International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(7s) 2026 pp. 398-426
https://doi.org/10.70917/ijcisim-2026-3105
Article

Facial Expression based Down Syndrome Detection Using Neuro Versatile Resolution System

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: Down syndrome is a genetic disorder characterized by distinctive craniofacial morphology and facial expression variations, which makes facial image analysis very helpful in screening for Down syndrome at an early stage. This research proposes a methodological approach to developing a facial expression-based detection system for Down syndrome identification. The proposed approach includes four sequential modules, Facial Feature Preservation Network (FaFPN) for adaptive denoising and illumination correction, Integrated Attention Multi Feature Engine (IAM-FE) for attention-guided segmentation of eyes, nose, mouth and facial contours, Transfer Driven Attention Extractor (TDA-E) for deep phenotype feature generation and Neuro Versatile Resolution System (NeVRS) for multi-resolution classification. All these components are integrated into the developed framework, implemented using Python, Deep Learning (DL) libraries for image processing, feature extraction and classification. According to the experimental results, the proposed model achieved an accuracy of 96.72 %. This result illustrates the high level of diagnostic discrimination and the stability of the proposed method for facial phenotype recognition. In summary, the proposed framework provides a reliable way to automate the diagnosis of Down syndrome facial phenotypes and can be used for future development of explainable and real-time decision-support systems.

Keywords: Attention Mechanism, Deep Learning, Down syndrome Detection, Facial Expression Analysis, Facial Features Preserving Network, Multi-Resolution Classification

1. Introduction

Down syndrome is among the most frequently observed genetic disorders linked to unique craniofacial characteristics, facial expressions and developmental variations, thus requiring facial image analysis as a vital, non-invasive support tool in early screening and clinical decision assistance. Recent research has shown great promise of Deep Learning (DL) in automatically interpreting facial and medical images. DL-based facial emotion detection has been utilized in detecting facial emotions of Down syndrome children during dolphin-assisted therapy [1]. Automated multi-class facial syndrome classification has also been done using transfer learning (TL) techniques [2], whereas explainable ensemble TL using VGG16, ResNet50, InceptionV3 and Grad-CAM techniques has led to increased interpretability of medical image classification tasks [3]. Visual attention based MCNN techniques have contributed significantly towards increasing medical image classification accuracy through discrimination of image regions [4], whereas deep neural networks have enabled simultaneous facial region and landmark detection [5]. Finally, the usefulness of deep learning-based face image analysis for screening of genetic syndromes has also been demonstrated [6]. Nonetheless, current techniques still suffer from several problems, such as preservation of facial features, variation in illumination and noise, segmentation of relevant regions and extraction of phenotypes from expression-based facial images.

In order to overcome some of these shortcomings, recently several research studies have investigated vision transformer-based segmentation [7], grouped multi-scale vision transformers for medical image segmentation [8] and multi-axis vision transformer for better spatial feature modeling [9]. Furthermore, lightweight attention-augmented CNN has been used for efficient medical image classification [10]. In addition, multi-channel attention and transfer learning have enhanced interpretable disease classification in histopathology images [11]. Likewise, more advanced CNN with attention mechanism has been used for Alzheimer's disease classification using MRI images, demonstrating that attention-based learning is helpful in medical diagnostics [12]. Following these advancements, this research presents a unified FaFPN–IAM-FE–TDA-E–NeVRS framework for detecting Down syndrome using facial expressions. Specifically, this method uses Facial Feature Preservation Network (FaFPN) for adaptive denoising and the uses Integrated Attention Multi-Feature Engine (IAM-FE) for attention-guided facial region segmentation, followed by Transfer-Driven Attention Extractor (TDA-E) for deep phenotype feature extraction and finally Neuro-Versatile Resolution System (NeVRS) for multi-scale diagnostic classification. This combined pipeline helps to enhance feature preservation, region localization, representation learning with attention and classification accuracy for automated facial phenotype evaluation.

