capturing complex data relationships through sequential decision steps, the model utilizes attention mechanisms to prioritize the most relevant features, using masks to maintain sparsity and reduce computational overhead. While TabNet performs well with single-modal data by processing selected features through a feature transformer (comprising fully connected layers) to extract intricate patterns and create higher-level representations, it struggles with precision and interpretability when dealing with multimodal data. This is because it may focus on the majority of features from only one or two modalities, potentially overlooking critical information from others, see supplementary figure S3. To mitigate these issues, the proposed model enhances TabNet's framework by incorporating transformer-based multiple encoders and multimodal feature fusion techniques tailored for multimodal data processing in prenatal health. This advancement retains TabNet's attention-driven decision-making process. The proposed design significantly boosts the model's capacity to handle diverse input modalities, enabling a more thorough analysis and improved retention of relevant features across various data sources. It causes enhanced interpretability and predictive performance, enabling easier integration into clinical decision support systems. The code for this work is publicly available on the GitHub Repository. The model's architecture is illustrated in Figure 1 and consists of the following key components:

1) Input Layer for Multimodal Data: The model processes inputs from multiple modalities independently. Let the input for modality m be $\mathbf{X}_m \in \mathbb{R}^{d_m}$, where $\mathbf{X}_m$ is the input tensor for modality m and d_m is the number of features.

Each input modality undergoes batch normalization $\mathbf{X}_m^{\text{norm}}$, normalizing the input tensor $\mathbf{X}_m$ to have zero mean and unit variance:

$$\mathbf{X}_m^{\text{norm}} = \frac{\mathbf{X}_m - \mu_m}{\sigma_m}, \tag{1}$$

where μ_m and σ_m are the mean and standard deviation of $\mathbf{X}_m$ across the batch. Batch normalization stabilizes training by ensuring consistent feature distribution. The normalized vector $\mathbf{X}_m^{\text{norm}}$ is used as the initial hidden representation $\mathbf{H}_m^{(1)}$ for each modality in the attention mechanism.

2) Attentive Transformer for Feature Selection: Each modality passes through its attentive transformer, which uses an attention mechanism to select the most relevant features. This block comprises a scaled dot-product attention mechanism followed by sparse feature selection.

Attention Mechanism: For modality m at step t, the attention mechanism computes:

$$\mathbf{H}_{\text{attn},m}^{(t)} = \text{softmax}\left(\frac{\mathbf{Q}_m^{(t)}\mathbf{K}_m^{(t)T}}{\sqrt{d_k}}\right)\mathbf{V}_m^{(t)}, \tag{2}$$

where $\mathbf{Q}_m^{(t)} = \mathbf{h}_m^{(t)}\mathbf{W}_{Q,m}$, $\mathbf{K}_m^{(t)} = \mathbf{h}_m^{(t)}\mathbf{W}_{K,m}$, and $\mathbf{V}_m^{(t)} = \mathbf{h}_m^{(t)}\mathbf{W}_{V,m}$ are the query, key, and value matrices for modality m at step t. $\mathbf{W}_{Q,m}, \mathbf{W}_{K,m}, \mathbf{W}_{V,m} \in \mathbb{R}^{d_h \times d_k}$ are learnable projection matrices and d_k is the key dimensionality.

Sparsemax and Prior Scale: To enforce sparsity, we apply a Sparsemax activation on the output of the attention block, modulated by a prior scale to encourage diversity across steps:

$$\mathbf{M}_{\text{sparse},m}^{(t)} = \text{Sparsemax}\left(\left(\prod_{j=1}^{t-1}(1 - \mathbf{M}_{\text{sparse},m}^{(j)})\right) \cdot \mathbf{H}_{\text{attn},m}^{(t)}\right), \tag{3}$$

where $\prod_{j=1}^{t-1}(1-\mathbf{M}_{\text{sparse},m}^{(j)})$ is the prior scale, encouraging the model to focus on unexplored features over steps. This design is inspired by TabNet and promotes interpretability through sparse selection.

This Sparsemax-based mask is primarily used for stepwise sparse selection of key features across time steps.

3) Masking for Computational Efficiency: To further refine the signal and promote computational efficiency, we apply a second masking operation using a learnable soft gate. This gate allows the model to dynamically suppress or enhance features in a differentiable manner.

The output $\mathbf{H}_{\text{attn},m}^{(t)}$ is passed through a masking network composed of Batch Normalization (BN), a Fully Connected (FC) layer, and a Sigmoid function:

$$\mathbf{M}_{\text{soft},m}^{(t)} = \sigma\left(\text{FC}\left(\text{BN}\left(\mathbf{H}_{\text{attn},m}^{(t)}\right)\right)\right), \tag{4}$$

The resulting mask is applied element-wise:

$$\mathbf{H}_{\text{masked},m}^{(t)} = \mathbf{M}_{\text{soft},m}^{(t)} \odot \mathbf{H}_{\text{attn},m}^{(t)}, \tag{5}$$

This soft masking mechanism allows finer control over information flow and complements the hard feature selection imposed by the Sparsemax block.

