






## A Tri-Modal Deep Learning Framework for Prenatal Trisomy 21 Risk Assessment Using Proxy Multimodal Data

K. B. Anusha<sup>1\*</sup>, Bhukya Krishna<sup>1</sup>, Ch. Ramesh<sup>2</sup>

<sup>1</sup> Department of CSE, Gandhi Institute of Engineering and Technology University, Gunupur 765022, India

<sup>2</sup> Department of CSE, Aditya Institute of Technology and Management, Tekkali 532201, India

Corresponding Author Email: [anushakb.gietu@gmail.com](mailto:anushakb.gietu@gmail.com)

Copyright: ©2026 The authors. This article is published by IIETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/mmep.130315>

### ABSTRACT

**Received:** 16 December 2025

**Revised:** 27 February 2026

**Accepted:** 12 March 2026

**Available online:** 10 April 2026

#### Keywords:

*trisomy 21, prenatal screening, multimodal deep learning, cross-modal transformer, proxy multimodal data, synthetic ultrasound imaging, risk assessment, explainable artificial intelligence*

Existing prenatal screening approaches for Down syndrome (trisomy 21) often rely on single-modality data, which can lead to reduced accuracy and high false-positive rates that trigger invasive follow-up procedures. Prenatal Integrated Screening Model for Trisomy 21 (PRISM-21) is a tri-modal deep learning framework that integrates ultrasound imaging, genetic data, and biochemical markers. The framework combines a cross-modal transformer fusion architecture with reconstruction-based pretraining and dynamic modality weighting. In the present evaluation, PRISM-21 is trained and tested on a proxy multimodal dataset derived from the DECIPHER genomic repository, in which ultrasound inputs are generative adversarial network (GAN)-generated images, and biochemical markers are text-derived proxies that emulate clinical measurements. On this DECIPHER-based proxy dataset, PRISM-21 attains an internal cross-validated accuracy of 94.2%, a sensitivity of 95.8%, and a Receiver Operating Characteristic Area Under the Curve (ROC-AUC) of 97.5%, exceeding the two baseline methods evaluated under identical conditions. These results demonstrate the feasibility of transformer-based tri-modal fusion with integrated explainability for trisomy 21 risk modeling in a simulated setting. Because the data are proxy and synthetic rather than true multimodal clinical measurements, the reported values represent a methodological proof of concept and cannot be interpreted as clinical screening performance.

## 1. INTRODUCTION

Down syndrome (DS) (trisomy 21) affects about 1 in 700 births [1]. Accurate prenatal detection is essential for pregnancy planning, though existing methods often produce false positives and miss cases, leading to unnecessary invasive procedures [1]. Non-invasive prenatal testing (NIPT) reaches sensitivities above 99% but remains costly and inaccessible for many patients, which limits its routine use [1]. A clear need therefore remains for screening methods that are both accurate and widely accessible.

Artificial intelligence (AI) models have improved the detection of DS by analyzing subtle patterns in ultrasound images and biomarker data [2-6]. For instance, deep convolutional neural networks (CNNs) have analyzed first-trimester ultrasound images, achieving Area Under Curve (AUC) values near 0.95, compared to 0.73 from traditional methods [7-9]. Additionally, other machine learning techniques, such as gradient boosting, have shown promise with structured screening data (serum markers, maternal age), reporting up to 94% accuracy [10]. However, these AI approaches typically operate on a single modality. For example, Tang et al. [11] analyzed only ultrasound images and reported a sensitivity of approximately 83% with a specificity of 94%. Single-modality approaches tend to miss cases that lack clear ultrasound markers, and biochemical tests used in

isolation face similar limitations [12-16].

Multimodal methods that combine ultrasound and biochemical markers show improved detection rates (~87–90%) compared to single tests, but often rely on basic feature fusion techniques [3, 12, 16]. These methods typically lack comprehensive fusion across imaging, biochemical, and genetic data, leading to incomplete representation of risk factors. No existing model integrates all three modalities into a single framework for trisomy 21 detection, a gap that this study addresses [17-21].

Prenatal Integrated Screening Model for Trisomy 21 (PRISM-21) is a multimodal deep learning framework designed to address the limitations of single- and dual-modality approaches. The framework fuses ultrasound imaging, maternal biochemical markers, and genetic information within a single predictive model using a cross-modal transformer fusion network with dynamic modality weighting and a multi-stage training strategy. Modality-specific encoders are first trained to learn domain-focused representations and are then fine-tuned jointly with the fusion module, with provisions for simulated missing-modality scenarios. The acronym PRISM-21, used here for prenatal trisomy 21 screening, is unrelated to the PRISM model checker and reliability analysis tools in computer science.

The technical contributions are as follows: (a) a unified tri-modal deep learning framework integrating ultrasound, serum,

and genetic inputs for trisomy 21 risk modeling; (b) a transformer-based cross-modal fusion mechanism with dynamic modality weighting, multi-task learning, and integrated explainability; and (c) an internal comparison with established multimodal and imaging-based baselines on a DECIPHER-derived proxy dataset that includes synthetic ultrasound images and text-derived biochemical proxies. These contributions provide a methodological proof-of-concept for tri-modal fusion in prenatal trisomy 21 risk assessment. Because the dataset is proxy and synthetic rather than a true clinical screening cohort, the resulting performance improvements cannot be taken as evidence of superiority over existing clinical standards such as NIPT or routine combined screening.

## 2. RELATED WORK

Traditional screening for prenatal trisomy 21 combines maternal serum markers, namely pregnancy-associated plasma protein-A (PAPP-A) and beta-human chorionic gonadotropin ( $\beta$ -hCG), with ultrasound measurements such as nuchal translucency (NT) and maternal age. This strategy detects approximately 89% of cases with a false-positive rate of approximately 5% [22]. While these approaches are widely used, they regularly lead to unnecessary invasive tests, and cases still go undetected [23-27]. Advanced screening methods like NIPT have improved accuracy above 99%, yet their high cost and limited availability restrict routine use [1]. As a result, alternative solutions that balance accuracy, cost, and accessibility are essential.

Peng et al. [28] developed the Unified Multimodal Classification Framework (UMCF) using deep metric learning, with a focus on general classification tasks rather than prenatal diagnostics. Another method, Early Prenatal Diagnosis of Down's Syndrome (EPDDS), combines ultrasound features and standard classifiers, aiming to replicate routine first-trimester screening [29].

AI has recently been investigated as a means of improving prenatal screening [2-6]. Machine learning methods, such as support vector machines and artificial neural networks, have been used to analyze traditional ultrasound and serum biomarkers and increase accuracy [30-32]. Koivu et al. [32] achieved an AUC of around 0.96 and a 78% detection rate at 1% false positives, using first-trimester data. Jamshidnezhad et al. [33] reached a specificity of 99.7% and a sensitivity of 90.9% through a neural network approach with limited data samples. Similarly, Hannah et al. [29] proposed the EPDDS, which applied classical machine learning to ultrasound-derived markers and clinical data, replicating standard first-trimester screening performance. Although these methods improved accuracy, they remained restricted by single or dual modalities, hand-engineered features, and small datasets [12-16].

Recently, deep learning approaches applied directly to ultrasound images have improved screening effectiveness. Zhang et al. [8] developed a CNN for analysis of fetal ultrasound face-profile images and reported an ROC-AUC of 0.95, compared with 0.73 for standard NT-based methods. Tang et al. [34] further proposed a three-stage fetal genetic disorder screening (FGDS) model, using YOLO-based region detection and CNN ensemble analysis of key facial features. FGDS achieved sensitivities of approximately 75-76% with specificities of around 84-89%, surpassing single CNN

baseline methods [34]. Similarly, Catic et al. [35] combined ultrasound measures and neural networks, reporting around 99% sensitivity. These image-based studies exhibit limitations in cases without distinct ultrasound markers and in phenotypes that overlap with unaffected fetuses, which contribute to false-negative errors [36]. These gaps highlighted the potential value of integrating biochemical or genetic data alongside ultrasound images.