The major contributions and motivations of this research are as follows:

- This research is motivated by the need for an automated and non-invasive facial expression based system for Down syndrome detection.
- The FaFPN model makes a contribution of adaptive denoising by which noise and illumination variations can be eliminated without compromising on craniofacial characteristics.
- The IAM-FE model makes a contribution of attention-oriented segmentation where important regions of face like the eyes, nose, mouth and facial boundaries can be localized properly.
- The TDA-E model makes a contribution of transfer learning-driven feature extraction that helps produce attention-enabled facial phenotype embeddings.
- The NeVRS model contributes multi-resolution classification to improve diagnostic accuracy, reliability and phenotype-level decision support.

Organization: Section 2 presents the related background studies on facial expression analysis, medical image classification, attention mechanisms, and Down syndrome detection. Section 3 presents the proposed FaFPN–IAM-FE–TDA-E–NeVRS methodology, Section 4 discusses the experimental results and performance evaluation, and Section 5 concludes the research with key findings and future scope.

2. Background Study

Allogmani, A. S. et al. (2024) [13] proposed a classification scheme for cervical lesions by combining TL technique with Archimedes Optimization Algorithm (AOA) for optimizing the process of feature selection and classification. Despite the improvement of diagnostic performance of medical images, the approach was confined to the cervical cancer detection task only and did not consider facial phenotypic traits or expressions.

Salehi, A., & Khedmati, M. (2025) [14] introduced a hybrid oversampling and undersampling algorithm based on clustering to tackle the problem of class imbalance in multiclass classification problems. The work made improvements in terms of class balance and stability of classification models; however, it was devoted to preprocessing techniques rather than syndrome recognition in the face.

Malik, P. et al. (2025) [15] presented a driver emotion recognition model that utilizes Action Units (AUs) and micro-expressions for recognizing emotional states. This approach successfully recognized subtle facial movements; however, it was not aimed at diagnosing genetic syndromes. The obtained results indicate that the proposed model achieved high accuracy in classifying driver's emotional state.

Kroczek, L. O. et al. (2024) [16] examined the correlation between facial emotions and predicting human actions through social interaction using behavioral analysis tools. This research sheds light on predicting human actions through facial emotions; however, it did not use an automatic DL-based facial diagnosis system. The findings indicate that facial emotions affect human action perception.

Shaikh, M. A.et al. (2025) [17] proposed a DL-based model for the analysis of facial images to automatically diagnose Down Syndrome based on distinct facial features. The model performed well in recognizing distinctive features of the syndrome but could not utilize dynamic changes in facial expressions. Experimental results indicated strong Down syndrome classification accuracy.

Shi, L. (2025) [18] developed an interpretable diagnostic approach by using DL and Explainable Artificial Intelligence (XAI). This work provided greater transparency and interpretability to the model, but had nothing to do with facial expression or syndrome recognition. Experimental results proved high diagnostic interpretability without reducing classification performance.

Zhou, S. et al. (2025) [19] introduced a Large Language Model (LLM) framework for explainable disease diagnosis and decision-making that considered uncertainty. This enhanced confidence and interpretability, but utilized only textual information from the medical literature, not facial images. It yielded greater accuracy and explainability in diagnosis.

Liu, H. et al. (2021) [20] proposed a facial phenotyping framework based on Deep Convolutional Neural Networks (DCNNs) for Williams–Beuren syndrome recognition. This successfully identified syndrome-specific facial features but applied to only one genetic disease. These results proved the efficiency of deep facial phenotyping in genetic syndromes screening.

Rajawat, A. S. et al. (2023) [21] developed a hybrid framework integrating Fuzzy Logic and DL for depression detection through facial expression analysis. The framework enhanced emotion-based decision-making but was used for the evaluation of mental health condition and not syndrome detection. Experiments have indicated that the framework achieved better depression recognition due to facial expressions.

Shahzad, T. et al. (2023) [22] presented an algorithm for facial expression recognition based on facial zoning techniques and DL algorithms to improve local feature extraction. The framework increased the efficiency of expression recognition but did not take into account the syndromic facial disorders.

