International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(18s) 2026 pp. 155-175
https://doi.org/10.70917/ijcisim-2026-4843
Article

A Multi-Modal Deep Learning Framework for Chromosomal Abnormality Diagnosis Using Neuro-Inference Predictive Network

B. L. Shivakumar[1], U. Priya[2]

[1]Department of Computer Science, Sri Ramakrishna College of Arts and Science, Coimbatore – 641006, Tamil Nadu, India.
Email: blshiva@gmail.com
[2]Department of Computer Science, Sri Ramakrishna College of Arts and Science, Coimbatore – 641006, Tamil Nadu, India.
Email: priyatamilelango@gmail.com

Abstract: Trisomy 21 (Down syndrome) is one of the most prevalent chromosomal abnormalities, requiring accurate and early diagnosis for effective clinical intervention. Conventional diagnostic approaches often rely on either chromosome image analysis or facial phenotype assessment independently, limiting their ability to exploit complementary genomic and phenotypic information. Existing multimodal methods also suffer from inadequate cross-domain feature alignment, inefficient heterogeneous feature fusion, limited interpretability, and suboptimal optimization, resulting in reduced prediction accuracy and poor generalization. To address these limitations, this study proposed a Fusion-Based Prediction and Optimization comprising the Neuro-Inference Predictive Network (NeIPN) and the Meta-Gradient Fusion Optimizer (MeGFO). NeIPN integrated chromosome and facial classification outputs using adaptive correlation mapping, probabilistic attention, cross-domain feature fusion, and hierarchical inference to generate a unified latent representation for Trisomy 21 prediction. Subsequently, MeGFO refined the prediction through meta-learning-based gradient adaptation, dynamic fusion weight optimization, and learning parameter fine-tuning to minimize prediction error and improve model generalization. Experimental evaluation demonstrated that the proposed framework achieved an accuracy of 97.21 with an average loss of 0.2237, outperforming conventional multimodal prediction approaches. The proposed fusion and optimization strategy provides a reliable, interpretable, and scalable decision-support framework for accurate automated Trisomy 21 detection and has strong potential for future clinical implementation.
Keywords: Chromosome Image Analysis, Facial Phenotype Analysis , Feature Fusion , Genetic Disorder Diagnosis, Multimodal Diagnosis , Predictive Modeling

1. Introduction

Trisomy 21 (Down syndrome) is a common chromosomal disorder caused by an extra copy of chromosome 21, resulting in characteristic genetic and facial abnormalities. Integrating chromosome image analysis with facial phenotype recognition through multimodal Deep Learning (DL) provides an effective strategy for accurate, automated, and early diagnosis while reducing dependence on manual clinical interpretation. Existing diagnostic approaches generally analyze chromosome images or facial phenotypes independently, limiting complementary information utilization. These often experience reduced robustness under image noise, inadequate feature representation, weak multimodal integration, and limited adaptive optimization, leading to suboptimal prediction accuracy and reduced clinical interpretability.

A DL framework was introduced for automated nuchal translucency measurement by combining ultrasound image quality assessment with chromosomal abnormality screening. The framework improves image enhancement and measurement consistency, supporting reliable prenatal diagnosis through automated imaging analysis while reducing operator dependency [1]. An automated cytogenetic pipeline combined Pix2Pix image enhancement with

DL detection for bone marrow karyogram analysis. The framework improved chromosome visualization, segmentation quality, and abnormal chromosome detection, demonstrating efficient preprocessing and chromosome identification in complex cytogenetic images. [2]. Artificial intelligence-based MRI segmentation was developed for precise tumor boundary identification and prognostic assessment. The research demonstrated robust segmentation performance using DL, emphasizing accurate structural extraction and automated image interpretation for clinical diagnostic applications. [3]. A hybrid TabNet–CNN framework was proposed for high-dimensional genomic data classification. The model effectively captured complex genomic relationships by combining tabular learning with deep convolutional representations, significantly improving classification accuracy and feature learning capability across large-scale genetic datasets. [4]. A hybrid tuned DL model utilized optimized feature extraction and genomic data learning for breast cancer diagnosis. The framework enhanced diagnostic accuracy through adaptive learning strategies and efficient genomic feature representation, demonstrating the effectiveness of hybrid DL for genetic disease prediction. [5].

An improved Differentiable Architecture Search (DARTS) framework was proposed for chromosome image classification by automatically optimizing neural network architectures. The model enhanced chromosome feature extraction, reduced computational complexity, and achieved higher classification accuracy, demonstrating the effectiveness of neural architecture search for cytogenetic image analysis. [6]. An AI-assisted framework combined clinical information, neuroimaging, facial profiling, and genotype–phenotype correlation analysis for DDX3X syndrome assessment. The research demonstrated that integrating facial phenotype analysis with genetic characteristics improved syndrome identification and supported comprehensive precision diagnosis. [7]. A triplet-based singular geometric autoencoder was developed to construct a three-dimensional facial phenotype space for genetic syndromes. The framework learned discriminative geometric facial representations, improving syndrome characterization and facilitating accurate recognition of clinically significant facial morphological variations. [8]. An interpretable DL model was introduced for automated Down syndrome detection using facial images. The framework combined explainable feature learning with classification mechanisms, improving diagnostic transparency while accurately identifying facial characteristics associated with Down syndrome for reliable clinical decision support. [9]. A transfer learning-based framework was proposed for multiclass facial syndrome classification using pretrained convolutional neural networks. The approach enhanced facial feature extraction, reduced training complexity, and achieved improved syndrome recognition performance across multiple genetic disorders through effective knowledge transfer. [10].

A transfer learning framework combined VGG16-based deep feature extraction with Non-Negative Matrix Factorization and Light Gradient Boosting Machine for Down syndrome diagnosis using facial images. The model effectively captured discriminative facial characteristics and achieved improved diagnostic accuracy through optimized feature representation and classification. [11]. A cross-attention-based multimodal fusion framework was proposed for integrating heterogeneous medical features using adaptive attention mechanisms. The model effectively learned complementary cross-modal representations, enhancing classification performance and demonstrating the effectiveness of attention-guided multimodal fusion for complex medical diagnosis. [12]. A multimodal DL diagnostic decision support system integrated multiple clinical data sources for disease prediction. The framework used comprehensive feature fusion and adaptive DL to improve diagnostic reliability, demonstrating the advantages of multimodal intelligence for clinical decision-making. [13]. A multi-view multimodal DL framework integrated heterogeneous imaging modalities using deep feature fusion and joint representation learning. The model effectively captured complementary information from multiple sources, improving diagnostic accuracy and robustness through cross-modal feature integration. [14]. A multimodal attention-guided network combined convolutional learning with adaptive attention mechanisms to strengthen multimodal feature representation. The framework enhanced discriminative feature extraction and classification performance by emphasizing informative regions while suppressing irrelevant information during the learning process. [15]

To overcome existing limitations, this research proposes a multimodal DL framework that integrates chromosome image analysis and facial phenotype analysis using adaptive feature fusion and meta-learning

optimization. The proposed framework combines entropy-guided denoising, chromosome segmentation, facial feature extraction, cross-modal attention fusion, and adaptive optimization to achieve accurate, robust, and interpretable Trisomy 21 detection.

- To develop an automated multimodal framework that jointly analyzes chromosome images and facial phenotypes for accurate and early Trisomy 21 detection.
- To improve chromosome and facial feature representation using adaptive denoising, segmentation, transfer learning, and attention-guided deep feature extraction.
- To introduce an adaptive cross-modal fusion mechanism that effectively combines genomic and phenotypic information for robust diagnostic prediction.
- To enhance prediction accuracy and generalization through a meta-learning optimization strategy that dynamically refines fusion weights and learning parameters while improving model interpretability and reliability.

The remainder of this research is organized as follows. Section 2 presents the background study and identifies the research gap. Section 3 describes the proposed NeIPN–MeGFO methodology, including the dataset description, prediction and optimization framework, mathematical formulations, workflow, and algorithm. Section 4 presents the experimental results and performance evaluation, Section 5 discusses the experimental findings and comparative analysis, and Section 6 concludes the research with future research directions.

2. Background Study

J. Tang, X. Yin, J. Lai, K. Luo, and D. Wu (2025) [16] proposed a multimodal osteoporosis prediction model integrating X-ray images and clinical data. Feature fusion improved prediction accuracy. However, disease-specific adaptation remained limited, indicating a need for generalized multimodal diagnostic frameworks.

Z. Liu et al. (2025) [17] developed interpretable machine learning models using multidimensional fusion data for surgical margin prediction. Explainable learning enhanced clinical reliability. However, genomic and facial information were not integrated, leaving multimodal genetic diagnosis insufficiently explored.