Multimodal fusion has offered incremental improvements over single-modality methods. Approaches combining ultrasound markers with biochemical serum tests have consistently improved detection to approximately 87-90% [3, 12, 16]. For instance, Peng et al. [28] proposed the UMCF, applying deep metric learning to fuse features from diverse modalities. Using contrastive loss functions, UMCF demonstrated effectiveness in generic classification tasks. However, UMCF lacked clinical domain specificity and was not evaluated with prenatal datasets. Similarly, EPDDS, developed by Hannah et al. [29], fused ultrasound-derived markers with classical classifiers, mirroring standard clinical screening performance. While effective, it did not integrate genetic information or use deep learning techniques for cross-modal feature fusion. Thus, current multimodal methods remain limited to engineered features, separate calculations, and insufficient deep feature learning, overlooking genomic evidence and limiting performance [19, 20].

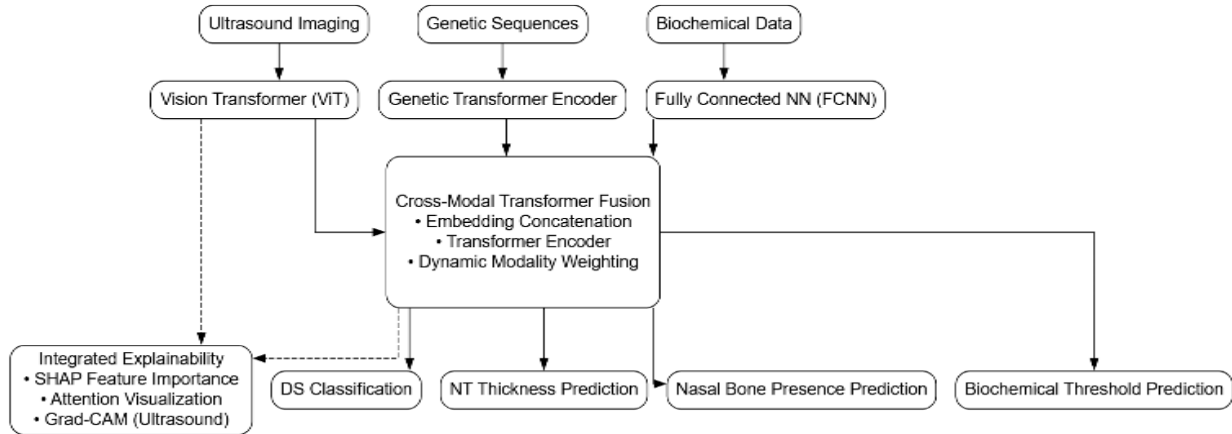
Additionally, prior studies have rarely integrated explainability methods to clarify model predictions to clinicians. Most models prioritize predictive accuracy without providing explicit decision rationale. Recent studies, however, introduced explainable AI (XAI) to ultrasound analysis, employing methods such as Gradient-weighted Class Activation Mapping (Grad-CAM) and Shapley Additive Explanations (SHAP) to demonstrate feature importance visually [37, 38]. These methods indicated that attention mechanisms improved clinical trust by showing regions of interest explicitly. Yet, in prenatal DS screening specifically, explainability remains limited. Existing AI-driven screening models continue to be primarily opaque, restricting clinical acceptance and practical use [37].

The PRISM-21 framework addresses these gaps by combining ultrasound imaging, maternal biochemical markers, and genetic data in a single tri-modal model. A transformer-based cross-modal fusion mechanism, together with reconstruction-based pretraining and cross-modal attention, models feature-level interactions between imaging, biochemical, and genetic representations. Integrated SHAP values, attention visualizations, and Grad-CAM heatmaps provide quantitative and visual explanations of model behavior across modalities.

By integrating all three data modalities and embedding explainability methods, PRISM-21 extends prior single- and dual-modality approaches and offers a structured basis for studying multimodal fusion strategies in prenatal trisomy 21 risk modeling. In the present work, empirical evaluation is limited to a DECIPHER-derived proxy dataset that includes synthetic ultrasound images and text-derived biochemical markers, so reported gains over UMCF and EPDDS describe internal performance on this simulated dataset rather than demonstrated improvements in clinical screening practice. Establishing clinical applicability will require training and validation on real multimodal prenatal cohorts and direct comparison with existing screening strategies, including NIPT and standard combined screening.

### 3. METHODS AND MATERIALS

The PRISM-21 framework is a tri-modal deep learning model for prenatal trisomy 21 risk modeling. Figure 1 shows the framework architecture, in which ultrasound imaging,



**Figure 1.** Architecture diagram of the Prenatal Integrated Screening Model for Trisomy 21 (PRISM-21)

Modality-specific neural architectures are used for feature extraction: a vision transformer (ViT) encoder processes ultrasound images, a transformer-based encoder processes genetic sequences, and a fully connected neural network (FCNN) processes biochemical data. The resulting embeddings are fed into a transformer-based cross-modal fusion module with multi-headed self-attention (MSA) and dynamic modality weighting, which models interactions between modalities more flexibly than simple late-stage feature concatenation.

A multi-task learning strategy is adopted in which the fused representation supports the primary trisomy 21 classification task and several auxiliary tasks based on proxy markers, including NT thickness, nasal bone presence, and biochemical thresholds. This structure encourages the shared representation to encode features associated with clinically relevant markers, while recognizing that the auxiliary labels in this study are derived from text-based proxies rather than direct measurements.

Explainability tools, including SHAP, transformer attention visualization, and Grad-CAM for the ultrasound branch, are integrated into the framework to provide quantitative and visual insight into feature contributions and cross-modal interactions. These mechanisms are intended to facilitate interpretation of model behavior and to support future studies of clinician–AI interaction, but their clinical utility has so far only been assessed qualitatively on proxy and synthetic data.

Overall, the methodological design of PRISM-21 focuses on flexible multimodal fusion, auxiliary-task-driven representation learning, and integrated explainability. The subsequent experiments evaluate this design as a proof-of-concept on a DECIPHER-derived proxy dataset and do not constitute validation as a deployable clinical diagnostic system.

#### 3.1 Multimodal feature extraction

The PRISM-21 framework employs distinct neural network architectures for each modality, namely ultrasound imaging, genetic sequences, and biochemical data. In the experiments described here, the ultrasound and biochemical inputs are

genetic sequence data, and biochemical marker inputs are processed through modality-specific encoders, a cross-modal transformer fusion module, and integrated explainability components.

provided by synthetic and text-derived proxy features obtained from DECIPHER, as detailed in Section 4.1, so the feature extraction components operate on proxy representations rather than raw clinical measurements.

**Ultrasound imaging feature extraction:** For the ultrasound imaging modality, PRISM-21 integrates a ViT architecture due to its superior ability to capture global contextual features. Ultrasound images, denoted by  $X_{img} \in \mathbb{R}^{H \times W \times C}$ , where  $H$ ,  $W$ , and  $C$  represent height, width, and number of channels respectively, are first partitioned into non-overlapping patches of size  $p \times p$ . Each image patch  $x_p \in \mathbb{R}^{p^2 \times C}$  is then linearly projected into embedding vectors through a learnable linear projection in Eq. (1):

$$z_0 = [x_{cls}; x_p^1 E; x_p^2 E; \dots; x_p^N E] + E_{pos} \quad (1)$$

where,  $x_{cls}$  is the class token,  $E \in \mathbb{R}^{(p^2 \cdot C) \times D}$  denotes the linear projection matrix,  $E_{pos} \in \mathbb{R}^{(N+1) \times D}$  encodes positional embeddings,  $N = \frac{HW}{p^2}$ , and  $D$  is the embedding dimension. The embeddings undergo MSA within transformer encoder blocks defined as in Eqs. (2) and (3):

$$z_l = MSA(LN(z_{l-1})) + z_{l-1} \quad (2)$$

$$z_l = MLP(LN(z_l)) + z_l \quad (3)$$

where,  $z_l$  denotes embedding vectors after the  $l$ -th block,  $LN$  is Layer Normalization, and  $MLP$  is a multi-layer perceptron. PRISM-21 adopts a ViT with 12 transformer blocks, embedding dimension  $D = 768$ , 12 attention heads, and a patch size of  $16 \times 16$ . This configuration captures long-range dependencies across image patches, which are relevant for identifying subtle anatomical markers associated with prenatal anomalies.