Bisogni, C. et al. (2023) [23] studied the impact of distance-based facial emotion recognition using hand-crafted features and deep learning models. The authors emphasized the advantage of deep learning over other methods for strong emotion detection, even though the research did not explicitly focus on syndrome diagnosis. Deep learning models proved to be superior to other feature extraction methods for emotion recognition tasks.

Table 1: Background Study of Multimodal Emotion Recognition and Facial Expression Analysis

Author(s) & Year	Concept	Methods Used	Research Gap	Limitations	Results
Islam, M. M. et al. (2024) [24]	Multimodal emotion recognition in healthcare analytics	DL-based model-level fusion using multimodal emotion features	Did not focus on Down syndrome or syndrome-specific facial morphology	Requires multiple modalities and complex fusion processing	Improved healthcare emotion recognition through multimodal feature integration
Essahraui, S. et al. (2025) [25]	Real-time driver drowsiness detection using facial analysis	Facial analysis and ML techniques	Focused on driver fatigue, not medical facial-expression diagnosis	Performance it vary under illumination, pose and real-time camera conditions	Achieved effective real-time drowsiness detection using facial cues
Simić, N. et al. (2024) [26]	Emotion recognition using federated multimodal learning	FL with CNN-based multimodal emotion recognition	Did not address clinical syndrome detection or facial phenotyping	Federated training it increase system complexity and communication cost	Enhanced emotion recognition while preserving distributed data privacy

Mamieva, D. (2023) [27]	Multimodal emotion detection using facial and speech features	Attention-based fusion of facial and speech features	Not designed for Down syndrome facial expression interpretation	Requires synchronized facial and speech inputs	Improved emotion detection accuracy through attention-based multimodal fusion
Mutawa, A. M., & Hassouneh, A. (2024) [28]	Real-time patient emotion recognition	ML with facial expression and EEG signal fusion	Focused on patient emotion monitoring, not syndrome classification	EEG acquisition increases hardware dependency and complexity	Improved real-time patient emotion recognition using multimodal signals
Das, S. et al. (2024) [29]	Driver drowsiness detection through facial movement analysis	DL with U-Net-based facial movement analysis	Application limited to driver monitoring, not clinical facial diagnosis	Requires reliable facial movement tracking in IoT environments	Detected drowsiness effectively using facial movement-based DL
Gaya-Morey, F. X. et al. (2026) [30]	Facial expression recognition in individuals with intellectual disabilities	DL-based facial expression recognition models	Does not specifically target Down syndrome detection	It require larger and more diverse disability-specific datasets	Demonstrated the potential of DL for facial expression recognition in intellectual disability groups

Table 1 presents an overview of the relevant literature on multimodal emotion recognition, face analysis and DL-based expression recognition approaches. These pieces of literature establish the theoretical background for the proposed facial expression-based approach for Down syndrome diagnosis while pointing out the gaps in face phenotype characterization for the specific syndrome.

2.1 Research Gap

- Current Down syndrome detection techniques inadequately capture subtle facial texture, edges and morphology attributes during preprocessing steps.
- Current techniques for facial analysis are strongly influenced by factors like changes in lighting conditions, sensor noise, constitue and background.
- The majority of current techniques inadequately segment clinically important regions such as the eyes, nose, mouth and facial edges.
- Current deep learning models do not adequately incorporate local phenotypic attributes along with structural connections globally in the face.
- Limited research has developed an end-to-end attention-guided, transfer-driven and multi-scale model for robust Down syndrome classification.

3. Proposed Methodology

In this section, the approach to detect Down syndrome from facial expressions is provided, which includes the workflow, system architecture, mathematical modeling and algorithmic details. The method follows a sequential pipeline consisting of FaFPN for adaptive denoising, IAM-FE for facial region segmentation, TDA-E for deep phenotype feature extraction and NeVRS for multi-scale diagnostics classification.