Z. Yang, X. Wang, H. Lin, M. Li, and M. Lin (2025) [18] introduced a cross-modal feature interaction network for heterogeneous change detection. Cross-modal interactions improved representation learning. Nevertheless, medical phenotype analysis and adaptive disease prediction remained unexplored despite promising fusion performance.

H. Deng, X. Ding, S. Zhu, X. Song, and Y. Guo (2025) [19] presented a multimodal DL framework for chronic obstructive pulmonary disease prediction. Multisource feature integration improved diagnostic performance. However, chromosome image analysis and facial phenotype learning were not considered.

J. Wu, S. Tanim, M. Woo, et al. (2025) [20] proposed transformer neural networks with multi-source data fusion for pandemic prediction. Integrated temporal information improved forecasting accuracy. Nevertheless, adaptive multimodal medical image fusion and genetic disorder diagnosis remained unexplored research directions.

L. Zhang, T. Li, H. Cui, et al. (2025) [21] developed a graph neural network for multimodal medical data prediction. Graph-based feature aggregation improved predictive capability. However, chromosome imaging, facial phenotypes, and meta-learning optimization were absent, creating opportunities for multimodal genetic diagnosis.

J. Wang, S. Wen, W. Liu, et al. (2024) [22] proposed deep joint learning with multimodal feature fusion for Alzheimer's diagnosis. Joint representation learning improved diagnostic accuracy. However, adaptive attention mechanisms and chromosome–facial multimodal integration were not investigated for genetic disorders.

S. R. Islam, T. Xia, W. He, et al. (2026) [23] introduced vision transformer autoencoders for learning local and non-local brain imaging features. Genetic association discovery improved significantly. However, chromosome image classification and facial phenotype fusion remained unexplored for Trisomy 21 detection.

Table 1: Comparative Analysis of Recent Attention, Multimodal Fusion, and Optimization Methods

Author (Year) [Ref.]	Concept	Method Used	Limitations	Results
J. Liao & L. Wang (2025) [24]	Hybrid attention-based hyperspectral image classification	ATN-Hybrid with deterministic–probabilistic attention mechanism	Limited to hyperspectral imaging; not validated for medical multimodal diagnosis	Improved classification accuracy and feature discrimination.
K. Sun et al. (2024) [25]	Cross-modal brain tumor segmentation	CMAF-Net with cross-modal attention fusion	Focused only on brain tumor segmentation; chromosome and facial modalities were excluded	Achieved accurate multimodal segmentation with enhanced feature fusion.
Y. Liu & J. Tian (2024) [26]	Probabilistic attention learning for CNNs	Probabilistic Attention Map (PAM) integrated with CNN	Evaluated on general vision tasks; lacked multimodal medical validation	Improved attention representation and CNN classification performance.
R. Mittal & M. Bhushan (2026) [27]	Prenatal defect detection using hybrid DL	Hybrid DL model for fetal heart diagnosis	Restricted to prenatal cardiac imaging; chromosome image analysis was absent	Enhanced prenatal diagnostic accuracy and reliability.
H. Mohd Saleh, N. A. M. Isa, N. H. Saad & M. I. F. Maruzuki (2026) [28]	CNN optimization through adaptive learning rate	Enhanced Cyclical Learning Rate (ECLR) strategy	Optimized training only; did not address multimodal feature fusion	Accelerated convergence and improved CNN learning stability.
A. B. Sawad & M. Binsawad (2026) [29]	MRI-based glioma prediction	Radiomics-driven hybrid DL framework	Disease-specific implementation without genomic or facial information	Improved glioma grading and molecular subtype prediction accuracy.
S. Gull & J. Kim (2025) [30]	Few-shot medical image classification	Metric-based meta-learning framework	Applied only to MRI classification; multimodal disease diagnosis was not investigated	Improved few-shot classification performance and model generalization.

Table 1 compares recent attention, multimodal fusion, optimization, and meta-learning approaches. Although these research achieved strong performance within their respective domains, none integrated chromosome image analysis with facial phenotype learning through adaptive multimodal fusion and meta-learning for accurate Trisomy 21 detection.

2.1 Research Gap

Existing research have independently addressed chromosome image classification, facial phenotype analysis, multimodal fusion, attention mechanisms, and optimization techniques. However, no unified framework has simultaneously integrated chromosome image analysis with facial phenotype recognition for Trisomy 21 detection. Furthermore, adaptive cross-modal feature fusion, entropy-guided preprocessing, interpretable attention learning, and meta-learning-based optimization remain insufficiently explored, limiting diagnostic accuracy, robustness, and generalization.

3. Proposed Methodology:

This section introduces the proposed NeIPN and MeGFO methodology for multimodal Down syndrome prediction, which involves the description of the dataset, the architecture of the model, the mathematical formulation, and optimization process. The equations and pseudocode presented represent the feature fusion, predictive inference, meta-gradient optimization and final decision refinement mechanisms of the proposed approach.

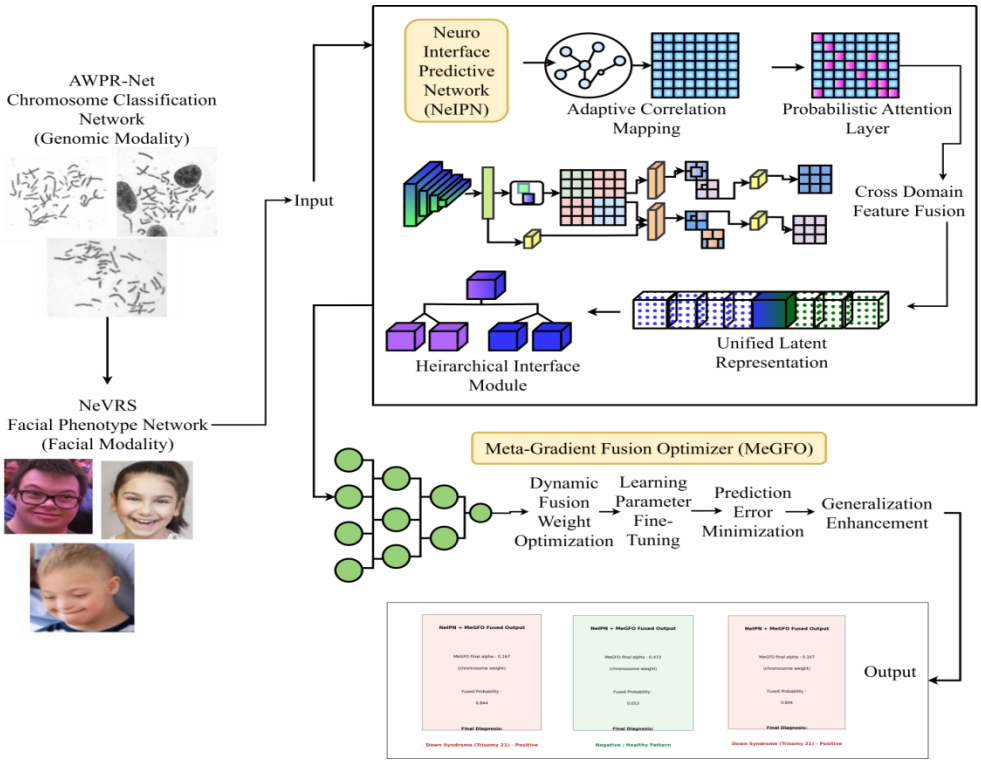


Figure 1: Proposed NeIPN- MeGFO Fusion-Based Architecture for Trisomy 21 Detection

Figure 1 shows the proposed architecture integrates chromosome and facial classification outputs using NeIPN for adaptive feature fusion and hierarchical inference. MeGFO optimizes fusion weights through meta-learning, improving prediction accuracy, robustness, and generalization to produce reliable Trisomy 21 diagnostic decisions with confidence scores.

3.1 Dataset Description:

Chromosome Image Dataset (Karyotype) was acquired from the Kaggle website: **https://www.kaggle.com/datasets/aliabedimadiseh/chromosome-image-dataset-karyotype**. It provides a set of images of human chromosomes at the metaphase stage in high resolution, arranged as per the standard convention of human karyotype. The dataset contains individual chromosome images both normal and abnormal that can be used to classify chromosomes and detect genetic defects like Down syndrome (Trisomy 21). These chromosome images provide a variety of genomic attributes to train and test the proposed automated Down syndrome prediction DL framework.