Note: The Sparsemax-based mask is applied first to select the most relevant features, followed by the application of the learnable soft mask to refine the feature selection further. In our implementation, the final representation $\mathbf{H}_{\text{masked},m}^{(t)}$ is used for downstream fusion.

4) Feature Transformer: After masking, the features for each modality m at step t are passed through a feature transformer. The feature transformer processes the masked features to extract higher-level representations:

$$\mathbf{H}_{\text{transformed},m}^{(t)} = \text{FeatureTransformer}(\mathbf{H}_{\text{masked},m}^{(t)}), \tag{6}$$

where $\mathbf{H}_{\text{transformed},m}^{(t)}$ is the transformed feature representation for modality m at step t, and FeatureTransformer is a neural network module that applies non-linear transformations to the input features. The output from each step t is processed sequentially, reflecting the recursive nature of the model over

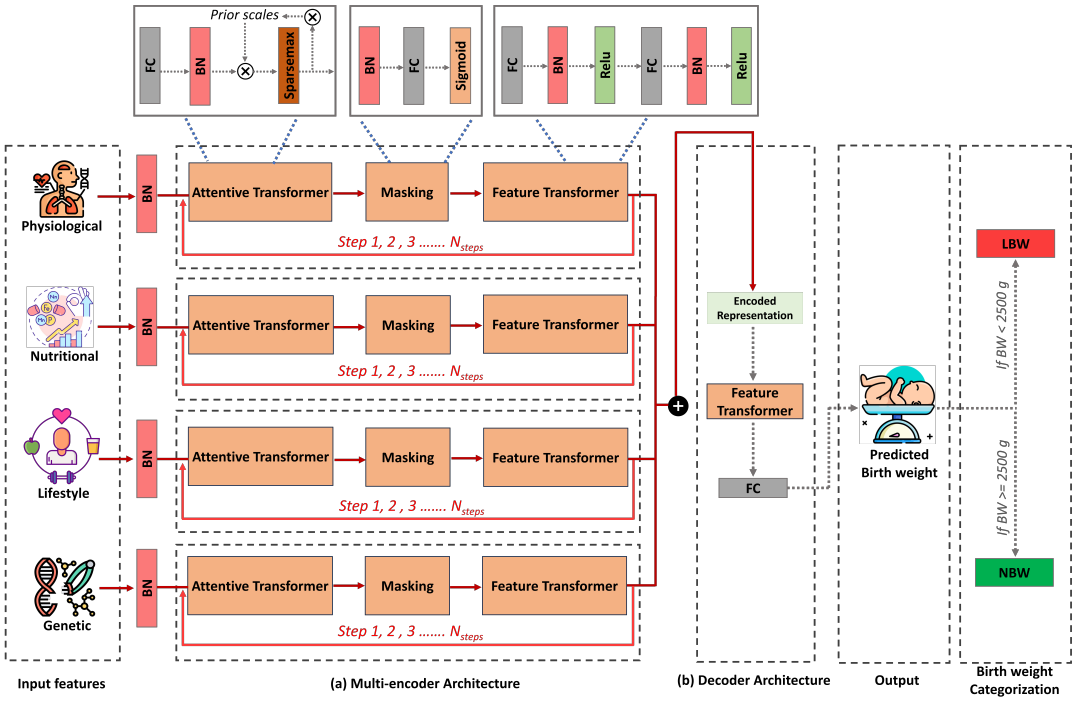

Fig. 1: Proposed Transformer-based Multi-encoder Architecture for Early Neonatal BW Prediction. The architecture integrates multiple encoders to process maternal nutritional, lifestyle, genetic, and health insights, facilitating more accurate prediction of BW and ultimately improving neonatal outcomes.

N_{steps} decision steps. At each step, the feature transformer adapts based on the transformed and masked features from the previous step, enabling progressively refined decision-making.

5) Multimodal Feature Fusion: Following the feature transformer, the features from each modality are concatenated to form a comprehensive representation that captures all the relevant information from each data stream:

$$\mathbf{H}_{\text{fused}} = \text{Fusion}(\mathbf{H}_{\text{transformed},1}, \mathbf{H}_{\text{transformed},2}, \ldots, \mathbf{H}_{\text{transformed},m}), \tag{7}$$

where $\mathbf{H}_{\text{fused}}$ represents the concatenated features.

This concatenation step ensures that the final decision-making process can access a full range of modality-specific features and allows the model to retain the distinct contributions of each modality. This multimodal feature fusion process is crucial in improving the model's performance, especially when the modalities provide diverse but complementary insights for BW prediction.