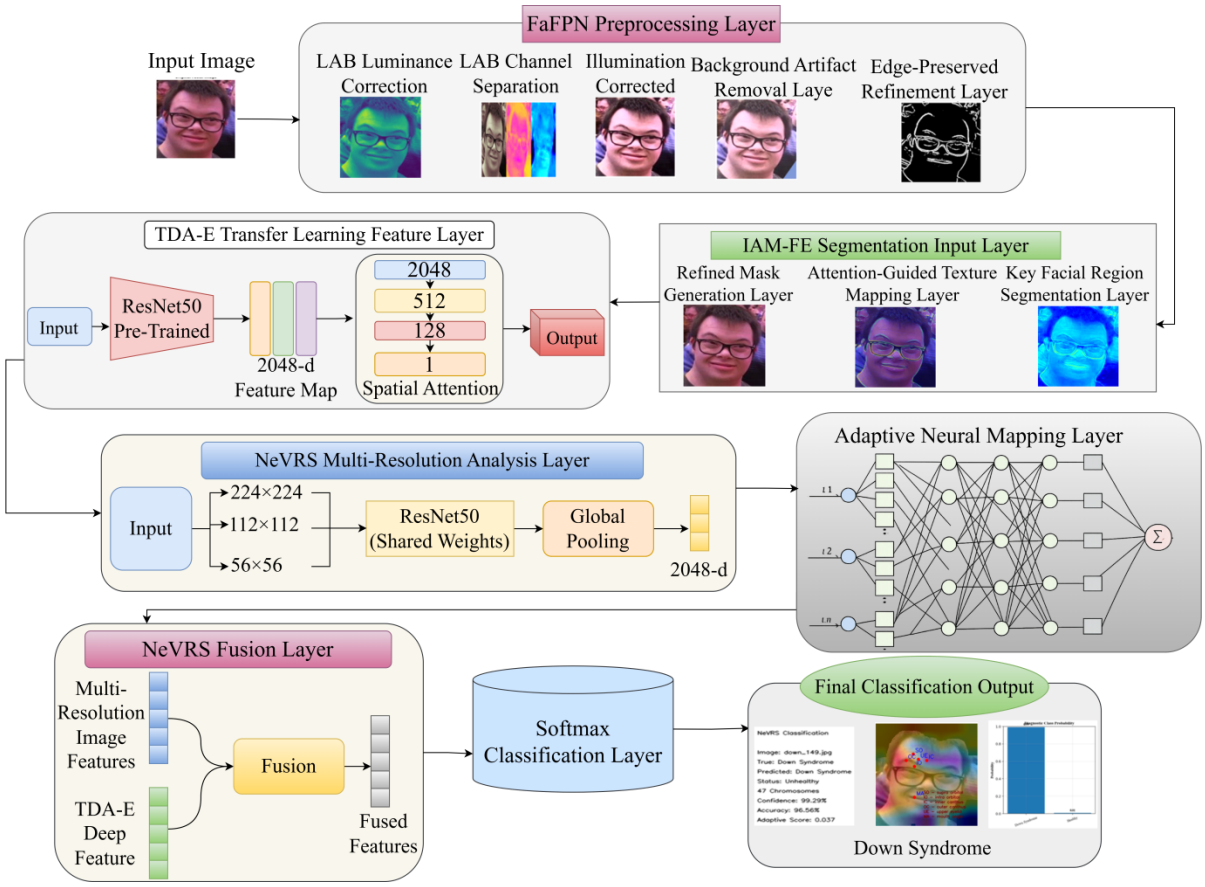


Figure 1: Architecture of the FaFPN–IAM-FE–TDA-E–NeVRS Framework for Facial Phenotype Classification

Figure 1 shows the entire process involved starting from the raw facial images as input, through FaFPN processing, IAM-FE segmentation, TDA-E transfer learning feature extraction, NeVRS multi-resolution analysis, adaptive neural mapping and finally fusion classification. Softmax classification determines the facial phenotype, like Down Syndrome, through the fused multi-resolution facial image and attention-enhanced deep facial features.