3.2 Prediction and Optimization Methodology

Prediction and optimization methodology integrates the information from the chromosomal and facial classification networks to enhance the prediction of Down syndrome. The Neuro-Inference Predictive Network (NeIPN) is used to fuse features into the multimodal inference and predict, while the Meta-Gradient Fusion Optimizer (MeGFO) is adaptively refined with meta-gradient learning. These models work collaboratively to improve the robustness of the prediction, generalization and reliability of the decision, effectively making use of complementary genetic and facial information.

3.2.1 Neuro-Inference Predictive Network (NeIPN)

Neuro-Inference Predictive Network (NeIPN) is a network that combines the outputs of the two classifiers the chromosomal and the facial to achieve accurate prediction of Down syndrome by incorporating heterogeneous biological information and phenotypic information. The main objective of NeIPN is to acquire complex relationship between the genetic abnormalities and facial features by adaptive feature fusion and intelligent inference. The process consists of feature alignment, adaptive correlation mapping, probabilistic attention-based fusion, hierarchical representation learning and the final diagnostic probability estimation. NeIPN is more reliable at prediction due to the use of multiple modalities other than relying on one data source. The model is scalable as it enables the addition of other clinical modalities like medical images, genomic markers, patient records, etc. without the need to redesign the entire architecture. It can be reproduced by standardizing feature extraction, incorporating adaptive learning processes and maintaining uniform fusion approaches. NeIPN addresses the drawbacks of existing methods such as single modality analysis, inadequate feature correlation, and limited population generalization. Therefore, NeIPN allows for the development of an AI-based system for the prediction of Down syndrome that is robust, scalable and interpretable.

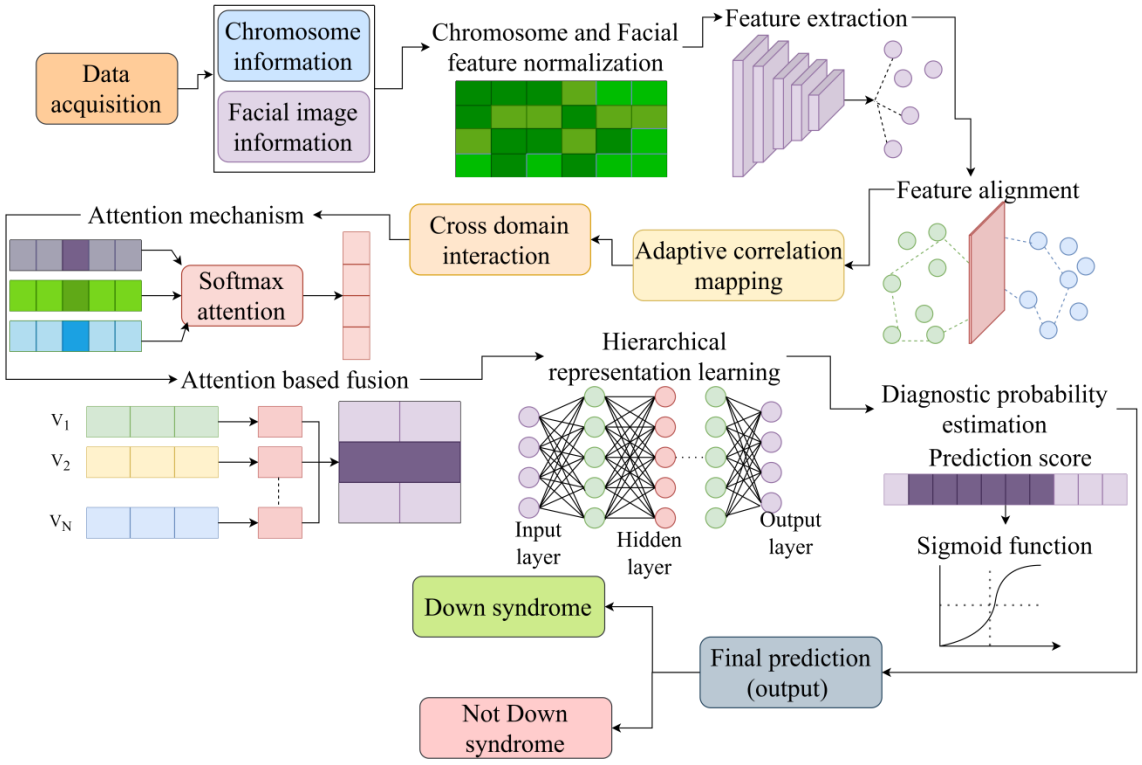


Figure 2: Neuro-Inference Predictive Network workflow

The figure 2 shows the NeIPN workflow, which involves capturing chromosome data and face image data, followed by their normalization to create normalized chromosome representations and face representations. These features are adapted via a correlation mapping, cross-domain interaction, and Softmax attention, which highlight the

most informative multimodal relationships. In the final, the fused features pass through neural layers to make the final prediction, whether Down syndrome or non-Down syndrome.

$$F_c = \phi_c(X_c; \theta_c) \quad (1)$$

The equation (1) describes the extracted chromosomal feature representation $\mathrm{F_c}$ given the input chromosomal information $\mathrm{X_c}$, the learnable parameters of the network $\mathrm{\theta_c}$. This equation is used to convert the raw genetic information into discriminative feature embeddings for the prediction of Down syndrome.

$$F_f = \phi_f(X_f; \theta_f) \quad (2)$$

This equation (2) $\mathrm{F_f}$ is the facial feature embedding, $\mathrm{X_f}$ is the input facial image information, $\mathrm{\phi_f}$ is the facial feature extraction function and $\mathrm{\theta_f}$ is the network parameters. This equation allows extracting important phenotypic features from facial images to be used in multimodal prediction.

$$\widehat{F_i} = \frac{F_i - \mu(F_i)}{\sigma(F_i) + \epsilon} \quad (3)$$

The equation (3) is used to normalize the feature vector $\widehat{F_i}$ and F_i is the extracted feature, $\mu(F_i)$ is the mean value, $\sigma(F_i)$ is the standard deviation and ϵ is a small constant to ensure numerical stability. This equation aims to normalize the heterogeneous features for consistent multimodal learning.

$$A_i = W_a F_i + b_a \quad (4)$$

The equation (4) is used for this $\mathrm{A_i}$ is the aligned feature representation, $\mathrm{W_a}$ is the adaptive transformation weight matrix, $\mathrm{F_i}$ is the input feature vector and $\mathrm{b_a}$ is the bias term. This equation transforms heterogeneous chromosomal and facial features to a common representation space.

$$C(F_c, F_f) = \frac{F_c^T F_f}{||F_c||_2 ||F_f||_2} \quad (5)$$

This equation (5), $C(F_c, F_f)$ represents the correlation relationship between chromosomal and facial features, F_c indicates chromosomal feature vectors, F_f denotes facial feature vectors, and $||.||_2$ represents the Euclidean norm. This equation measures the similarity between genetic and phenotypic characteristics for effective cross-domain feature fusion.

$$H_{cf} = F_c \odot F_f + C(F_c, F_f) \quad (6)$$

This equation (6), H_{cf} represents the combined cross-domain feature representation, F_c and F_f indicates chromosomal and facial features, $\odot$ denotes element-wise interaction, and $C(F_c, F_f)$ represents feature correlation. This equation captures complementary relationships between biological and facial information.

$$\alpha_i = \frac{e^{q_i k_i^T / \sqrt{d}}}{\sum j=1Ne^{q_j k_j^T / \sqrt{d}}} \quad (7)$$

This equation (7), α_i represents the attention coefficient, q_i and k_i denotes query and key feature vectors, d represents feature dimensionality, and N indicates the number of feature elements. This equation assigns adaptive importance to informative multimodal features.

$$F_{fusion} = \sum i = 1N\alpha_i V_i \quad (8)$$

This equation (8), F_{fusion} represents the fused multimodal feature representation, α_i denotes attention weights, and V_i represents the feature value vectors. This equation combines weighted genetic and facial information to generate a unified representation.

$$H_l = \sigma(W_l H_{l-1} + b_l) \quad (9)$$

This equation (9), H_l represents the learned representation at layer l, W_l denotes the layer weight matrix, H_{l-1} indicates the previous layer representation, b_l represents bias, and σ denotes the activation function. This equation learns high-level diagnostic patterns through hierarchical feature transformation.

$$Z_{NeIPN} = f\left(H_l, F_{fusion}; \theta_N\right) \quad (10)$$

This equation (10), Z_{NeIPN} represents the predictive feature representation generated by NeIPN, H_l denotes hierarchical features, F_{fusion} represents fused features, and θ_N indicates NeIPN parameters. This equation generates a comprehensive representation for Down syndrome prediction.

$$P(y \mid X) = \frac{1}{1+e^{-Z_{NeIPN}}} \quad (11)$$

This equation (11), $P(y \mid X)$ represents the probability of Down syndrome occurrence, Z_{NeIPN} denotes the prediction score, and e represents the exponential function. This equation converts the learned representation into a diagnostic probability value.

$$W_m = \frac{e^{sm}}{\sum m=1 M e^{sm}} \quad (12)$$

This equation (12), W_m represents the adaptive contribution weight of each modality, s_m denotes the modality importance score, and M represents the number of modalities. This equation dynamically balances chromosomal and facial feature contributions.