6) Final Prediction Layer for BW Prediction: The fused features $\mathbf{H}_{\text{fused}}$ are passed through a fully connected (FC) layer with a linear activation function to predict BW:

$$\hat{y} = \mathbf{W}_{\text{out}}\mathbf{H}_{\text{fused}} + \mathbf{b}_{\text{out}}, \tag{8}$$

where $\mathbf{W}_{\text{out}} \in \mathbb{R}^{1\times d_{\text{fused}}}$ is the weight matrix for the output layer, $\mathbf{b}_{\text{out}} \in \mathbb{R}$ is the bias term, d_{fused} is the dimensionality of the fused feature vector. To enhance clinical relevance, the predicted BW $\hat{y}$ is classified into distinct BW groups as follows:

$$\hat{c} = \begin{cases} \text{LBW}, & \text{if } \hat{y} < 2500 \text{ grams} \\ \text{NBW}, & \text{otherwise} \end{cases} \tag{9}$$

E. Model Implementation, training and Evaluation

The model was implemented in Python (version 3.12) and executed on PyCharm. The experiment was run on a Windows system with an Intel(R) Core(TM) i7-5930K CPU @ 3.50 GHz, 32 GB RAM. We used five-fold cross-validation to evaluate the proposed model's performance and generalization. In this approach, 80% of the dataset was divided into five equal subsets. In each iteration, four subsets were used for training and one for validation. This process was repeated five times, ensuring that each sample served as a validation case once. The remaining 20% of the dataset was set aside as a separate test set, which was not used during training or validation. After completing the cross-validation, the test set was used to evaluate the model's final performance. Performance metrics were computed exclusively on the test set after the cross-validation process. We trained our model on the Reus-Tarragona birth cohort and IEEE dataset separately to assess the generalizability of the proposed model. To ensure optimal performance, we employed GridSearchCV for hyperparameter optimization, which allowed us to exhaustively search through a predefined set of hyperparameter values and identify the best-performing combination. Supplementary Table S1 reports the optimal values of the hyperparameters that led to the highest model performance. We used MAE and R^2 as performance metrics for the evaluation. Additionally, sensitivity, specificity, and AUROC and 95% confidence interval (CI),

were calculated based on the model's predictions to evaluate its robustness in distinguishing between BW categories, thereby supporting its reliability in identifying at-risk neonates.

F. Model Interpretability

In this study, model interpretability was achieved using feature importance, sensitivity analysis, and SHAP analysis to gain insights into the factors influencing neonatal BW predictions. Feature importance was used to identify which maternal health factors had the most significant impact on the model's predictions. A sensitivity analysis was performed by holding all features at their mean value and varying one feature at a time. Four representative values, minimum, mean-to-minimum median, mean-to-maximum median, and maximum, were used for each continuous feature, and the average changes in BW (%) from the baseline were measured. For categorical features, sensitivity analysis was assessed on all values to evaluate their impact. This approach helped assess the influence of different features on the model's output. SHAP analysis further provided a detailed understanding of how each feature contributed to individual predictions. Together, these analyses provided a comprehensive approach to interpreting the model, highlighting key insights into neonatal health.

IV. Results and Discussions

A. Model Predictive Analysis

The predictive analysis was conducted on the Reus-Tarragona birth cohort dataset, using neonatal BW as the target label. Three evaluation stages were performed to compare different models: the original TabNet, the proposed M- TabNet with four variations (i.e., with all modalities, without genetic factors, without nutritional factors, and without both genetic and nutritional factors), and the comparative evaluation of M-TabNet against XGBoost, CatBoost, and TabTransformer, as shown in Table III. In the first stage, the TabNet model was trained on individual maternal factors. Physiological data, which is most likely available to clinicians, achieved the best performance among single modalities, with an MAE of 236.4 g, an R^2 value of 0.8182, a sensitivity of 80.12%, a specificity of 76.68% and AUROC of 78.28. Nutritional factors, which are less likely available, followed by an MAE of 280.1 g, an R^2 of 0.7512, a sensitivity of 68.78%, a specificity of 65.84% and AUROC of 69.91. Lifestyle and genetic data, that were most and least availability, respectively, performed worst, with MAE values of 314.9 g and 549.1 g, R^2 values of 0.7035 and 0.2805, sensitivities of 70.22% and 52.67%, specificities of 68.46% and 50.28%, and AUROC values of 71.54 and 56.26 respectively. The low R^2 for genetic data indicates weak predictive power when used in isolation, suggesting the limited individual contribution of the studied genotypes to neonatal BW prediction. Then, all maternal factors (physiological, nutritional, lifestyle, and genetic) were integrated into the TabNet model, improving performance. The MAE dropped to 181.6 g, the R^2 increased to 0.8685, and the sensitivity, specificity and AUROC improved to 86.85%, 84.64%, and 90.72 respectively.