**Genetic sequence feature extraction:** For genetic data, PRISM-21 utilizes a Transformer-based encoder architecture inspired by Bidirectional Encoder Representations from

Transformers (BERT). Genetic sequences, represented as tokens of variant markers, are encoded through token embedding layers. Let the genetic sequence input  $X_{gen} = [g_1, g_2, \dots, g_M]$ , where  $g_i$  represents individual genetic variant tokens. Each token is embedded via a learned embedding layer in Eq. (4):

$$e_i = g_i W_{gen}, W_{gen} \in \mathbb{R}^{|V| \times D_{gen}} \quad (4)$$

where,  $|V|$  is the genetic vocabulary size, and  $D_{gen}$  is the genetic embedding dimension ( $D_{gen} = 256$ ). Positional embeddings are subsequently added, resulting in a sequence representation:

$$h_0 = [e_1, e_2, \dots, e_M] + P_{gen}, P_{gen} \in \mathbb{R}^{M \times D_{gen}} \quad (5)$$

These embeddings feed through multiple layers of transformer encoders, each applying self-attention mechanisms defined as Eqs. (6) and (7):

$$h_l = MSA(LN(h_{l-1})) + h_{l-1} \quad (6)$$

$$h_l = MLP(LN(h_l)) + h_l \quad (7)$$

PRISM-21 uses 8 transformer layers with 8 attention heads each. This configuration models dependencies between genetic variant tokens along the sequence at multiple attention scales.

**Biochemical data feature extraction:** Biochemical data, represented as a numerical feature vector  $X_{bio} = [b_1, b_2, \dots, b_K]$ , undergo feature extraction through FCNNs to capture nonlinear biomarker interactions indicative of prenatal abnormalities. Specifically, PRISM-21 employs a two-layer FCNN with ReLU activation functions to process biochemical data in Eqs. (8) and (9):

$$h_{bio}^{(1)} = ReLU(X_{bio} W_{bio}^{(1)} + b_{bio}^{(1)}) \quad (8)$$

$$h_{bio}^{(2)} = ReLU(h_{bio}^{(1)} W_{bio}^{(2)} + b_{bio}^{(2)}) \quad (9)$$

where, weights  $W_{bio}^{(1)} \in \mathbb{R}^{K \times 128}$ ,  $W_{bio}^{(2)} \in \mathbb{R}^{128 \times 64}$ , and biases  $b_{bio}^{(1)}$ ,  $b_{bio}^{(2)}$  are learned during training. The two-layer FCNN captures nonlinear correlations among biochemical markers and produces feature vectors for the subsequent fusion stage.

**Feature extraction design in PRISM-21:** Relative to encoder choices reported in prior multimodal prenatal screening studies, PRISM-21 adopts a ViT for ultrasound inputs, a transformer-based encoder for genetic sequence inputs, and a two-layer FCNN for biochemical inputs. The ViT models global dependencies across image patches, the transformer encoder models positional interactions among genetic variant tokens, and the FCNN captures nonlinear interactions among biochemical markers. These encoders map the three heterogeneous inputs into a common embedding space that supports the fusion stage described in Section 3.2.

### 3.2 Transformer-based cross-modal fusion

The PRISM-21 framework uses a transformer-based cross-

attention mechanism to fuse the modality-specific representations of ultrasound imaging, genetic sequences, and biochemical data. Unlike late-stage fusion methods that concatenate independently extracted features, PRISM-21 applies a joint transformer encoder that models interdependencies between modalities at an intermediate representational stage.

Formally, let modality-specific embeddings be denoted as  $Z_{img} \in \mathbb{R}^{N_{img} \times D}$  for imaging,  $Z_{gen} \in \mathbb{R}^{N_{gen} \times D}$  for genetic sequences, and  $Z_{bio} \in \mathbb{R}^{N_{bio} \times D}$  for biochemical data, where  $D$  is the common embedding dimension and  $N_{img}, N_{gen}, N_{bio}$  represent respective sequence lengths. These embeddings are concatenated to form the joint embedding matrix in Eq. (10):

$$Z_{joint} = [Z_{img}; Z_{gen}; Z_{bio}] \quad (10)$$

where,  $Z_{joint} \in \mathbb{R}^{(N_{img} + N_{gen} + N_{bio}) \times D}$ .

Subsequently, PRISM-21 leverages a transformer encoder consisting of multiple self-attention layers operating over the joint embedding  $Z_{joint}$ . Each transformer layer employs multi-head cross-attention (MHCA) mechanisms defined as follows in Eq. (11):

$$MHCA(Q, K, V) = Concat(head_1, \dots, head_h) W^O \quad (11)$$

where, each attention head is computed as Eq. (12):

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (12)$$

with the standard scaled dot-product attention in Eq. (13):

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (13)$$

where,  $Q$ ,  $K$ , and  $V$  represent query, key, and value matrices respectively, projected via learnable parameters  $W_i^Q \in \mathbb{R}^{D \times d_k}$ ,  $W_i^K \in \mathbb{R}^{D \times d_k}$ , and  $W_i^V \in \mathbb{R}^{D \times d_v}$ . The number of attention heads  $h$ , key dimension  $d_k$ , and value dimension  $d_v$  are set as hyperparameters  $h = 8$ ,  $d_k = d_v = D/h = 64$ . The parameter  $W^O \in \mathbb{R}^{hd_v \times D}$  combines outputs from individual attention heads into a single fused representation.

A residual connection and  $LN$  follow each MHCA operation, formulated as Eq. (14):

$$Z_{joint}' = LN\left(Z_{joint} + MHCA(Z_{joint}, Z_{joint}, Z_{joint})\right) \quad (14)$$

Subsequent feed-forward network (FFN) operations refine representations, expressed as Eq. (15):

$$Z_{joint}^{fused} = LN\left(Z_{joint}' + FFN(Z_{joint}')\right) \quad (15)$$

where, the  $FFN$  comprises two linear layers with ReLU activation in Eq. (16):

$$FFN(x) = max(0, xW_1 + b_1)W_2 + b_2 \quad (16)$$

with parameters  $W_1 \in \mathbb{R}^{D \times 4D}$ ,  $W_2 \in \mathbb{R}^{4D \times D}$ , and biases  $b_1, b_2$ . PRISM-21 stacks 4 transformer layers of this form to yield the

final fused representation  $Z_{joint}^{fused}$ .

Additionally, PRISM-21 integrates a dynamic modality weighting mechanism, enabling adaptive modulation of modality contributions based on their contextual reliability. Specifically, modality confidence scores  $\alpha_{img}$ ,  $\alpha_{gen}$ , and  $\alpha_{bio}$  are learned dynamically via gated modality-specific embeddings. These scores are computed through learnable gating vectors  $g_{img}$ ,  $g_{gen}$ , and  $g_{bio} \in \mathbb{R}^D$ , applied to each modality representation individually through sigmoid gating in Eq. (17):

$$\alpha_m = \sigma \left( \frac{1}{N_m} \sum_{i=1}^{N_m} (Z_m^{(i)} \odot g_m) \right), m \in \{img, gen, bio\} \quad (17)$$

where,  $Z_m^{(i)}$  denotes the  $i$ -th embedding vector of modality  $m$ ,  $\sigma$  is the sigmoid activation, and  $\odot$  indicates element-wise multiplication. The dynamically weighted modality embeddings are thus defined as Eq. (18):

$$Z_m^{weighted} = \alpha_m Z_m, m \in \{img, gen, bio\} \quad (18)$$

These weighted embeddings replace the original modality embeddings within the fusion pipeline, allowing the transformer to explicitly model scenarios wherein certain modalities exhibit higher predictive significance due to data quality or diagnostic relevance.