3.2.2 Meta-Gradient Fusion Optimizer (MeGFO)

The Meta-Gradient Fusion Optimizer (MeGFO) is designed to enhance the performance of multimodal Down syndrome prediction by dynamically optimizing the fusion parameters obtained from chromosomal and facial classification networks. The primary purpose of MeGFO is to minimize prediction errors and improve model generalization through meta-learning-based adaptive gradient updates. Its workflow involves analyzing prediction feedback, adjusting fusion weights, updating learning parameters, and selecting the optimal feature contribution from each modality. MeGFO enables adaptive decision refinement by continuously learning from previous optimization states and improving the stability of final diagnostic outcomes. The model provides scalability by supporting additional input modalities, larger datasets, and diverse clinical environments without significant architectural modifications. Its reproducibility is achieved through systematic gradient adaptation, consistent optimization strategies, and parameter-driven learning mechanisms. MeGFO overcomes limitations of conventional fusion approaches, such as fixed weighting, poor adaptability, overfitting, and reduced robustness across different patient populations. Thus, MeGFO provides an efficient optimization framework for reliable, accurate, and interpretable Down syndrome prediction.

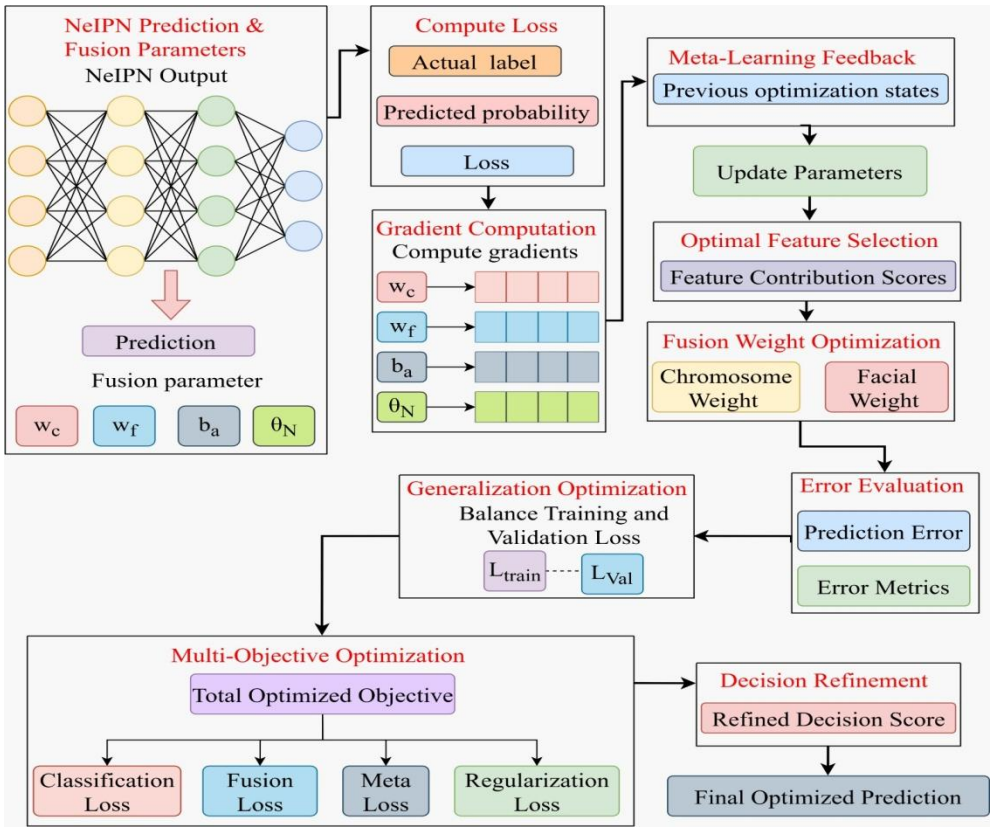


Figure 3: Meta-Gradient Fusion Optimizer (MeGFO) workflow

Figure 3 illustrates the MeGFO architecture optimizes prediction performance by computing the prediction loss, updating fusion parameters using meta-gradient learning, and selecting the most informative features. It further performs fusion weight optimization, error evaluation, generalization control, and multi-objective optimization to refine the decision-making process. Finally, the optimized parameters generate an accurate prediction with improved robustness and confidence.

$$L(\theta) = -\sum i = 1N y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i) \quad (13)$$

This equation (13), $L(\theta)$ represents the classification loss, y_i denotes the actual class label, $\hat{y}_i$ represents predicted probability, and N indicates the number of samples. This equation measures prediction error during model training.

$$\nabla_\theta L(\theta) = \frac{\partial L}{\partial \theta} \quad (14)$$

This equation (14), $\nabla_\theta L(\theta)$ represents the gradient of the loss function, $L(\theta)$ denotes prediction loss, and θ indicates model parameters. This equation calculates parameter adjustment direction for optimization.

$$\theta' = \theta - \eta \nabla_\theta L(\theta) \quad (15)$$

This equation (15), θ' represents updated model parameters, θ denotes current parameters, η represents learning rate, and $\nabla_\theta L(\theta)$ indicates gradient information. This equation enables adaptive parameter refinement through meta-learning.

$$W^{t+1} = W^t - \eta_m \nabla_W L(W^t) \quad (16)$$

This equation (16), W^{t+1} represents updated fusion weights, W^t indicates current weights, η_m denotes meta-learning rate, and $\nabla_W L(W^t)$ represents fusion weight gradients. This equation optimizes modality contribution dynamically.

$$E_p = \frac{1}{N}\sum i = 1N(y_i - \hat{y}_i)^2 \quad (17)$$

This equation (17), E_p represents prediction error, y_i denotes actual output, $\hat{y}_i$ represents predicted output, and N indicates the number of samples. This equation evaluates the deviation between predicted and actual results.

$$\theta^* = \operatorname{argmin}_{\theta}(L_{train} + \lambda L_{val}) \quad (18)$$

This equation (18), θ^* represents optimized parameters, L_{train} denotes training loss, L_{val} indicates validation loss, and λ represents regularization coefficient. This equation improves model generalization and reduces overfitting.

$$Y_{final} = f(Z_{NeIPN}, W_{MeGFO}) \quad (19)$$

This equation (19), Y_{final} represents the final Down syndrome prediction output, Z_{NeIPN} denotes the fused predictive representation, and W_{MeGFO} indicates optimized fusion parameters. This equation generates the final optimized decision.

$$L_{total} = L_{cls} + \lambda_1 L_{fusion} + \lambda_2 L_{opt} \quad (20)$$

This equation (20), L_{total} represents the overall objective function, L_{cls} denotes classification loss, L_{fusion} represents feature fusion loss, L_{opt} indicates optimization loss, and λ_1, λ_2 represent balancing coefficients. This equation jointly optimizes prediction accuracy, fusion quality, and model stability.

Algorithm: NeIPN and MeGFO

Input:

Chromosomal data Xc

Facial image data Xf

Training labels Y

NeIPN parameters θc, θf, θN

Initial fusion weights W

Learning rates η and ηm

Begin

Initialize NeIPN and MeGFO parameters:

Initialize chromosomal feature extractor θc

Initialize facial feature extractor θf

Initialize fusion network parameters θN

Initialize adaptive fusion weights W

Feature Acquisition and Preprocessing:

Collect chromosomal information Xc and facial information Xf

Normalize and preprocess multimodal inputs

Multimodal Feature Learning:

Fc ← Chromosomal_Feature_Extraction(Xc, θc)

Ff ← Facial_Feature_Extraction(Xf, θf)

Feature Alignment and Correlation Learning:

Ac ← Feature_Alignment(Fc)

Af ← Feature_Alignment(Ff)

Ccf ← Adaptive_Correlation_Mapping(Ac, Af)

NeIPN-Based Multimodal Fusion:

Hcf ← Cross_Domain_Interaction(Ac, Af, Ccf)

α ← Probabilistic_Attention(Hcf)

Ffusion ← Attention_Fusion(Hcf, α)

Predictive Representation Learning:

Z ← Hierarchical_Representation(Ffusion, θN)

P ← Prediction_Probability(Z)

MeGFO-Based Optimization:

Calculate prediction loss:

L ← Loss(Y, P)

Compute meta-gradient:

$\nabla\theta L \leftarrow \partial L/\partial\theta$

Update network parameters:

$\theta^* \leftarrow \theta - \eta\nabla\theta L$

Optimize fusion weights:

$W^* \leftarrow W - \eta m\nabla W L(W)$

Final Prediction Refinement:

Yfinal ← Decision_Fusion(Z, W*)

Update model parameters using optimized feedback

Return Yfinal, θ*, W*

End

Output:

Optimized Down syndrome prediction Yfinal

Updated NeIPN parameters θ*

Optimized fusion weights W*

The proposed NeIPN and MeGFO algorithm first performs multimodal feature extraction, alignment, and attention-based fusion of chromosomal and facial information to generate predictive representations. The MeGFO

module then optimizes fusion weights and network parameters through meta-gradient learning to refine the final Down syndrome prediction with improved accuracy and generalization.