In the second stage, the proposed M-TabNet model was introduced. This model separately processes each modality before integrating them, leading to a substantial performance improvement. M-TabNet achieved an MAE of 122.3 grams, an R^2 of 0.9432, a sensitivity of 97.55%, a specificity of 94.48%, and AUROC of 97.48, demonstrating its superior predictive ability. Further ablation studies revealed that removing genetic data slightly reduced performance (MAE = 151.9 g, R^2 = 0.9069, sensitivity = 93.54%, specificity = 90.63%, AUROC = 95.98), while eliminating nutritional data increased the MAE to 178.5 g (R^2 = 0.8736, sensitivity = 89.91%, specificity = 87.65%, AUROC = 93.08), highlighting its significant predictive contribution. Additionally, excluding both nutritional and genetic factors further deteriorated performance (MAE = 190.2 g, R^2 = 0.8698, sensitivity = 89.91%, specificity = 87.65%, AUROC of 90.88). These results emphasize the complementary role of different maternal factors in neonatal BW prediction. While physiological data alone provides strong predictive power, integrating multiple factors, especially nutritional and lifestyle data, further improves accuracy. Adding genetic features to M-TabNet reduced MAE from 151.9 g to 122.3 g (19.2%) and improved R^2 from 0.9069 to 0.9432 (4.4%). However, given the potential logistical limitations of genotyping (sample availability, specialised preparation and cost), the model still performs robustly without genetic data (MAE = 151.9 g, R^2 = 0.9069), making it a practical option for resource-limited settings. Moreover, M-TabNet's high sensitivity (97.55%), specificity (94.48%) and {AUROC score (97.48) demonstrate its strong ability to distinguish between LBW and NBW neonates, making it highly effective in clinical settings.

In the last stage, we performed a comparative evaluation of M-TabNet against XGBoost, CatBoost, and TabTransformer. CatBoost, XGBoost, and TabTransformer predicted BW with MAEs of 206.4 g, 200.2 g, and 189.5 g, respectively; R^2 scores of 0.8003, 0.8061, and 0.8477; sensitivities of 85.46%, 85.95%, and 86.22%; specificities of 82.97%, 83.07%, and 84.49%; and AUROC values of 88.92, 89.57, and 90.36, respectively. The superior performance of M-TabNet underscores the importance of multimodal data fusion in predictive modeling and establishes it as a complementary tool in clinical decision-making.

Supplementary Figure S2(a) is a scatter plot of the predicted versus actual BW values, demonstrating the model's accuracy. The close alignment of points along the diagonal indicates that the model effectively captures neonatal BW variations. Supplementary Figure S2(b) shows the MAE distribution across the 5-fold cross-validation. The MAE remains consistently below 128 g across folds, confirming the model's stability, robustness and generalizability.

Model Performance Analysis Across Maternal Data: To evaluate the performance of the proposed model in different maternal health settings, we trained it from scratch on the IEEE childbirth dataset, which includes three maternal data modalities (physiological, nutritional, and lifestyle). The proposed model achieved an MAE of 105.4 grams and an R^2 value of 0.95, maintaining high predictive accuracy despite training on a different dataset. This result demonstrates the consistent and efficient performance of the proposed model, suggesting its potential for use in diverse clinical settings with different distributions of maternal data.

TABLE III: PERFORMANCE COMPARISON OF MODELS FOR NEONATAL BW PREDICTION USING MATERNAL FACTORS FROM THE REUS-TARRAGONA AND IEEE DATASETS.

Model	Maternal Factor	Availability to clinicians	MAE (g)	R^2	Sensitivity (%)	Specificity (%)	AUROC (95% CI)
Reus-Tarragona Birth Cohort							
TabNet	Physiological	Most Likely	236.4	0.8182	80.12	76.68	78.28 (0.75 - 0.80)
TabNet	Lifestyle	Most Likely	314.9	0.7035	70.22	68.46	71.54 (0.67 - 0.73)
TabNet	Nutritional	Less Likely	280.1	0.7512	68.78	65.84	69.91 (0.65 - 0.70)
TabNet	Genetic	Least Likely	549.1	0.2805	52.67	50.28	56.26 (0.51 - 0.58)
TabNet	Integrated all four modalities	-	181.6	0.8635	86.85	84.64	90.72 (0.86 - 0.92)
XGBoost	Integrated all four modalities	-	200.2	0.8061	85.95	83.07	89.57 (0.87 - 0.91)
CatBoost	Integrated all four modalities	-	206.4	0.8003	85.46	82.97	88.92 (0.86 - 0.90)
TabTransformer	Integrated all four modalities	-	189.5	0.8477	86.22	84.49	90.36 (0.87 – 0.92)
M-TabNet	**Integrated all four modalities**	-	**122.3**	**0.9432**	**97.55**	**94.48**	**97.48 (0.95 - 0.98)**
M-TabNet	without genetic	-	151.9	0.9069	93.54	90.32	95.98 (0.92 - 0.96)
M-TabNet	without nutritional	-	178.5	0.8736	89.91	87.65	93.08 (0.88 - 0.95)
M-TabNet	without nutritional and genetic	-	190.2	0.8698	89.91	87.65	90.88 (0.86 - 0.93)
IEEE Child Birth Data							
M-TabNet	Integrated all three modalities	-	105.4	0.9502	97.94	96.70	97.63 (0.94 - 0.98)