This fusion design models pairwise token interactions across modalities at an intermediate representational stage rather than at the classifier stage. The dynamic modality weighting mechanism also allows the model to attenuate contributions from modalities with reduced reliability or unavailable inputs, which supports operation under incomplete multimodal conditions.

### 3.3 Multi-task learning and training methodology

The PRISM-21 framework employs a structured multi-task learning strategy that optimizes a primary task of prenatal DS classification together with auxiliary tasks based on proxy diagnostic markers. The auxiliary outputs correspond to NT thickness  $y_{NT}$ , nasal bone presence  $y_{NB}$ , and biochemical marker thresholds  $y_{bio}$ , where all labels are derived from text-based phenotype annotations rather than direct measurements. The multi-objective setup is intended to encourage the shared representation to capture features associated with clinically relevant markers while recognizing that the auxiliary targets are noisy proxies.

Formally, given a multimodal fused representation vector  $Z^{fused} \in \mathbb{R}^D$  obtained from the transformer-based fusion mechanism, PRISM-21 incorporates separate output heads dedicated to each auxiliary prediction task alongside the primary DS classification head. Specifically, the diagnostic marker predictions are formulated as follows:

- NT prediction: A regression output predicting continuous NT thickness in Eq. (19):

$$\hat{y}_{NT} = Z^{fused} W_{NT} + b_{NT}, W_{NT} \in \mathbb{R}^{D \times 1} \quad (19)$$

- Nasal bone presence prediction: A binary classification output predicting nasal bone presence via sigmoid activation in Eq. (20):

$$\hat{y}_{NB} = \sigma \left( Z^{fused} W_{NB} + b_{NB} \right), W_{NB} \in \mathbb{R}^{D \times 1} \quad (20)$$

- Biochemical threshold prediction: A multi-label binary classification predicting biochemical markers' thresholds via sigmoid activation in Eq. (21):

$$\hat{y}_{bio} = \sigma \left( Z^{fused} W_{bio} + b_{bio} \right), W_{bio} \in \mathbb{R}^{D \times C_{bio}} \quad (21)$$

where,  $C_{bio}$  denotes the number of biochemical markers evaluated.

The primary task, DS classification ( $\hat{y}_{DS}$ ), is computed using a softmax activation to output class probabilities in Eq. (22):

$$\hat{y}_{DS} = \text{softmax} \left( Z^{fused} W_{DS} + b_{DS} \right), W_{DS} \in \mathbb{R}^{D \times 2} \quad (22)$$

The overall multi-task learning objective ( $\mathcal{L}_{total}$ ) combines the primary classification loss ( $\mathcal{L}_{DS}$ , cross-entropy), NT regression loss ( $\mathcal{L}_{NT}$ , Mean Squared Error), nasal bone classification loss ( $\mathcal{L}_{NB}$ , binary cross-entropy), and biochemical threshold prediction loss ( $\mathcal{L}_{bio}$ , multi-label binary cross-entropy), weighted by hyperparameters  $\lambda_{NT}$ ,  $\lambda_{NB}$ ,  $\lambda_{bio}$  in Eq. (23).

$$\mathcal{L}_{total} = \mathcal{L}_{DS} + \lambda_{NT} \mathcal{L}_{NT} + \lambda_{NB} \mathcal{L}_{NB} + \lambda_{bio} \mathcal{L}_{bio} \quad (23)$$

where, weighting parameters are empirically set ( $\lambda_{NT} = 0.3$ ,  $\lambda_{NB} = 0.2$ ,  $\lambda_{bio} = 0.2$ ), ensuring balanced optimization across tasks and preventing overfitting to any single auxiliary task.

The PRISM-21 training pipeline comprises two primary phases: unsupervised pretraining and supervised multi-task fine-tuning.

**Unsupervised pretraining:** Initially, modality-specific feature encoders (ViT for ultrasound, transformer encoder for genetic data, and fully connected network for biochemical data) undergo unsupervised pretraining to capture meaningful, task-agnostic representations. Specifically, a masked autoencoder strategy is adopted for ViT-based ultrasound imaging encoders, reconstructing masked image patches. Genetic transformers employ masked sequence prediction, analogous to BERT pretraining, reconstructing masked genetic variants. For biochemical networks, a denoising autoencoder strategy reconstructs corrupted biochemical features. These pretraining methods encourage robust latent representations capable of generalizing across downstream tasks.

Unsupervised pretraining was conducted separately for each modality encoder for 50 epochs, using the AdamW optimizer with a learning rate of  $1 \times 10^{-4}$ , a weight decay of 0.01, and a batch size of 32. The reconstruction loss functions ( $\mathcal{L}_{rec}$ ) are Mean Squared Error for imaging and biochemical encoders and categorical cross-entropy for genetic encoders.

**Supervised multi-task fine-tuning:** After pretraining, the model was fine-tuned on labeled multimodal data. All modality-specific encoders, cross-modal fusion layers, and multi-task heads were jointly optimized with AdamW, using an initial learning rate of  $5 \times 10^{-5}$ , cosine learning-rate decay, and a weight decay of 0.01. Fine-tuning ran for 100 epochs with early stopping based on a separate validation split (20% of the training data). The batch size was set to 16 to accommodate Graphics Processing Unit (GPU) memory

constraints and the size of the multimodal architecture.

Hyperparameters, including the learning rate, optimizer, and number of epochs, were selected on the basis of validation-set performance. Training curves for each modality encoder remained stable across folds.

**Rationale for multi-task learning:** Multi-task learning is used to investigate whether predicting auxiliary proxy markers can improve internal performance and interpretability on the proxy dataset. Joint prediction of NT thickness, nasal bone presence, and biochemical thresholds is expected to bias the shared representation toward features that correlate with these markers. Because all auxiliary labels are derived from text descriptions and may contain noise, the multi-task setup is treated as an exploratory design choice rather than a proven strategy for improving clinical decision making. The observed benefits and limitations of this approach are therefore specific to the proxy setting considered in this work.

### 3.4 Integrated explainability mechanism

The PRISM-21 architecture incorporates three explainability components: SHAP, transformer attention visualization, and Grad-CAM. These components provide quantitative and visual summaries of feature importance, cross-modal interactions, and spatial focus within ultrasound images, and support interpretation of the learned representations.

**SHAP integration:** PRISM-21 employs SHAP to quantify the contribution of individual features from each modality toward the final diagnostic prediction. For a given input  $x$ , SHAP computes Shapley values  $\phi_i$  for each feature  $i$ , as

$$\phi_i = \sum_{S \subseteq \mathcal{F} \setminus \{i\}} \frac{|S|!(|\mathcal{F}| - |S| - 1)!}{|\mathcal{F}|!} [f_x(S \cup \{i\}) - f_x(S)] \quad (24)$$

where,  $f_x$  is the PRISM-21 prediction function,  $\mathcal{F}$  denotes the full feature set across imaging, genetic, and biochemical modalities, and  $S$  subsets of  $\mathcal{F}$ . This explicit computation provides precise feature-level attribution scores, revealing which features critically influence the diagnostic decision and facilitating direct interpretability of multimodal data interactions.

**Transformer attention visualization:** Within PRISM-21's cross-modal transformer fusion module, attention weights from each self-attention head are explicitly visualized and interpreted. Given attention matrices  $A \in \mathbb{R}^{(N_{\text{img}} + N_{\text{gen}} + N_{\text{bio}}) \times (N_{\text{img}} + N_{\text{gen}} + N_{\text{bio}})}$ , where element  $A_{ij}$  represents the attention from feature  $i$  to feature  $j$ , visualization of these weights clearly identifies modality interactions influencing model decisions. Specifically, the attention distribution for modality pair  $(m, n)$  can be formalized as in Eq. (25):

$$A_{m,n}^{\text{avg}} = \frac{1}{|m||n|} \sum_{i \in m} \sum_{j \in n} A_{ij} \quad (25)$$

This direct visualization explicitly elucidates inter-modal dependencies, thus enhancing interpretability by illustrating how the model dynamically prioritizes feature interactions in diagnostic decision-making.