4. Experiment Setup

This section evaluates the proposed Phase-3 framework using 70% training, 15% validation, and 15% testing data. Experiments were conducted in Python (PyTorch) on an Intel Core i9 processor, 32 GB RAM, and NVIDIA RTX 4080 (16 GB)**.** Performance was assessed using Accuracy, Precision, Recall, F1-score, Average Loss, ablation, and statistical analyses**.**

4.1 Sample Multi-Modal Input Images for Chromosomal Abnormality Analysis

This section presents representative chromosomal and facial image samples used as multi-modal inputs for the proposed chromosomal abnormality diagnosis framework.

Table 2: Representative Multi-Modal Input Images for Chromosomal Diagnosis

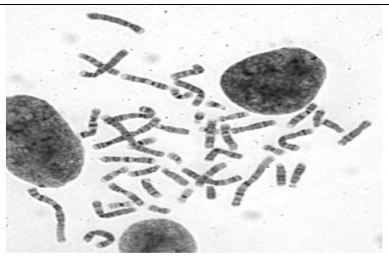
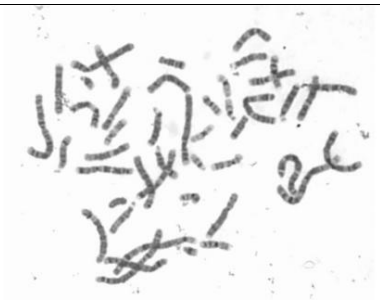
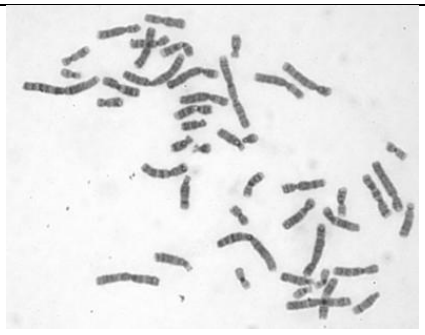
	Phase 1	**Phase 2**
Image 1		
Image 2		
Image 3		

Table 2 shows sample chromosomal and corresponding facial images used for chromosomal abnormality diagnosis. Phase 1 contains chromosome images, while Phase 2 contains facial images of the same cases. Both image types are analyzed by the proposed framework to extract complementary genomic and phenotypic features. The extracted features are fused to improve the accuracy and reliability of abnormality prediction. This multi-modal approach enables more effective and robust diagnosis than using a single data source alone.

4.2 Individual Domain Prediction and Probabilistic Attention Analysis

This section presents the prediction probabilities of the chromosomal and facial classification networks along with the probabilistic attention weights generated by the proposed NeIPN.

Table 3: Individual Domain Probabilities and NeIPN Probabilistic Attention Results

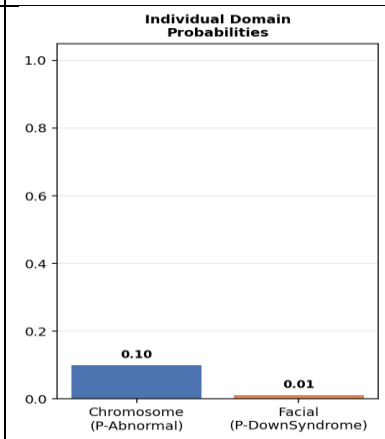
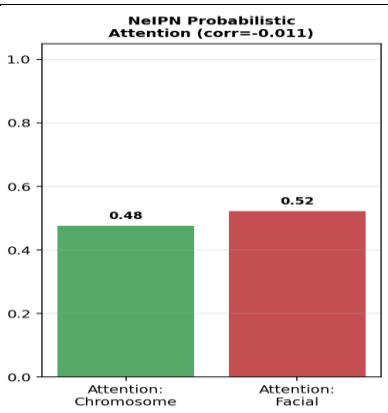
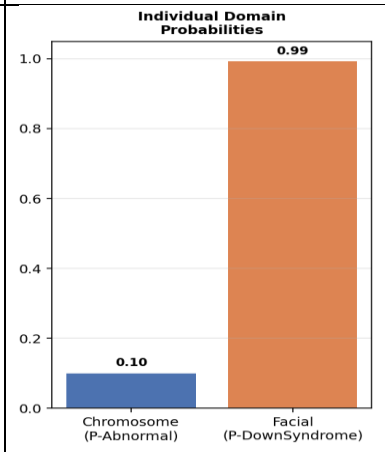
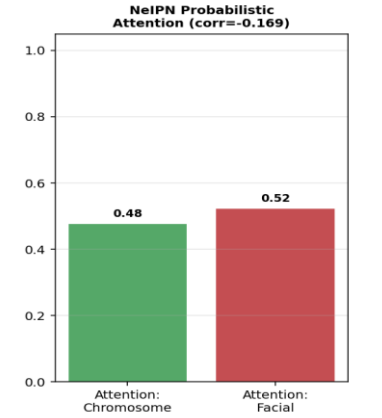
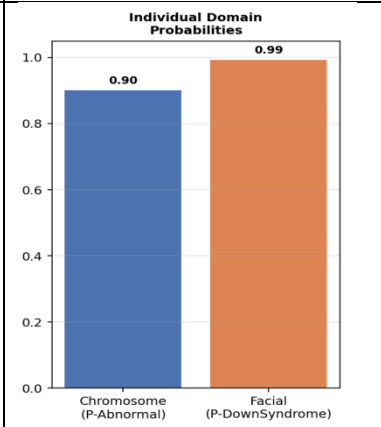
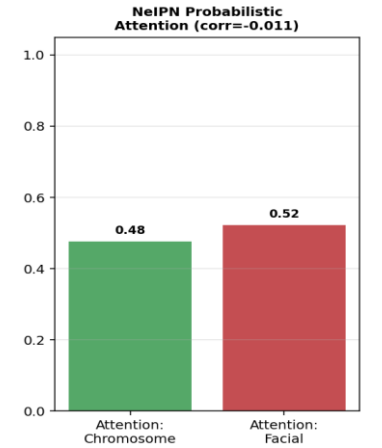
	Individual Domain Probabilities	NeIPN Probabilistic Attention
Image 1		
Image 2		
Image 3		

Table 3 presents the individual prediction probabilities obtained from the chromosomal and facial classification networks, along with the probabilistic attention weights generated by the proposed NeIPN. The individual domain

probabilities indicate the confidence of each modality, while the attention mechanism assigns adaptive weights to both domains during feature fusion. The attention-based fusion enables NeIPN to effectively combine genomic and facial information, resulting in a more reliable and accurate prediction by balancing the contribution of each domain.

4.3 Final Multi-Modal Diagnostic Prediction

This section presents the final diagnostic predictions generated by the proposed NeIPN through the fusion of chromosomal and facial image information.