Paired t-test Analysis: A paired t-test was performed on the results achieved from the Reus-Tarragona cohort to determine statistical significance. The proposed model significantly outperformed TabNet, with a t-statistic of 9.48 and a p-value of 0.0024 ($p < 0.05$) for MAE. The R^2 comparison yielded a t-statistic of -9.79 and a p-value of 0.0022 ($p < 0.05$), confirming that the proposed model provides a more accurate and explicable BW prediction.

Model's Inference Latency: The proposed model's inference latency, incorporating all modalities, was also measured to assess its computational efficiency during use. With the above specified hardware, the average time taken to process a batch and generate predictions was 0.0047 seconds per batch. Given that the batch size was 64, the average latency per sample was approximately 0.000073 seconds (or 73 microseconds). This shows the capability of the model to provide real-time predictions, suitable for clinical settings where fast decision-making is critical.

B. Ablation Studies

To assess the contributions of individual model components, we conducted a series of ablation experiments as shown in Table IV. The base model, which incorporates all four maternal data modalities (physiological, nutritional, lifestyle, and genetic) using attention mechanism, ReLU activation function, MinMax scaling, full training and concatenation for feature fusion, achieved an MAE of 122.3 g.

The attention mechanism played a critical role in feature selection and interaction modeling. Disabling attention led to the most significant performance drop, with the MAE increasing to 391.3 g (+269), demonstrating that the model heavily relies on attention to extract meaningful relationships between features. Other architectural and optimization choices also affected performance. Switching the activation function from ReLU to GELU increased the MAE to 129.5 g (+7.2), indicating that while activation functions contribute to predictive accuracy, their impact is relatively minor. Feature scaling was also evaluated, with Z-score normalization increasing the MAE to 126.2 g (+3.9), suggesting that MinMax preserves relevant feature distributions more effectively. Early stopping provided a minor benefit, reducing the MAE to 119.8 g (-2.5), preventing overfitting while maintaining performance. Finally, we compared different fusion strategies. Replacing concatenation with aggregation increased the MAE to 140 g (+17.7), confirming that concatenation better retains cross-modal information, leading to superior performance.

TABLE IV: ABLATION STUDY ON MODEL COMPONENTS AND TECHNIQUES, SHOWING THE IMPACT OF ATTENTION MECHANISM EXCLUSION, ACTIVATION FUNCTIONS, SCALING, EARLY STOPPING, AND FUSION METHODS ON TEST MAE.

Ablation Cases	MAE (Difference)
M-TabNet (Base Model)	122.3 g
Without Attention Mechanism	391.3 g (+269)
Activation Function (ReLU vs. GELU)	129.5 g (+7.2)
Feature Scaling (MinMax vs. Z-score)	126.2 g (+3.9)
Early Stopping vs. Full Training	119.8 g (-2.5)
Fusion (Concatenation vs. Aggregation)	140 g (+17.7)

C. Model Interpretability Analysis

To understand the decision-making process of the proposed model, we employed feature importance analysis, sensitivity analysis, and SHAP analysis. Figure 2 presents the feature importance scores, showing that maternal age (0.91), tobacco exposure (0.87), and vitamin B12 (0.83) are the most influential factors in predicting neonatal BW. Sun exposure (0.81) and folate (0.71) also contribute significantly, whereas anaemia and MTHFD1 105TC genotype exhibit minimal influence. The relatively low impact of anaemia may be attributed to preventive maternal care protocols in the studied cohort.

Figure 3 illustrates sensitivity analysis results, which quantify the effect of varying maternal factors on BW. Maternal age was associated with the most significant changes (26.7% variation), followed by tobacco exposure (25.3%), folate status (22%), and vitamin B12 status (18.7%). These results emphasize the role of lifestyle and nutritional factors in BW prediction, while anaemia and MTHFD1 105TC contributed minimally ($\leq$3% variation).