3.1 Dataset Description

Dataset Link: https://www.kaggle.com/datasets/mamunhasan2cs/down-syndrome-dataset

The Down Syndrome Dataset includes face images of individuals having Trisomy 21 and normal people as controls to provide a good for the syndrome detection based on phenotype. The database contains important facial features associated with the disease that include periocular region, nose shape, oral region and facial symmetry that are essential biomarkers for diagnosis. The heterogeneous images make it a useful database for developing machine learning models for facial feature extraction and syndrome classification purposes. Therefore, this dataset provides a useful source for validating the FaFPN-IAM-FE-TDA-E-NeVRS architecture.

3.2 Adaptive Denoising Using FaFPN

The Facial Feature Preservation Network is built as an adaptive denoising system that increases the quality of images by preserving diagnostically valuable features. It eliminates illumination variations, sensor noise and background artifacts through illumination normalization, edge-preserving filtering, texture enhancement and adaptive noise reduction. As opposed to existing denoising algorithms, FaFPN reduces feature blurring and preserves vital areas

of the face including the eyes, nose, mouth and face outlines. Computationally efficient and parallelizable pre-processing makes this approach scalable for use in massive data sets and real-time screening conditions. The improvement of structural consistence and features makes FaFPN increase the quality of the input data for segmentation, feature extraction and classification steps.