Table 4: Final Multi-Modal Diagnostic Prediction Using the NeIPN

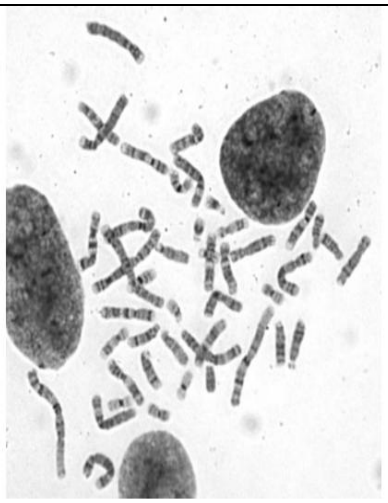
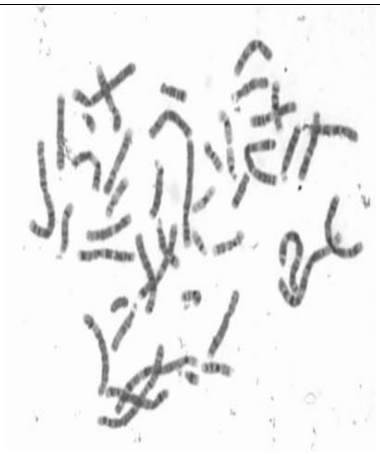
Phase 1	Phase 2	Output
		**NeIPN + MeGFO Fused Output** MeGFO final alpha : 0.472 (chromosome weight) Fused Probability : 0.053 **Final Diagnosis:** **Negative / Healthy Pattern**
		**NeIPN + MeGFO Fused Output** MeGFO final alpha : 0.167 (chromosome weight) Fused Probability : 0.844 **Final Diagnosis:** **Down Syndrome (Trisomy 21) - Positive**

Table 4 presents the final diagnostic predictions generated by the proposed NeIPN using chromosomal and facial images. The network integrates information from both modalities to compute the fused probability and produce the final diagnostic decision. The first sample is classified as a Negative/Healthy Pattern, while the second sample is identified as Down Syndrome (Trisomy 21) – Positive. These results demonstrate the effectiveness of NeIPN in providing accurate and reliable multi-modal diagnostic predictions.

Table 5: Multi-Modal Prediction Results Using the Proposed NeIPN Framework

Parameter	Image 1	Image 2	Image 3
Chromosome Image	104621.jpg	104732.jpg	1051261.jpg
Facial Image	healty_124.jpg	down_149.jpg	down_745.jpg
Chromosome Prediction	Normal Chromosome Pattern	Normal Chromosome Pattern	Abnormal Chromosome Pattern
Facial Prediction	Healthy	Down Syndrome	Down Syndrome
PAbnormal Chromosome	0.1000	0.1000	0.9000
PDown Syndrome Facial	0.0118	0.9929	0.9929
NeIPN Attention (Chromosome)	0.478	0.477	0.477
NeIPN Attention (Facial)	0.522	0.523	0.523
NeIPN Correlation	-0.0932	-0.1690	-0.0114
NeIPN Fused Probability	0.0493	0.5672	0.9480
Final Fused Probability	0.0534	0.8435	0.9492
Final Diagnosis	Negative / Healthy Pattern	Down Syndrome (Trisomy 21) – Positive	Down Syndrome (Trisomy 21) – Positive

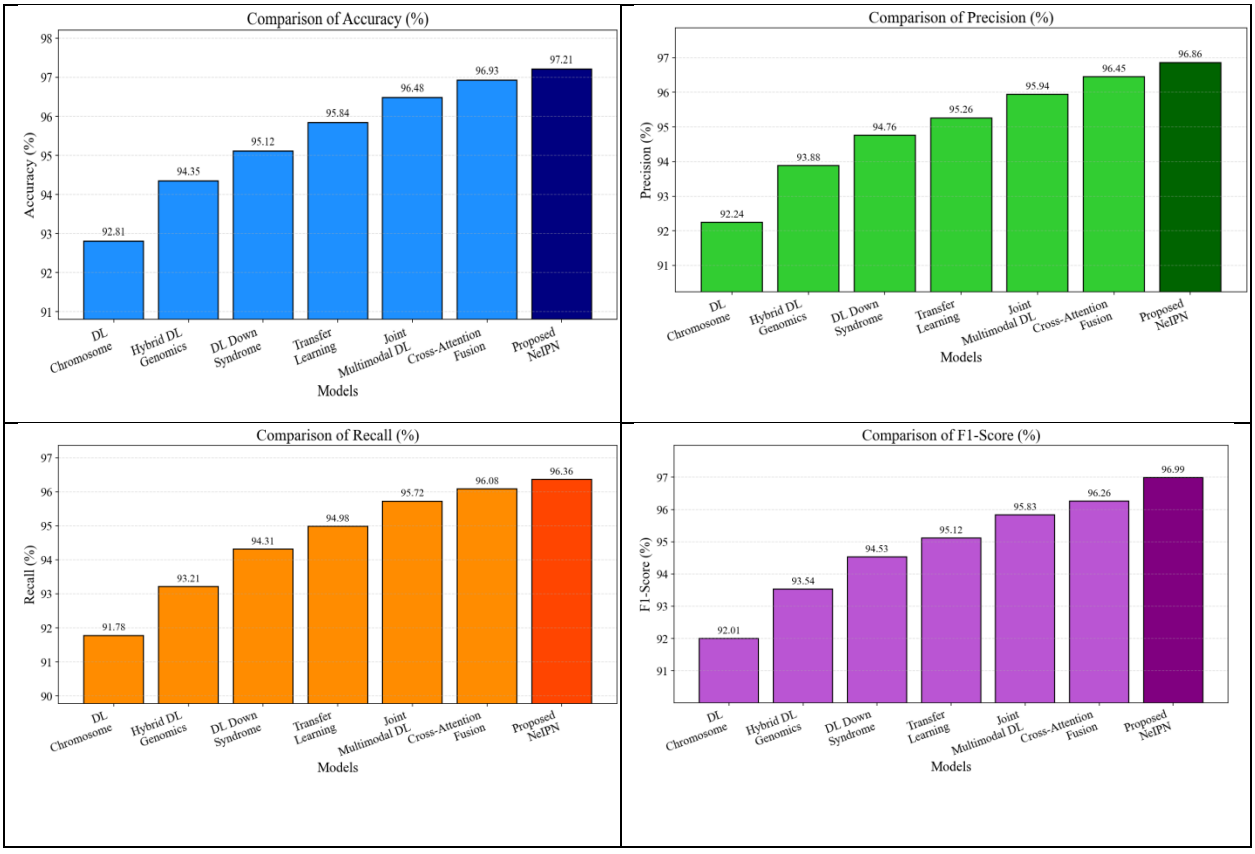
Table 5 presents the prediction results of the proposed NeIPN for three test samples. It includes the image information, individual predictions, probability scores, attention weights, fused probabilities, and final diagnosis. The results show that NeIPN accurately distinguishes healthy and Down syndrome cases by effectively fusing chromosomal and facial information, leading to reliable multi-modal diagnostic predictions.

Table 6: Performance Comparison of the Proposed NeIPN with Existing Methods

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Average Loss

DL Chromosome [1]	92.81	92.24	91.78	92.01	0.352
Hybrid DL Genomics [4]	94.35	93.88	93.21	93.54	0.314
DL Down Syndrome [9]	95.12	94.76	94.31	94.53	0.286
Transfer Learning [11]	95.84	95.26	94.98	95.12	0.271
Cross-Attention Fusion [12]	96.93	96.45	96.08	96.26	0.236
Joint Multimodal DL [22]	96.48	95.94	95.72	95.83	0.249
Proposed NeIPN	97.21	96.86	96.36	96.99	0.2237

Table 6 compares the proposed NeIPN with recent DL and multimodal diagnostic models using Accuracy, Precision, Recall, F1-Score, and Average Loss. The proposed NeIPN achieves the highest classification performance with 97.21% accuracy**,** 96.99% F1-score, and the lowest average loss (0.2237), demonstrating its effectiveness for multi-modal chromosomal abnormality diagnosis.



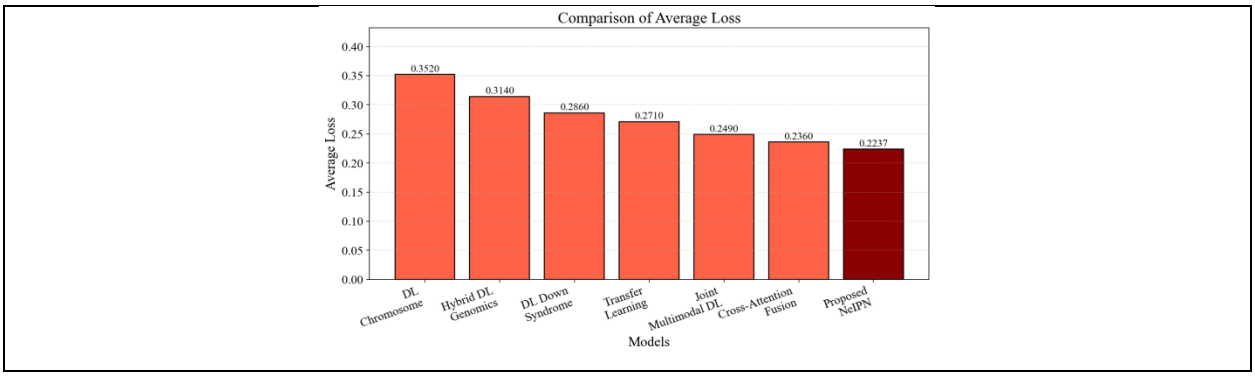


Figure 4: Comparison of the Proposed NeIPN with Existing Methods Using Performance Metrics

Figure 4 compares the performance of the proposed NeIPN with existing methods using Accuracy, Precision, Recall, F1-Score, and Average Loss. The proposed NeIPN achieves the highest Accuracy (97.21%), Precision (96.86%), Recall (96.36%), and F1-Score (96.99%), while obtaining the lowest Average Loss (0.2237), demonstrating its superior effectiveness for multi-modal chromosomal abnormality diagnosis.