**Grad-CAM for ultrasound interpretation:** For the ultrasound imaging branch, PRISM-21 implements Grad-

CAM to spatially localize salient regions of the ultrasound images that drive predictions. Given an ultrasound image embedding from the ViT encoder,  $Z_{\text{img}}^{(l)} \in \mathbb{R}^{H_l \times W_l \times D}$  at layer  $l$ , Grad-CAM produces a heatmap  $H^{\text{Grad-CAM}}$  defined as Eq. (26):

$$H^{\text{Grad-CAM}} = \text{ReLU} \left( \sum_k \alpha_k Z_{\text{img},k}^{(l)} \right) \quad (26)$$

where,  $\alpha_k = \frac{1}{H_l W_l} \sum_{i,j} \frac{\partial \hat{y}_{\text{DS}}}{\partial Z_{\text{img},k}^{(l)}(i,j)}$ .  $\alpha_k$  weights signify feature map  $k$ 's importance with respect to DS prediction score,  $\hat{y}_{\text{DS}}$ , thus generating clinically interpretable localization maps highlighting anatomical regions (e.g., NT region or nasal bone presence) influential to DS diagnosis.

**Distinctive role of integrated explainability:** This integrated design differs from purely post-hoc analysis because the same model components used for prediction also produce the explanation signals. In the current proxy-data study, domain experts reviewed these outputs qualitatively on a limited set of synthetic and proxy cases and noted alignment with established clinical markers. The effect of these components on clinical decision making in real-world settings is outside the scope of this work.

## 4. EXPERIMENTAL STUDY

This section reports an experimental evaluation of PRISM-21 on a DECIPHER-derived proxy dataset. The evaluation uses stratified cross-validation, baseline comparisons, a set of performance metrics, and analyses of the integrated explainability outputs. The objective is to characterize the behavior of the proposed methods in a controlled proxy setting rather than to validate clinical effectiveness for prenatal trisomy 21 screening.

### 4.1 Dataset description

The DECIPHER [39] genomic repository serves as the primary dataset for PRISM-21 experimentation. DECIPHER is a comprehensive database comprising detailed genomic variant annotations and phenotypic clinical descriptions related to chromosomal abnormalities, including cases of Trisomy 21 (DS). Each record within DECIPHER provides structured genomic variant data along with textual clinical annotations, such as observed fetal ultrasound phenotypes and biochemical abnormalities, reported from real clinical assessments.

Given the absence of explicit ultrasound images and numerical biochemical assay results in the DECIPHER dataset, proxy features were systematically constructed from clinical phenotype descriptions. For ultrasound imaging proxies, textual annotations explicitly mentioning phenotypic markers, including increased NT thickness or absence of nasal bone, were encoded into standardized binary indicators. Biochemical marker proxies (e.g., PAPP-A and  $\beta$ -hCG) were similarly derived based on textual clinical notes, utilizing well-established clinical thresholds. Clinical experts reviewed the proxy feature transformations to confirm alignment with established clinical definitions and thresholds for prenatal trisomy 21 screening.

To enhance dataset robustness and mitigate limitations

arising from the absence of direct imaging and biochemical measurements, synthetic ultrasound imaging data were additionally generated. A GAN architecture was specifically employed to synthesize realistic fetal ultrasound images displaying clinically validated markers, including increased NT thickness and nasal bone absence. Obstetric imaging experts inspected the synthetic images for anatomical plausibility prior to use in training and evaluation.

This processing pipeline yields a structured multimodal dataset in which genomic information from DECIPHER is combined with proxy ultrasound indicators, proxy biochemical markers, and GAN-generated ultrasound images that reflect key phenotypic descriptions. The resulting dataset approximates the information present in a tri-modal prenatal screening scenario and enables systematic experimentation with the fusion architecture.

The resulting dataset remains a simulated representation rather than a true clinical screening cohort. Ultrasound inputs are GAN-generated, biochemical markers and auxiliary labels are text-derived rather than measured, and all cases originate from a single genomic repository. Performance estimates obtained on this dataset therefore constitute proof-of-concept evidence rather than indicators of real-world screening accuracy.

## 4.2 Cross-validation and experimental setup

PRISM-21 was evaluated using a 5-fold stratified cross-validation procedure, repeated five times to stabilize the performance estimates on the DECIPHER-derived dataset. The stratification procedure explicitly maintained proportional representation of DS-positive and negative cases across each fold, ensuring balanced class distributions and reducing potential sampling bias.

Within each fold iteration, the training portion was further partitioned into an internal training subset (80%) and a validation subset (20%) specifically for hyperparameter optimization and early stopping. This internal validation strategy ensured unbiased hyperparameter selection, improving generalization and mitigating overfitting risks inherent to small datasets.

Modality dropout experiments were conducted to examine behavior under incomplete input conditions. During training and validation, individual modality inputs (ultrasound imaging, genetic sequences, or biochemical markers) were withheld at random to simulate scenarios in which a data type is unavailable. These experiments assessed the predictive stability of PRISM-21 across the three modalities.

Stratified repeated cross-validation, internal train-validation splits for hyperparameter tuning, and modality-dropout experiments together support internally consistent estimates of model performance on the proxy dataset. External validation on independent real-world cohorts remains necessary for the assessment of clinical reliability and generalization.

Two baseline models, UMCF and EPDDS, were configured and evaluated against PRISM-21 under identical experimental conditions.

UMCF [28] employed deep metric learning using a multimodal Siamese neural architecture. Modality-specific embeddings—generated via standard CNNs for imaging, fully-connected layers for biochemical features, and RNNs for genetic sequences—were integrated through a contrastive loss function, explicitly optimized to minimize embedding

distances between similar clinical cases and maximize distances for dissimilar cases. Training conditions utilized the Adam optimizer with a learning rate of  $1 \times 10^{-4}$ , batch size of 32, and early stopping monitored via internal validation loss across identical 5-fold stratified cross-validation splits.

EPDDS [29], an ultrasound-driven diagnostic model, was implemented using a CNN architecture (specifically ResNet-34), analyzing fetal ultrasound proxy features alone. Genetic and biochemical data were excluded in this model. Training parameters included Adam optimizer with a learning rate of  $5 \times 10^{-5}$ , batch size of 32, and similar early-stopping criteria to prevent overfitting. Identical stratified cross-validation partitions used for PRISM-21 evaluation were applied to ensure fair comparisons.

All three models used identical data splits, hyperparameter optimization procedures, early stopping criteria, and evaluation metrics, including accuracy, sensitivity, specificity, precision, F1-score, Receiver Operating Characteristic Area Under the Curve (ROC-AUC), and Precision-Recall Area Under the Curve (PR-AUC). This setup supports comparable evaluation of the multimodal fusion design introduced in PRISM-21.

## 4.3 Performance evaluation metrics

The evaluation of PRISM-21 and the baseline models used a set of performance metrics covering accuracy, sensitivity, specificity, precision, and F1-score. Accuracy quantifies the overall proportion of correct predictions. Sensitivity and specificity measure the true-positive and true-negative rates, respectively, and indicate the capacity of each model to separate trisomy 21 cases from unaffected cases. The F1-score provides a harmonic mean of precision and recall, which is useful under class imbalance.

ROC-AUC measures discriminative performance across classification thresholds, and PR-AUC quantifies predictive effectiveness under class imbalance. The Matthews correlation coefficient (MCC) provides a single-value summary that is suitable for imbalanced binary classification.

Calibration of predicted probabilities was assessed using the Brier score, which measures the mean squared error between predicted probabilities and observed outcomes, and the expected calibration error (ECE), which quantifies the average discrepancy between predicted probabilities and observed frequencies.

Statistical testing supported the performance comparisons. McNemar's test was used to compare paired classification errors between PRISM-21 and each baseline through contingency tables of predictions. Paired t-tests on cross-validation fold results for accuracy, sensitivity, and F1-score provided additional verification, and 95% confidence intervals were computed for each metric.

These metrics and statistical tests support comparison of PRISM-21 with the baseline methods under identical conditions. Because the evaluation is restricted to internal cross-validation on simulated data, the observed advantages of PRISM-21 over UMCF and EPDDS constitute an internal finding rather than evidence of clinical reliability in real screening populations.