Figure 4 presents the SHAP analysis, which provides individualized feature impact on BW predictions. The ranking of predictors aligns with previous analyses, with tobacco use, sun exposure, vitamin B12, and maternal age emerging as key determinants. Tobacco exposure was associated with a reduction of approximately 350 grams in BW, while consistent

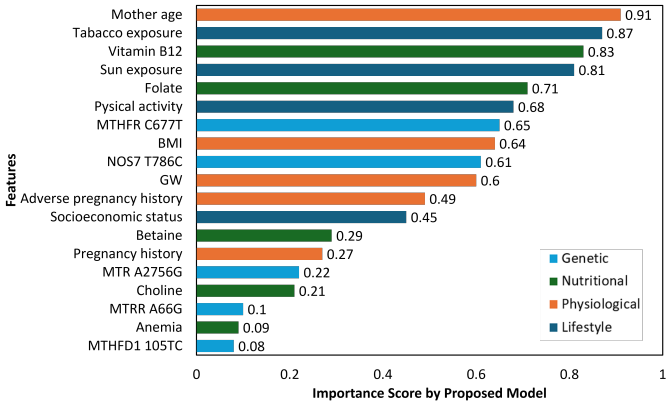

Fig. 2: Feature importance scores for neonatal BW prediction, highlighting age, tobacco use, and sun exposure as dominant factors. Anemia and the MTHFD1 105TC variant show minimal influence.

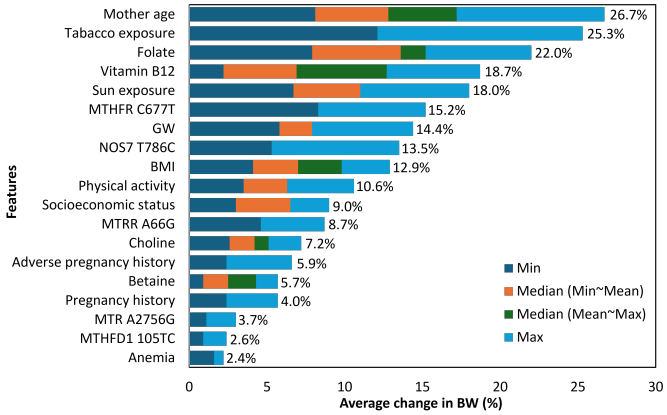

Fig. 3: Sensitivity analysis of maternal factors on neonatal BW shows age, tobacco use, and folate as key influencers, while anemia and MTHFD1 105TC had minimal impact.

sun exposure led to an increase of 410 grams. Vitamin B12 deficiency contributed to a BW increase of 250 grams, whereas maternal age was linked to a BW reduction of 430 grams. Anaemia and MTHFD1 105TC showed minimal impact.

The combined findings from these analyses reinforce the interpretability of our model, offering actionable insights for clinicians. The identification of maternal age, tobacco exposure, and vitamin B12 as dominant predictors suggests potential intervention strategies for mitigating risks associated with LBW. This interpretability allows healthcare professionals to assess risk factors early, personalize maternal care, and improve neonatal health outcomes.

D. Clinical Applications and Limitations

The proposed model can serve as an effective decision-support tool for clinicians, providing early neonatal BW predictions even in clinical settings where ultrasonography is unavailable. Integrating diverse maternal factors enables early identification of at-risk pregnancies, facilitating timely interventions such as nutritional support and lifestyle modifications. A potential application of this model includes a mobile or web-

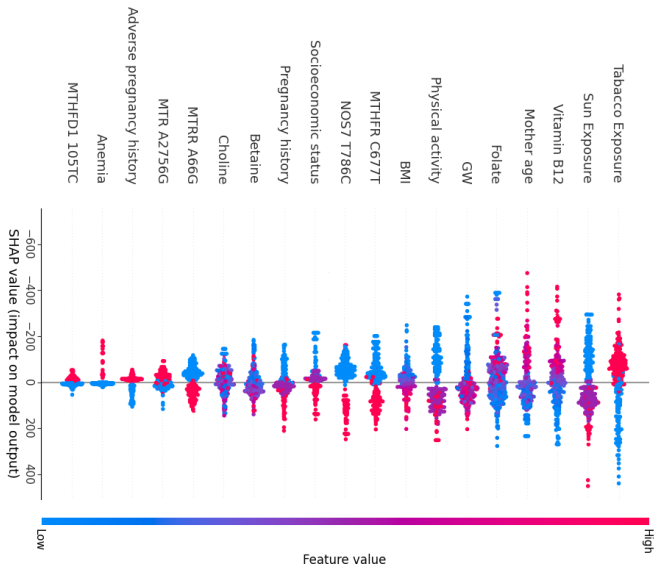

Fig. 4: SHAP analysis shows highest positive and negative impact of tobacco, sun exposure, vitamin B12, and maternal age in neonatal BW prediction.

based platform that allows expectant mothers to monitor risk factors and receive personalized recommendations.

Despite its advantages, this study has some limitations. The relatively small sample size may limit its ability to detect subtle associations, and potential biases due to population stratification or confounding factors could affect model generalizability. Further validation in larger and more diverse cohorts is essential to ensure broader clinical applicability and enhance the robustness of the proposed approach. Gestational age at birth and foetal sex are well-established predictors of BW. However, these parameters are unavailable until the end of pregnancy, excluding their practical use from early prediction models. To assess their impact, we trained models including completed gestational weeks at birth and foetal sex. While this inclusion improved predictive performance (MAE = 107 g, $R^2 = 0.9477$), it did not significantly alter the ranking of top feature importance, which remained strong predictors of BW. This indicates that despite their predictive value, the absence of these variables in early pregnancy does not substantially diminish our model's performance, reinforcing its robustness in real-world clinical settings.