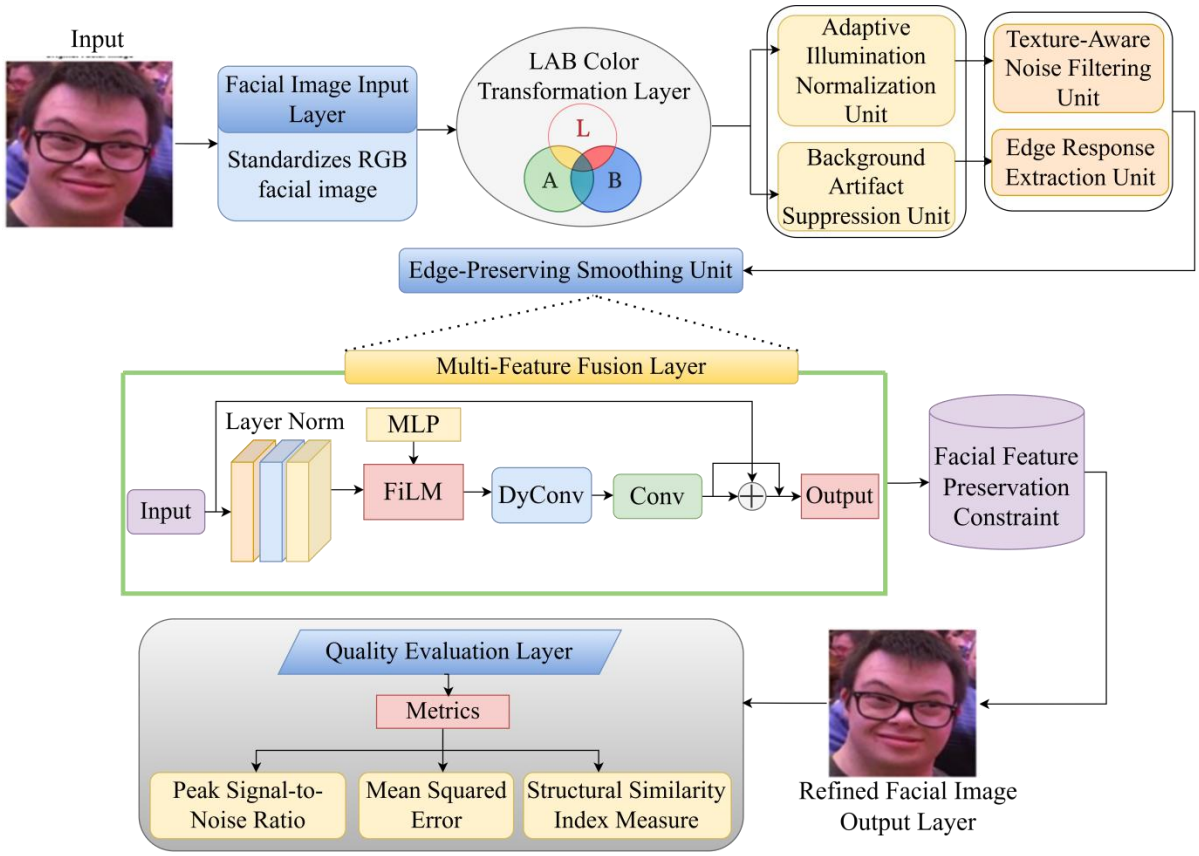


Figure 2: Architecture of the FaFPN for Facial Image Denoising

Figure 2 shows the structure of FaFPN network, whereby the noisy image of the face is subjected to standardization, conversion into LAB color space, illumination correction, noise removal, removal of background artifacts and extraction of edge-aware features. The output of this network is an enhanced version containing vital facial information.

$$I_n(x,y) = \frac{I(x,y)-\mu_I}{\sigma_I+\epsilon} \quad (1)$$

In equation 1 represents that $I(x,y)$ refers to original facial image intensity at pixel (x,y), $I_n(x,y)$ denotes normalized image intensity, μ_I indicate a mean image intensity, σ_I is a standard deviation of image intensity, ϵ is a small constant preventing division by zero. This equation equals the variations in illumination and inconsistencies in brightness prior to denoising.

$$D(x,y) = I_n(x,y) - \lambda N(x,y) \quad (2)$$

In equation 2 represents that $D(x,y)$ denotes a denoised facial image, $N(x,y)$ indicate an estimated noise component, λ is an adaptive noise suppression coefficient. This equation is able to eliminate the sensor and environment noise and keep the meaningful information of faces.

$$E_p = \sum_{x,y} |\nabla D(x,y) - \nabla I(x,y)| \quad (3)$$

In equation 3 represents that E_p denotes an edge preservation loss, $\nabla D(x, y)$ indicate a gradient of denoised image, $\nabla I(x, y)$ is a gradient of original image. In this way, it is guaranteed that relevant facial boundaries and contours are maintained during denoising.

$$T(x, y) = D(x, y) + \alpha \cdot G(x, y) \quad (4)$$

In equation 4 represents that $T(x, y)$ denotes a texture-enhanced image, $G(x, y)$ is a local texture gradient map, α indicate a texture enhancement weight. This equation strengthens subtle facial textures and expression-related details important for diagnosis.

$$F_{out} = \beta T(x, y) + (1 - \beta) M(x, y) \quad (5)$$

In equation 5 represents that F_{out} denotes a final FaFPN output image, $M(x, y)$ is a morphological facial feature map, β indicate a feature fusion coefficient. The last reconstructed face picture is produced using this equation that integrates texture and morphological facial features.

Algorithm: Facial Feature Preservation Network

Input:
- I → Raw facial image
- λ → Adaptive noise suppression coefficient
- α → Texture enhancement weight
- β → Feature fusion coefficient
- ε → Small constant for numerical stability

Step 1: Image Normalization
- Compute mean intensity μI from input image I
- Compute standard deviation σI from input image I
- For each pixel (x, y) in I do
  - In(x, y) ← (I(x, y) - μI) / (σI + ε)
- End For

Step 2: Noise Estimation
- Estimate noise component N(x, y) from normalized image In
- using local variance or adaptive filtering

Step 3: Adaptive Denoising
- For each pixel (x, y) in In do
  - D(x, y) ← In(x, y) - λ × N(x, y)
- End For