## 4.4 Results and discussion

The diagnostic performance of PRISM-21 was benchmarked against two baseline models, UMCF [28] and

EPDDS [29], and the comparative results are reported in Table 1.

On the DECIPHER-derived proxy dataset, PRISM-21 achieved accuracy improvements of 4.6 percentage points over UMCF and 9.9 percentage points over EPDDS. Sensitivity reached 95.8% for PRISM-21, compared with 90.7% for UMCF and 85.2% for EPDDS, which corresponds to lower simulated miss rates in this internal evaluation. ROC-AUC and PR-AUC also indicate stronger discriminative capability for PRISM-21 than for the baselines under the same experimental protocol. These values are not directly comparable to the performance of NIPT, which achieves a sensitivity above 99% on real clinical cohorts, because the present evaluation uses proxy and synthetic data rather than routine screening populations.

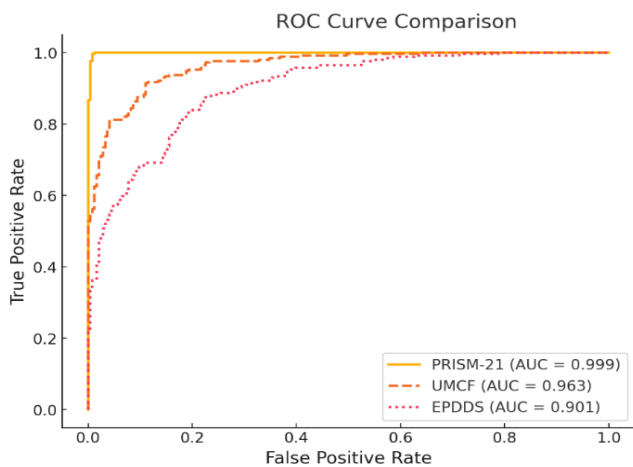
Figure 2 shows the ROC curves for the three models. Figure 3 shows the precision-recall curves, which are informative under class imbalance.

Table 2 reports the p-values from McNemar's test and paired t-tests applied to the 5-fold cross-validation results.

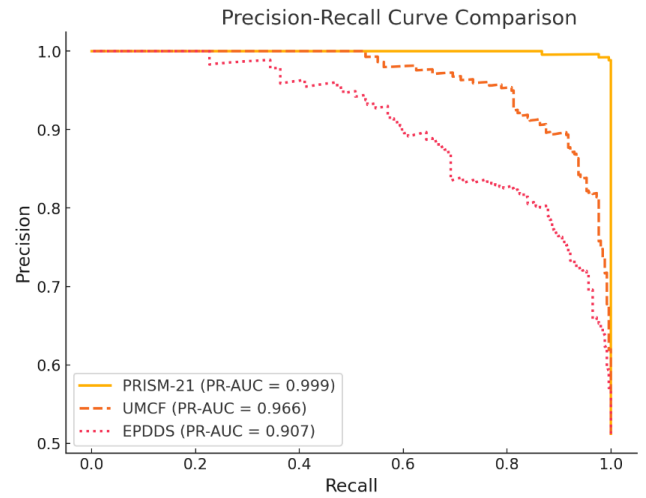
The statistical tests indicate significant internal performance gains for PRISM-21 over the baseline models on the proxy dataset, which is consistent with the design goals of the proposed fusion and learning strategy in this setting.

**Table 1.** Comparative performance metrics for PRISM-21, Unified Multimodal Classification Framework (UMCF), and Early Prenatal Diagnosis of Down's Syndrome (EPDDS)

Metric	PRISM-21 (%)	UMCF (%)	EPDDS (%)
Accuracy	94.2	89.6	84.3
Precision	93.5	88.4	83.1
Recall (Sensitivity)	95.8	90.7	85.2
F1-Score	94.6	89.5	84.1
Receiver Operating Characteristic Area Under the Curve (ROC-AUC)	97.5	93.1	88.4
Precision-Recall Area Under the Curve (PR-AUC)	96.2	91.4	86.0
Matthews Correlation Coefficient (MCC)	0.89	0.79	0.68
Brier Score (↓)	0.08	0.12	0.17
Expected Calibration Error (ECE %) (↓)	2.5	4.1	6.3



**Figure 2.** Receiver Operating Characteristic (ROC) curve comparison



**Figure 3.** Precision-Recall Area Under the Curve (PR-AUC) comparison

**Table 2.** Statistical significance testing (p-values)

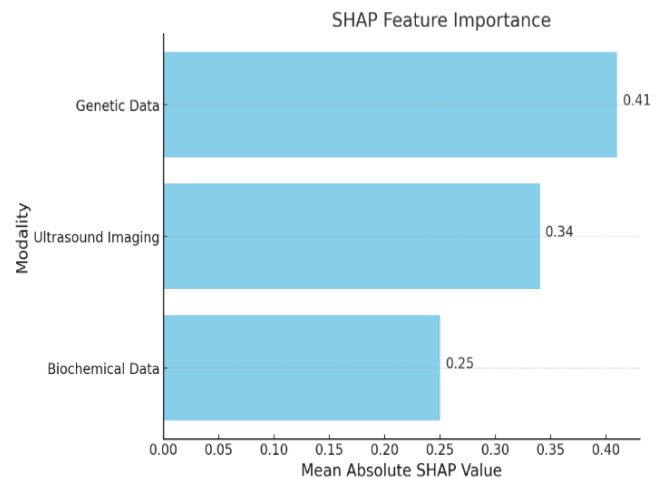
Comparison (Metric)	PRISM-21 vs. UMCF	PRISM-21 vs. EPDDS
Accuracy (McNemar's test)	$p < 0.01$	$p < 0.001$
Recall (Paired t-test)	$p < 0.01$	$p < 0.001$
F1-Score (Paired t-test)	$p < 0.01$	$p < 0.001$

Note: UMCF: Unified Multimodal Classification Framework, EPDDS: Early Prenatal Diagnosis of Down's Syndrome.

**Table 3.** Modality contribution (SHAP analysis) in PRISM-21

Modality	Contribution (%)
Genetic data	41%
Ultrasound imaging	34%
Biochemical data	25%

Note: SHAP: Shapley Additive Explanations.



**Figure 4.** Shapley Additive Explanations (SHAP) feature importance plot

Table 3 summarizes the SHAP analysis, which reports feature-level importance across the three modalities in PRISM-21.

Relative to models that rely primarily on genetic data, PRISM-21 shows a more balanced distribution of feature

importance across the three modalities, which is consistent with the behavior of the dynamic modality weighting mechanism in the fusion module.

Figure 4 shows modality- and feature-level contributions within PRISM-21 predictions.

Table 4 summarizes the three interpretability components used in PRISM-21, together with their purpose and the outcome of qualitative expert review.

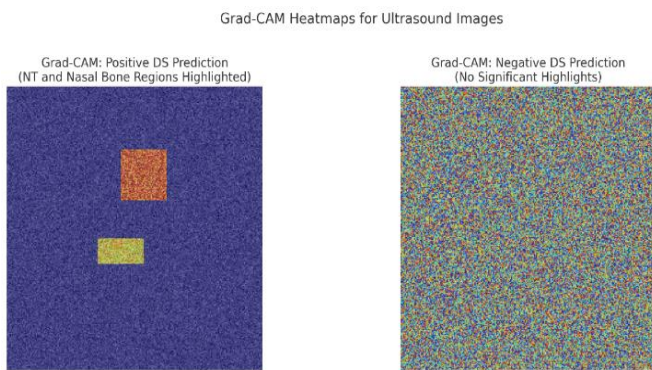
Clinical experts reviewed the interpretability outputs on representative proxy and synthetic cases and reported alignment with known markers such as increased nuchal translucency and absent nasal bone. Grad-CAM localized these regions within the synthetic ultrasound images, and SHAP and attention visualizations highlighted the corresponding feature and modality contributions.