V. Conclusion

This study introduces a novel transformer-based multi-encoder model for the early prediction of neonatal BW using a comprehensive set of multimodal maternal data encompassing physiological, nutritional, lifestyle, and genetic factors. By addressing the limitations of traditional methods, widely used ultrasonography scans, and existing deep learning models, such as TabNet, our proposed model significantly improves both the predictive accuracy and interpretability of BW predictions at an early gestational stage. The model's superior performance, demonstrated by a MAE of 122 grams and an R^2 value of 0.94, highlights its potential to make reliable early predictions, facilitating timely medical interventions.

Furthermore, the study emphasizes the underexplored role of maternal genotype and nutrition, in conventional models. The model's clinical relevance is enhanced by its transparent interpretability, achieved through sensitivity analysis and SHAP, which provide valuable insights into how variations in maternal factors affect neonatal health outcomes. The results suggest that maternal age, tobacco exposure, and vitamin B12 status are the most significant determinants of neonatal BW, with genetic factors playing a secondary role. This work represents a significant advancement in maternal health and child development, offering a promising tool for improving neonatal outcomes, particularly in settings where early intervention could mitigate complications associated with abnormal BWs. Future research may explore further model refinements and real-world clinical validations to expand its applicability and impact.

Acknowledgment

Supported by grants from The Spanish Interministerial Science and Technology Committee (SAF2005-05096); The Carlos III Health Institute, National Scientific Research, Development and Technological Innovation Program Health Investigation Resources, cofinanced by The European Regional Development Fund (10/00335, 13/02500, 16/00506, 19/00844, CIBEROBn) The EPIBRAIN project (funded by the Joint Programming Initiative 'A Healthy Diet for a Healthy Life', JFA2 Nutrition and the Epigenome, Horizon2020 grant agreement number 696300 and The Spanish State Agency for Investigation PCI2018093098/AEI). Pere Virgili Health Research Institute (IISPV2010/21); Agency for Management of University and Research grants, Generalitat de Catalunya (Support to Research Groups: 2009-1237, 2014332) and Italfarmaco S.A., Spain. Predoctoral research fellowship from AGAUR et FI to Muhammad Mursil.

References

[1] M. Z. Islam, M. R. K. Chowdhury, M. Kader, B. Billah, M. S. Islam, and M. Rashid, "Determinants of low birth weight and its effect on childhood health and nutritional outcomes in bangladesh," *Journal of Health, Population and Nutrition*, vol. 43, no. 1, p. 64, 2024.

[2] L. Shen, J. Wang, Y. Duan, and Z. Yang, "Prevalence of low birth weight and macrosomia estimates based on heaping adjustment method in china," *Scientific Reports*, vol. 11, no. 1, p. 15016, 2021.

[3] M. Mursil, H. A. Rashwan, P. Cavallé-Busquets, L. A. Santos-Calderón, M. M. Murphy, and D. Puig, "Maternal nutritional factors enhance birthweight prediction: A super learner ensemble approach," *Information*, vol. 15, no. 11, p. 714, 2024.

[4] V. S. Nakshine and S. D. Jogdand, "A comprehensive review of gestational diabetes mellitus: impacts on maternal health, fetal development, childhood outcomes, and long-term treatment strategies," *Cureus*, vol. 15, no. 10, 2023.

[5] W. H. Organization *et al.*, "Unicef-who low birthweight estimates: levels and trends 2000-2015," World Health Organization, Tech. Rep., 2019.

[6] J. L. Niyi, Z. Li, and F. Zumah, "Association between gestational weight gain and maternal and birth outcomes in northern ghana," *BioMed Research International*, vol. 2024, no. 1, p. 5526942, 2024.

[7] A. Sotiriadis and A. Odibo, "Systematic error and cognitive bias in obstetric ultrasound," *Ultrasound in Obstetrics & Gynecology*, vol. 53, no. 4, pp. 431–435, 2019.

[8] H. Sun, A. Wu, Lu, and Cao, "Liability, risks, and recommendations for ultrasound use in the diagnosis of obstetrics diseases," *Heliyon*, 2023.

[9] N. Dudley, "A systematic review of the ultrasound estimation of fetal weight," *Ultrasound in Obstetrics and Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology*, vol. 25, no. 1, pp. 80–89, 2005.

[10] L. J. Ewington, O. Hugh, E. Butler, S. Quenby, and J. Gardosi, "Accuracy of antenatal ultrasound in predicting large-for-gestational-age babies: Population-based cohort study," *American Journal of Obstetrics and Gynecology*, vol. 232, no. 2, pp. 210–e1, 2025.