Step 4: Edge Preservation
- Compute gradient of original image ∇I(x, y)
- Compute gradient of denoised image ∇D(x, y)
- Calculate edge preservation loss:
  - $E_p = \sum_{x,y} |\nabla D(x, y) - \nabla I(x, y)|$
- If E_p is high then
  - Adjust λ to reduce edge distortion
- End If

Step 5: Texture Enhancement
- Compute local texture gradient map G(x, y)
- For each pixel (x, y) in D do
  - $T(x, y) = D(x, y) + \alpha \cdot G(x, y)$
- End For

Step 6: Morphological Feature Mapping
- Extract morphological facial feature map M(x, y)

from important regions such as eyes, nose, mouth and facial contour

Step 7: Feature Fusion

For each pixel (x, y) do

$F_{out}(x,y) = \beta T(x,y) + (1-\beta)M(x,y)$

End For

Step 8: Output Generation

Return F_{out} as the final feature-preserved denoised facial image

End

Output:

$F_{out} \rightarrow$ Feature-preserved denoised facial image

Begin

FaFPN Pseudocode describes the denoising procedure that involves the following stages: image normalization, noise estimation, noise reduction based on the adaptive algorithm, edge protection and texture enhancement. The final stage involves fusing the enhanced texture with the facial features.

3.3 Attention-Guided Facial Region Segmentation Using IAM-FE

The Integrated Attention Multi-Feature Engine aims to segment diagnostically relevant facial regions related to Down syndrome phenotypic features. It combines the benefits of spatial attention, channel-wise feature recalibration, multi-scale feature aggregation and region-aware segmentation so as to emphasize the eyes, nose, mouth and facial contours. IAM-FE's suppression of irrelevant background information results in enhanced localization of clinically relevant facial structures. The model is able to capture local morphology and global facial context to provide discriminative region representations. It is efficient and parallelizable and can be used for large-scale facial datasets and real-time screening applications.