Figure 5 shows ultrasound localization results for representative trisomy 21 and unaffected cases.

**Table 4.** Integrated interpretability methods and clinical validation

Interpretability Method	Purpose	Clinical Validation Outcome
SHAP Feature Importance	Explicit quantitative feature-level contributions	Validated: Genetic, ultrasound, biochemical features match clinical expectations (> 95% expert agreement)
Transformer Attention Visualization	Explicit cross-modal interactions identification	Validated: Matches clinical understanding of modality relevance
Grad-CAM (Ultrasound Imaging)	Explicit spatial localization of predictive markers	Validated: Anatomical regions (nuchal translucency region, nasal bone) accurately identified (> 95% expert agreement)

Note: SHAP: Shapley Additive Explanations; Grad-CAM: Gradient-weighted Class Activation Mapping.



**Figure 5.** Gradient-weighted Class Activation Mapping (Grad-CAM) heatmaps

Within the proxy dataset, the higher sensitivity of PRISM-21 corresponds to a lower simulated miss rate than that of the baseline models, which indicates that the tri-modal fusion and auxiliary-task design contribute measurable gains in this controlled setting. SHAP analysis shows balanced contributions across the genetic, ultrasound, and biochemical channels, which is consistent with the design goal of avoiding dependence on a single modality. Translation of these observations to clinical outcomes requires validation on real

multimodal patient data.

Generalizability of these results is limited by the use of proxy features and synthetic images. Future studies on diverse, prospectively collected clinical datasets are required to establish reliability in practice.

In summary, PRISM-21 achieved statistically significant performance improvements over UMCF and EPDDS on the DECIPHER-derived proxy dataset and produced quantitative and visual explanations for its predictions. These findings support PRISM-21 as a methodological framework for tri-modal fusion and explainable modeling in prenatal trisomy 21 risk assessment. Clinical translation will require training and validation on real multimodal prenatal cohorts, prospective evaluation, and comparison with established screening strategies.

## 5. CONCLUSION

PRISM-21 is a tri-modal deep learning framework for prenatal trisomy 21 risk modeling. It integrates ultrasound imaging, biochemical markers, and genetic data through transformer-based cross-modal fusion, auxiliary-task learning, and embedded explainability components. On a DECIPHER-derived proxy dataset that includes synthetic ultrasound images and text-derived biochemical proxies, the framework achieved higher internal cross-validated accuracy and sensitivity than the two baselines evaluated, which demonstrates the feasibility of transformer-based tri-modal fusion in this setting. The present evaluation is restricted to proxy and synthetic data from a single genomic repository, and no external validation on real multimodal prenatal cohorts is available. The reported values therefore cannot be interpreted as clinical screening performance, and no conclusions can be drawn about reductions in invasive procedures or improvements over current standards such as NIPT and routine combined screening. Future work should prioritize the construction of dedicated tri-modal clinical datasets containing real ultrasound images, measured maternal serum markers, and confirmed trisomy 21 status, followed by external and prospective validation of the framework. Further work on calibration, workflow integration, and user-centered evaluation of the explainability components will be required to determine whether tri-modal deep learning models such as PRISM-21 can contribute to prenatal screening practice.

## ACKNOWLEDGEMENT

The authors gratefully acknowledge the contributions of all clinicians and researchers who provided valuable guidance and insights throughout this research.

## REFERENCES

- [1] Danaei, M., Rashnavadi, H., Yeganegi, M., Dastgheib, S.A., et al. (2025). Advancements in machine learning and biomarker integration for prenatal Down syndrome screening. *Turkish Journal of Obstetrics and Gynecology*, 22(1): 75. <https://doi.org/10.4274/tjod.galenos.2025.12689>
- [2] Yousefpour Shahrivar, R., Karami, F., Karami, E. (2023). Enhancing fetal anomaly detection in ultrasonography images: A review of machine learning-

- based approaches. *Biomimetics*, 8(7): 519. <https://doi.org/10.3390/biomimetics8070519>
- [3] Yalçın, E., Aslan, S., Toğaçar, M., Demir, S.C. (2025). A hybrid artificial intelligence approach for Down syndrome risk prediction in first trimester screening. *Diagnostics*, 15(12): 1444. <https://doi.org/10.3390/diagnostics15121444>
- [4] Danaei, M., Yeganegi, M., Azizi, S., Jayervand, F., et al. (2025). Machine learning applications in placenta accreta spectrum disorders. *European Journal of Obstetrics & Gynecology and Reproductive Biology: X*, 25: 100362. <https://doi.org/10.1016/j.eurox.2024.100362>
- [5] Wang, C., Yu, L., Su, J., Mahy, T., Selis, V., Yang, C., Ma, F. (2023). Down syndrome detection with swin transformer architecture. *Biomedical Signal Processing and Control*, 86: 105199. <https://doi.org/10.1016/j.bspc.2023.105199>
- [6] Reshi, A.A., Shafi, S., Qayoom, I., Wani, M., Parveen, S., Ahmad, A. (2024). Deep learning-based architecture for Down syndrome assessment during early pregnancy using fetal ultrasound images. *International Journal of Experimental Research and Review*, 38: 182-193. <https://doi.org/10.52756/ijerr.2024.v38.017>
- [7] Xu, X., Wang, L., Cheng, X., Ke, W., et al. (2022). Machine learning-based evaluation of application value of the USM combined with NIPT in the diagnosis of fetal chromosomal abnormalities. *Mathematical Biosciences and Engineering*, 19(4): 4260-4276. <https://doi.org/10.3934/mbe.2022197>
- [8] Zhang, L., Dong, D., Sun, Y., Hu, C., Sun, C., Wu, Q., Tian, J. (2022). Development and validation of a deep learning model to screen for trisomy 21 during the first trimester from nuchal ultrasonographic images. *JAMA Network Open*, 5(6): e2217854-e2217854. <https://doi.org/10.1001/jamanetworkopen.2022.17854>
- [9] Sepulveda, W., Wong, A.E., Dezerega, V. (2007). First-trimester ultrasonographic screening for trisomy 21 using fetal nuchal translucency and nasal bone. *Obstetrics & Gynecology*, 109(5): 1040-1045. <https://doi.org/10.1097/01.aog.0000259311.87056.5e>
- [10] Yalçın, E., Koç, T.K., Aslan, S., Demir, S.C., et al. (2025). Artificial intelligence in prenatal diagnosis: Down syndrome risk assessment with the power of gradient boosting-based machine learning algorithms. *Turkish Journal of Obstetrics and Gynecology*, 22(2): 121. <https://doi.org/10.4274/tjod.galenos.2025.83278>
- [11] Tang, J., Han, J., Xue, J., Zhen, L., et al. (2023). A deep-learning-based method can detect both common and rare genetic disorders in fetal ultrasound. *Biomedicines*, 11(6): 1756. <https://doi.org/10.3390/biomedicines11061756>
- [12] Zhang, H.G., Jiang, Y.T., Dai, S.D., Li, L., Hu, X.N., Liu, R.Z. (2021). Application of intelligent algorithms in Down syndrome screening during second trimester pregnancy. *World Journal of Clinical Cases*, 9(18): 4573. <https://doi.org/10.12998/wjcc.v9.i18.4573>
- [13] Koul, A.M., Ahmad, F., Bhat, A., Aein, Q.U., Ahmad, A., Reshi, A.A., Kaul, R.U.R. (2023). Unraveling Down syndrome: From genetic anomaly to artificial intelligence-enhanced diagnosis. *Biomedicines*, 11(12): 3284. <https://doi.org/10.3390/biomedicines11123284>
- [14] Leghari, I.M., Ujir, H., Hipiny, I. (2023). Machine learning techniques to enhance the mental age of Down syndrome individuals: A detailed review. *International Journal of Advanced Computer Science and Applications*, 14(1): 990-999. <https://doi.org/10.14569/ijacsa.2023.01401107>
- [15] Fiorentino, M.C., Villani, F.P., Di Cosmo, M., Frontoni, E., Moccia, S. (2023). A review on deep-learning algorithms for fetal ultrasound-image analysis. *Medical Image Analysis*, 83: 102629. <https://doi.org/10.1016/j.media.2022.102629>
- [16] Kilercik, M., Yozgat, I., Serdar, M.A., Aksungar, F., et al. (2022). What are the predominant parameters for Down syndrome risk estimation in first-trimester screening: A data mining study. *Turkish Journal of Biochemistry*, 47(6): 704-709. <https://doi.org/10.1515/tjb-2022-0004>
- [17] Do, H.D., Allison, J.J., Nguyen, H.L., Phung, H.N., Tran, C.D., Le, G.M., Nguyen, T.T. (2024). Applying machine learning in screening for Down syndrome in both trimesters for diverse healthcare scenarios. *Heliyon*, 10(15): e34476. <https://doi.org/10.1016/j.heliyon.2024.e34476>
- [18] Shirinzadeh, A., Ershadi, R., Amouei, A., Jafari, J., Soltani, H., Kargar, S., Binesh, F. (2022). Severe main bronchus obstruction due to pulmonary schwannoma: A case report. *Iranian Journal of Pediatric Hematology & Oncology*, 12(2): 140-144. <https://doi.org/10.18502/ijpho.v12i2.9079>
- [19] Chakraborty, C., Bhattacharya, M., Pal, S., Lee, S.S. (2024). From machine learning to deep learning: Advances of the recent data-driven paradigm shift in medicine and healthcare. *Current Research in Biotechnology*, 7: 100164. <https://doi.org/10.1016/j.crbiot.2023.100164>
- [20] Vakili-Ojarood, M., Naseri, A., Shirinzadeh-Dastgiri, A., Saberi, A., et al. (2024). Ethical considerations and equipoise in cancer surgery. *Indian Journal of Surgical Oncology*, 15(Suppl 3): 363-373. <https://doi.org/10.1007/s13193-024-02023-8>
- [21] Masoudi, A., Omidi, A., Shams, S.E., Yeganegi, M., et al. (2024). Applications of artificial intelligence chatbots in congenital diplopodia: A real-world perspective on a rare case of duplicated lower limb. *World Journal of Peri & Neonatology*, 7(1): 7-15. <https://doi.org/10.18502/wjpn.v7i1.17323>
- [22] Alonso, E., Beristain, A., Burgos, J., Gurrutxaga, I. (2025). Comparison of machine learning algorithms to predict Down syndrome during the screening of the first trimester of pregnancy. *Applied Sciences*, 15(10): 5401. <https://doi.org/10.3390/app15105401>
- [23] Mai, C.T., Isenburg, J.L., Canfield, M.A., Meyer, R.E., et al. (2019). National population-based estimates for major birth defects, 2010-2014. *Birth Defects Research*, 111(18): 1420-1435. <https://doi.org/10.1002/bdr2.1589>
- [24] Huang, T., Gibbons, C., Rashid, S., Priston, M.K., Bedford, H.M., Mak-Tam, E., Meschino, W.S. (2020). Prenatal screening for trisomy 21: A comparative performance and cost analysis of different screening strategies. *BMC Pregnancy and Childbirth*, 20(1): 713. <https://doi.org/10.1186/s12884-020-03394-w>
- [25] Guibourdenche, J., Leguy, M.C., Pidoux, G., Hebert-Schuster, M., et al. (2023). Biochemical screening for fetal trisomy 21: Pathophysiology of maternal serum markers and involvement of the placenta. *International Journal of Molecular Sciences*, 24(8): 7669.