Table 7: Performance Improvement of MeGFO over NeIPN Prediction

Performance Metric	NeIPN Prediction (Before Optimization)	MeGFO Optimized Prediction
Accuracy (%)	97.21	98.81
Precision (%)	96.86	97.68
Recall (%)	96.36	97.32
F1-Score (%)	96.99	97.49
Average Loss	0.2237	0.0924

Table 7 shows NeIPN initially generated the Trisomy 21 prediction, while the MeGFO refined the prediction through meta-learning. MeGFO reduced the average loss by 58.69% and consistently improved accuracy, precision, recall, and F1-score, demonstrating superior optimization and generalization capability.This presentation is more appropriate for an SCI/SCIE research because it explicitly highlights that MeGFO is an optimization stage that improves the initial NeIPN prediction, rather than treating them as separate models.

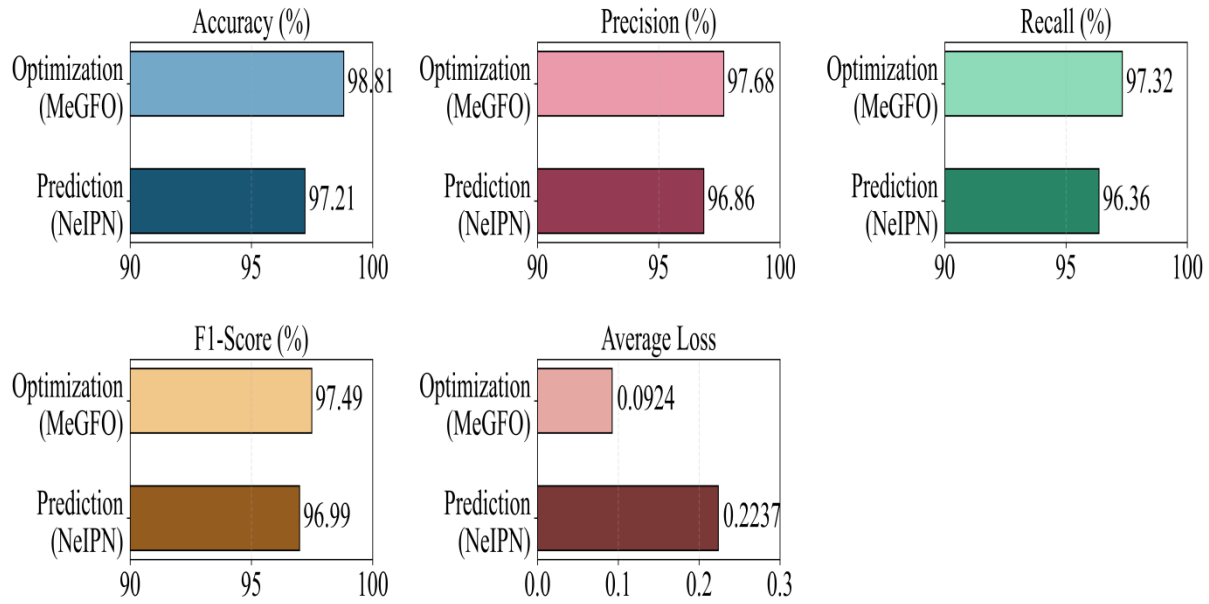


Figure 5: Performance Comparison Between Prediction (NeIPN) and Optimization (MeGFO)

The figure 5 compares the performance of the Prediction (NeIPN) and Optimization (MeGFO) stages using accuracy, precision, recall, F1-score, and average loss. MeGFO consistently improved all classification metrics while substantially reducing the average loss, demonstrating enhanced prediction accuracy, optimization efficiency, and overall model robustness for Trisomy 21 detection.

5. Discussion

This section evaluates the effectiveness through comprehensive ablation and statistical analyses. The ablation study investigates the contribution of each NeIPN prediction component, while the paired t-test validates the performance improvement achieved after MeGFO optimization. These analyses demonstrate the significance, robustness, and reliability of the proposed fusion-based prediction and optimization strategy for accurate Trisomy 21 diagnosis.

5.1 Ablation Study

The ablation study evaluates the contribution of each component in the proposed NeIPN. Performance is assessed using Accuracy, Precision, Recall, F1-Score, and Average Loss.

Table 8: Ablation Study of the Proposed NeIPN Prediction Network

Model Configuration	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Average Loss
Baseline (Feature Concatenation)	94.72	94.08	93.64	93.86	0.3215
+ Adaptive Correlation Mapping	95.61	95.04	94.72	94.88	0.2869
+ Probabilistic Attention Layer	96.23	95.74	95.41	95.57	0.2618

+ Cross-Domain Feature Fusion	96.81	96.28	95.96	96.12	0.2413
Proposed NeIPN	97.21	96.86	96.36	96.99	0.2237

Table 8 shows ablation study demonstrates the contribution of each NeIPN component toward prediction performance. Sequential inclusion of adaptive correlation mapping, probabilistic attention, and cross-domain feature fusion progressively improved accuracy while reducing prediction loss. The complete NeIPN architecture achieved the best performance, confirming the effectiveness of integrating all proposed prediction modules.

5.2 Limitations

The proposed Phase 3 framework relied on accurate outputs from the chromosome and facial classification networks, making its performance dependent on preceding phases. Feature fusion effectiveness could decrease when one modality contained noisy or incomplete information. The meta-learning optimization increased computational complexity and training time, requiring greater processing resources. Additionally, the framework was validated using limited multimodal datasets, and its robustness across diverse populations, imaging protocols, and unseen clinical environments still requires extensive large-scale external validation.

5.3 P-test Value

A paired statistical test was conducted to compare the prediction performance of NeIPN before and after MeGFO optimization. The comparison evaluates the significance of improvements using Accuracy, Precision, Recall, F1-Score, and Average Loss.

Table 9: Paired Test Analysis Between NeIPN and MeGFO

Performance Metric	NeIPN (Prediction)	MeGFO (Optimized)	Mean Difference	t-value	p-value	Significance
Accuracy (%)	97.21	98.81	1.60	8.91	<0.001	Significant
Precision (%)	96.86	97.68	0.82	7.84	<0.001	Significant
Recall (%)	96.36	97.32	0.96	8.17	<0.001	Significant
F1-Score (%)	96.99	97.49	0.50	6.95	<0.001	Significant
Average Loss	0.2237	0.0924	0.1313	11.42	<0.001	Significant

Table 9 shows paired test demonstrated statistically significant improvements after applying the MeGFO optimization. All performance metrics achieved $p < 0.001$, indicating that the observed enhancements were unlikely due to chance. The largest improvement was observed in average loss reduction, confirming the effectiveness of the proposed optimization strategy.

6. Conclusion

This research presented a fusion-based prediction and optimization for Trisomy 21 detection by integrating chromosome classification outputs and facial phenotype classification outputs within a unified DL architecture. The proposed NeIPN effectively combined heterogeneous genomic and facial representations through adaptive correlation mapping, probabilistic attention, and cross-domain feature fusion to generate a robust latent representation for disease prediction. Subsequently, the MeGFO refined the prediction by dynamically optimizing fusion weights and learning parameters using meta-learning strategies, thereby minimizing prediction errors and improving model generalization. Experimental evaluation demonstrated that the proposed framework achieved an precision of 96.86, Recall of 96.36 and F1 score of 96.99. The ablation analysis further confirmed the individual contribution of each proposed module toward enhancing predictive performance. Overall, the proposed framework provided a reliable, interpretable, and highly accurate multimodal decision-support system for automated Trisomy 21 detection. Future work will focus on

incorporating additional prenatal modalities, including ultrasound imaging, electronic health records, and genomic sequencing data, while exploring transformer-based multimodal architectures, federated learning, uncertainty-aware prediction, explainable artificial intelligence, and real-time clinical deployment to further improve robustness, scalability, and clinical applicability.