[11] D. Ramiro-Cortijo, M. De la Calle, V. Benitez, A. Gila-Diaz, B. Moreno-Jiménez, S. M. Arribas, and E. Garrosa, "Maternal psychological and biological factors associated to gestational complications," *Journal of Personalized Medicine*, vol. 11, no. 3, p. 183, 2021.

[12] M. A. Gebremichael and T. B. Lema, "Effect of behavior change communication through the health development army on birth weight of newborns in ambo district, ethiopia: a cluster randomized controlled community trial," *BMC Women's Health*, vol. 24, no. 1, p. 200, 2024.

[13] L. A. Santos-Calderón, P. Cavallé-Busquets, C. Ramos-Rodríguez, C. Grifoll, A. Rojas-Gómez, M. Ballesteros, P. M. Ueland, and M. M. Murphy, "Folate and cobalamin status, indicators, modulators, interactions, and reference ranges from early pregnancy until birth: the reus–tarragona birth cohort study," *The American Journal of Clinical Nutrition*, vol. 120, no. 5, pp. 1269–1283, 2024.

[14] L. Megaw, T. Clemens, K. Daras, R. B. Weller, C. Dibben, and S. J. Stock, "Higher sun exposure in the first trimester is associated with reduced preterm birth; a scottish population cohort study using linked maternity and meteorological records," *Frontiers in Reproductive Health*, vol. 3, p. 674245, 2021.

[15] S. Kobayashi, F. Sata, and R. Kishi, "Gene-environment interactions related to maternal exposure to environmental and lifestyle-related chemicals during pregnancy and the resulting adverse fetal growth: a review," *Environmental Health and Preventive Medicine*, vol. 27, pp. 24–24, 2022.

[16] 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," *Scientific Reports*, vol. 9, no. 1, p. 10834, 2019.

[17] Płotka and et al, "Deep learning for estimation of fetal weight throughout the pregnancy from fetal abdominal ultrasound," *American journal of obstetrics & gynecology MFM*, vol. 5, no. 12, p. 101182, 2023.

[18] M. Feng, L. Wan, Z. Li, L. Qing, and X. Qi, "Fetal weight estimation via ultrasound using machine learning," *IEEE Access*, vol. 7, pp. 87 783–87 791, 2019.

[19] D. A. Alabbad, S. Y. Ajibi, R. B. Alotaibi, N. K. Alsqer, R. A. Alqahtani, N. M. Felemban, A. Rahman, S. S. Aljameel, M. I. B. Ahmed, and M. M. Youldash, "Birthweight range prediction and classification: A machine learning-based sustainable approach," *Machine Learning and Knowledge Extraction*, vol. 6, no. 2, pp. 770–788, 2024.

[20] W. Khan, N. Zaki, M. M. Masud, A. Ahmad, L. Ali, N. Ali, and L. A. Ahmed, "Infant birth weight estimation and low birth weight classification in united arab emirates using machine learning algorithms," *Scientific reports*, vol. 12, no. 1, p. 12110, 2022.

[21] A. Ranjbar, F. Montazeri, M. V. Farashah, V. Mehrnoush, F. Darsareh, and Roozbeh, "Machine learning-based approach for predicting low birth weight," *BMC Pregnancy and Childbirth*, vol. 23, no. 1, p. 803, 2023.

[22] S. Ö. Arik and T. Pfister, "Tabnet: Attentive interpretable tabular learning," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 35, no. 8, 2021, pp. 6679–6687.

[23] L. Camargo-Marín, M. Guzmán-Huerta, O. Piña-Ramirez, and J. Perez-Gonzalez, "Multimodal early birth weight prediction using multiple kernel learning," *Sensors*, vol. 24, no. 1, p. 2, 2023.

[24] E. Taeidi, A. Ranjbar, F. Montazeri, V. Mehrnoush, and F. Darsareh, "Machine learning-based approach to predict intrauterine growth restriction," *Cureus*, vol. 15, no. 7, 2023.

[25] H. Qi, Y. Hu, R. Fan, and Deng, "Tab-cox: An interpretable deep survival analysis model for patients with nasopharyngeal carcinoma based on tabnet," *IEEE Journal of Biomedical and Health Informatics*, 2024.

[26] W. Li, "Tabnet for high-dimensional tabular data: advancing interpretability and performance with feature fusion," in *IET Conference Proceedings CP915*, vol. 2025, no. 2. IET, 2025, pp. 168–173.

[27] Z. Hussain and M. D. Borah, "Child birth weight dataset," 2022. [Online]. Available: https://dx.doi.org/10.21227/dvd4-3232

[28] P. Branco, L. Torgo, and R. P. Ribeiro, "Smogn: a pre-processing approach for imbalanced regression," in *First international workshop on learning with imbalanced domains: Theory and applications*. PMLR, 2017, pp. 36–50.