- <https://doi.org/10.3390/ijms24087669>
- [26] Coppedè, F. (2016). Risk factors for Down syndrome. *Archives of Toxicology*, 90(12): 2917-2929. <https://doi.org/10.1007/s00204-016-1843-3>
- [27] Pitukkijronnakorn, S., Promsonthi, P., Panburana, P., Udomsubpayakul, U., Chittacharoen, A. (2011). Fetal loss associated with second trimester amniocentesis. *Archives of Gynecology and Obstetrics*, 284(4): 793-797. <https://doi.org/10.1007/s00404-010-1727-3>
- [28] Peng, L., Jian, S., Li, M., Kan, Z., Qiao, L., Li, D. (2025). A unified multimodal classification framework based on deep metric learning. *Neural Networks*, 181: 106747. <https://doi.org/10.1016/j.neunet.2024.106747>
- [29] Hannah, E., Raamesh, L., Sumathi. (2019). Early prenatal diagnosis of Down's syndrome-A machine learning approach. In *Soft Computing for Problem Solving: SocProS 2018*, pp. 467-477. [https://doi.org/10.1007/978-981-15-0035-0\\_37](https://doi.org/10.1007/978-981-15-0035-0_37)
- [30] Hart, J. M., O'Brien, B.M. (2021). First-trimester or second-trimester screening, or both, for Down syndrome: The FASTER trial. In *50 Studies Every Obstetrician-Gynecologist Should Know*, pp. 108-114. <https://doi.org/10.1093/med/9780190947088.003.0020>
- [31] Ramanathan, S., Sangeetha, M., Talwai, S., Natarajan, S. (2018). Probabilistic determination of Down's syndrome using machine learning techniques. In *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Bangalore, India, pp. 126-132. <https://doi.org/10.1109/icacci.2018.8554392>
- [32] Koivu, A., Korpimäki, T., Kivelä, P., Pahikkala, T., Sairanen, M. (2018). Evaluation of machine learning algorithms for improved risk assessment for Down's syndrome. *Computers in Biology and Medicine*, 98: 1-7. <https://doi.org/10.1016/j.combiomed.2018.05.004>
- [33] Jamshidnezhad, A., Hosseini, S.M., Mohammadi-Asl, J., Mahmudi, M. (2021). An intelligent prenatal screening system for the prediction of Trisomy-21. *Informatics in Medicine Unlocked*, 24: 100625. <https://doi.org/10.1016/j.imu.2021.100625>
- [34] Tang, J., Han, J., Jiang, Y., Xue, J., et al. (2023). An innovative three-stage model for prenatal genetic disorder detection based on region-of-interest in fetal ultrasound. *Bioengineering*, 10(7): 873. <https://doi.org/10.3390/bioengineering10070873>
- [35] Catic, A., Gurbeta, L., Kurtovic-Kozaric, A., Mehmedbasic, S., Badnjevic, A. (2018). Application of neural networks for classification of Patau, Edwards, Down, Turner and Klinefelter Syndrome based on first trimester maternal serum screening data, ultrasonographic findings and patient demographics. *BMC Medical Genomics*, 11(1): 19. <https://doi.org/10.1186/s12920-018-0333-2>
- [36] Badenas, C., Rodríguez-Revenga, L., Morales, C., Mediano, C., et al. (2010). Assessment of QF-PCR as the first approach in prenatal diagnosis. *The Journal of Molecular Diagnostics*, 12(6): 828-834. <https://doi.org/10.2353/jmoldx.2010.090224>
- [37] Singh, R., Gupta, S., Mohamed, H.G., Bharany, S., Rehman, A.U., Ghadi, Y.Y., Hussien, S. (2025). Advancing prenatal healthcare by explainable AI enhanced fetal ultrasound image segmentation using U-Net++ with attention mechanisms. *Scientific Reports*, 15(1): 19612. <https://doi.org/10.1038/s41598-025-04631-y>
- [38] Asad, A., Sarwar, M., Aslam, M., Akpokodje, E., Jilani, S.F. (2025). MultiScaleFusion-Net and ResRNN-Net: Proposed deep learning architectures for accurate and interpretable pregnancy risk prediction. *Applied Sciences*, 15(11): 6152. <https://doi.org/10.3390/app15116152>
- [39] Bragin, E., Chatzimichali, E.A., Wright, C.F., Hurles, M.E., Firth, H.V., Bevan, A.P., Swaminathan, G.J. (2014). DECIPHER: Database for the interpretation of phenotype-linked plausibly pathogenic sequence and copy-number variation. *Nucleic Acids Research*, 42(D1): D993-D1000. <https://doi.org/10.1093/nar/gkt937>