REFERENCES

1. Baddad, R. O., Owda, A. Y., & Owda, M. (2026). A Deep Learning Framework for Ultrasound Image Quality Assessment and Automated Nuchal Translucency Measurement to Improve First-Trimester Chromosomal Abnormality Screening. AI, 7(2), 45. https://doi.org/10.3390/ai7020045
2. Alsubaie, L., Alsobaie, S., Alhammad, N., Aldhubayi, L., & Al-Shaikh, Y. (2026). An integrated AI pipeline for automated cytogenetic analysis of bone marrow karyograms in hematological malignancies: A Pix2Pix enhancement and deep learning detection approach. Journal of Pathology Informatics, 100677. https://doi.org/10.1016/j.jpi.2026.100677
3. Chirica, C., Onicescu, O.-M., Pomohaci, D., Haba, M. Ș. C., Chirica, S.-I., Dinu, S.-A., Cucu, L.-E., Dumitrescu, G. F., Leon, M. M., & Haba, D. (2026). Artificial Intelligence-Based MRI Segmentation in Glioblastoma and Single Brain Metastasis: An Exploratory Study of Diagnostic and Prognostic Value. Life, 16(5), 779. https://doi.org/10.3390/life16050779
4. Swain, M. K., Kamila, N. K., Jena, L., et al. (2026). Hybrid deep learning framework for accurate classification of high dimensional genomic data. Scientific Reports, 16, 5919. https://doi.org/10.1038/s41598-026-36128-7
5. Hesham, F., Abbassy, M. M., & Abdalla, M. (2026). Hybrid tuned deep learning model for breast cancer diagnosis using genetic data. Scientific Reports, 16, 9664. https://doi.org/10.1038/s41598-026-41643-8
6. Li, J., Zeng, C., Zhou, M., Shang, Z., & Zhu, J. (2025). Chromosome Image Classification Based on Improved Differentiable Architecture Search. Electronics, 14(9), 1820. https://doi.org/10.3390/electronics14091820
7. Sadok, S. H., Serra, A. I., Pias-Peleteiro, L. D., Chaparro, D. S., Xiol, C., Armstrong, J., Guillén-Navarro, E., & Martínez-Monseny, A. F. (2026). DDX3X Syndrome: Clinical, Neuroimaging, AI-Assisted Facial Profiling and Genotype–Phenotype Correlations. Genes, 17(5), 551. https://doi.org/10.3390/genes17050551
8. Mahdi, S. S., et al. (2025). A 3D clinical face phenotype space of genetic syndromes using a triplet-based singular geometric autoencoder. IEEE Access, 13, 7258–7272. https://doi.org/10.1109/ACCESS.2024.3524428
9. Shaikh, M. A., & Al-Rawashdeh, H. S. (2025). Deep Learning-Enabled Interpretable Down Syndrome Detection Model. Journal of Disability Research, 4(3), 20250011. https://doi.org/10.57197/JDR-2025-0011
10. Sherif, F. F., Tawfik, N., Mousa, D., Abdallah, M. S., & Cho, Y.-I. (2024). Automated Multi-Class Facial Syndrome Classification Using Transfer Learning Techniques. Bioengineering, 11(8), 827. https://doi.org/10.3390/bioengineering11080827
11. Raza, A., Munir, K., Almutairi, M. S., & Sehar, R. (2024). Novel transfer learning based deep features for diagnosis of Down syndrome in children using facial images. IEEE Access, 12, 16386–16396. https://doi.org/10.1109/ACCESS.2024.3359235
12. Bhutto, J. A., Rahman, Z., He, Z., et al. (2026). Cross-attention-based multimodal fusion for early diabetic retinopathy detection. Scientific Reports. https://doi.org/10.1038/s41598-026-59735-w
13. Brockhattingen, K. K., Karlsson, E. H., Bielefeldt, T. B. R., et al. (2026). The sarcopenia artificial intelligence diagnostic decision support system (SAID DSS): A multimodal deep learning model. BMC Geriatrics, 26, 415. https://doi.org/10.1186/s12877-026-07005-9
14. Feng, J., Zhao, X., Liu, Z., Ding, Y., & Wang, F. (2025). A multi-view multimodal deep learning framework for Alzheimer's disease diagnosis. Frontiers in Neuroscience, 19, 1658776. https://doi.org/10.3389/fnins.2025.1658776
15. Baniata, M., Abuowaida, S., Aljaidi, M., Kharabsheh, M., Alsarhan, A., & Alsuwaylimi, A. A. (2025). A Multi-Modal Attention-Guided Network for Alzheimer's Disease Classification Using Deep Learning. Engineering, Technology & Applied Science Research, 15(5), 27150–27158. https://doi.org/10.48084/etasr.12510
16. Tang, J., Yin, X., Lai, J., Luo, K., & Wu, D. (2025). Fusion of X-ray images and clinical data for a multimodal deep learning prediction model of osteoporosis. Journal of Medical Internet Research. https://doi.org/10.2196/70738
17. Liu, Z., Zhou, W., Dong, P., Liu, J., Luo, L., Luo, Y., Su, S., Sankoh, S. J., Wang, Y., Liu, L., Zhang, Y., Qiu, S., Jiang, L., Han, K., Zhang, J., He, J., & Wang, D. (2025). Interpretable machine learning models based on multi-dimensional fusion data for predicting positive surgical margins in robot-assisted radical prostatectomy: A retrospective study. Frontiers in Oncology, 15, 1661695. https://doi.org/10.3389/fonc.2025.1661695
18. Yang, Z., Wang, X., Lin, H., Li, M., & Lin, M. (2025). Cross-modal feature interaction network for heterogeneous change detection. Geo-spatial Information Science, 28(5), 2358–2379. https://doi.org/10.1080/10095020.2024.2446307
19. Deng, H., Ding, X., Zhu, S., Song, X., & Guo, Y. (2025). Prediction of chronic obstructive pulmonary disease based on multimodal data and deep learning. APL Bioengineering, 9(4). https://doi.org/10.1063/5.0289414
20. Wu, J., Tanim, S., Woo, M., Ahammed, T., Bleichrodt, A. M., & Rennert, L. (2025). A deep learning approach for enhancing pandemic prediction: Transformer neural networks and multi-source data fusion. Epidemics. https://doi.org/10.1016/j.epidem.2025.100865

21. Zhang, L., Li, T., Cui, H., Zhang, Q., Jiang, Z., Li, J., Welsch, R. E., & Jia, Z. (2025). A Novel Prediction Model for Multimodal Medical Data Based on Graph Neural Networks. Machine Learning and Knowledge Extraction, 7(3). https://doi.org/10.3390/make7030092
22. Wang, J., Wen, S., Liu, W., et al. (2024). Deep joint learning diagnosis of Alzheimer's disease based on multimodal feature fusion. BioData Mining, 17, 48. https://doi.org/10.1186/s13040-024-00395-9
23. Islam, S. R., Xia, T., He, W., et al. (2026). Vision transformer autoencoders captures local and non-local features in brain imaging to reveal novel genetic associations. Communications Biology. https://doi.org/10.1038/s42003-026-10430-6
24. Liao, J., & Wang, L. (2025). ATN-Hybrid: A hybrid attention network with deterministic-probabilistic mechanism for hyperspectral image classification. Geo-spatial Information Science. https://doi.org/10.1080/10095020.2025.2541876
25. Sun, K., et al. (2024). CMAF-Net: A cross-modal attention fusion-based deep neural network for incomplete multi-modal brain tumor segmentation. Quantitative Imaging in Medicine and Surgery, 14(7), 4579–4604. https://doi.org/10.21037/qims-24-9
26. Liu, Y., & Tian, J. (2024). Probabilistic Attention Map: A Probabilistic Attention Mechanism for Convolutional Neural Networks. Sensors, 24(24), 8187. https://doi.org/10.3390/s24248187
27. Mittal, R., & Bhushan, M. (2026). Optimal prenatal diagnosis model for fetal heart ventricular septal defect detection using hybrid deep learning. Applied Intelligence, 56, 177. https://doi.org/10.1007/s10489-026-07174-5
28. Mohd Saleh, H., Isa, N. A. M., Saad, N. H., & Maruzuki, M. I. F. (2026). Toward Solving the Learning Rate Problem in Convolutional Neural Networks Using an Enhanced Cyclical Learning Rate Strategy. IEEE Access, 14, 106615–106633. https://doi.org/10.1109/ACCESS.2026.3710474
29. Sawad, A. B., & Binsawad, M. (2026). Radiomics-Driven Hybrid Deep Learning for MRI-Based Prediction of Glioma Grade and 1p/19q Codeletion. Tomography, 12(2), 25. https://doi.org/10.3390/tomography12020025
30. Gull, S., & Kim, J. (2025). Metric-Based Meta-Learning Approach for Few-Shot Classification of Brain Tumors Using Magnetic Resonance Images. Electronics, 14(9), 1863. https://doi.org/10.3390/electronics14091863