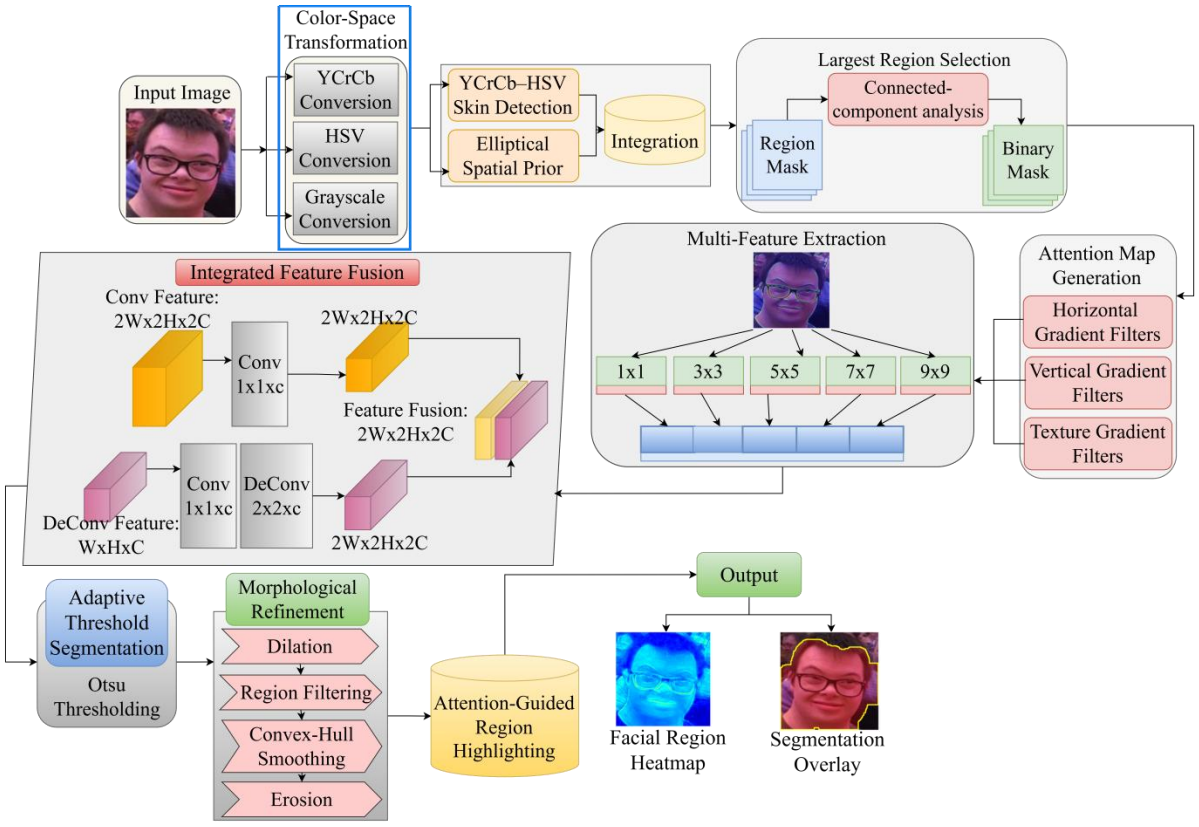


Figure 3: Architecture of the Proposed IAM-FE for Facial Region Segmentation

Figure 3 presents the proposed architecture of IAM-FE, in which the input face image undergoes conversion to the YCrCb, HSV and grayscale forms. Multi-scale feature extraction, attention generation, Otsu segmentation and morphological processing generate the facial region heatmap and the refined segmentation map, respectively.

$$A_s = \sigma\left(f^{7\times7}([Avg(F); Max(F)])\right) \quad (6)$$

In equation 6 represents that A_s denotes a spatial attention map, F is an input facial feature map, $Avg(F)$ refers to average pooled feature map, $Max(F)$ indicates a maximum pooled feature map, [;]Concatenation operation, $f^{7\times7}$ Convolution operation with kernel size 7×7, σ Sigmoid activation function. This equation determines the informative Spatial Face Regions (SFLs) like eyes, nose and mouth.

$$A_c = \sigma\left(W_2\left(ReLU\left(W_1(F)\right)\right)\right) \quad (7)$$

In equation 7 represents that A_c Channel attention vector, W_1, W_2 Learnable weight matrices, $ReLU$ Rectified Linear Unit activation. This equation puts more weight on discriminative channels of facial features and suppresses redundant information.

$$F_a = (A_s \odot F) + (A_c \odot F) \quad (8)$$

In equation 8 represents that F_a Attention-enhanced feature map, A_s Spatial attention map, $\odot$ Element-wise multiplication. This equation integrates spatial and channel attention to provide a very informative facial representation.

$$F_m = \sum_{i=1}^{n} \omega_i F_i \quad (9)$$

In equation 9 represents that F_m Aggregated multi-scale feature map, F_i Feature map at scale i, ω_i Adaptive scale weight, n Number of feature scales. This equation combines local and global information from the face for better segmentation robustness.

$$S(x,y) = \begin{cases} 1, & P(x,y) \geq T \\ 0, & P(x,y) < T \end{cases} \quad (10)$$

In equation 10 represents that $S(x,y)$ Segmentation mask, $P(x,y)$ Predicted region probability, T Segmentation threshold, 1 Target facial region, 0 Background region. This equation creates the final segmentation mask which is used to separate the important facial regions from the background.

Algorithm: Integrated Attention Multi-Feature Engine (IAM-FE)

Input:
- F_{out} → Feature-preserved facial image from FaFPN
- T → Segmentation threshold
- n → Number of multi-scale feature levels

Begin

Step 1: Feature Map Generation
- Extract input facial feature map F from F_{out} using convolutional layers

Step 2: Spatial Attention Generation
- Compute average-pooled feature map $Avg(F)$
- Compute maximum-pooled feature map $Max(F)$
- Concatenate $Avg(F)$ and $Max(F)$
- Apply 7×7 convolution operation
- Apply sigmoid activation to generate spatial attention map A_s

Step 3: Channel Attention Refinement
- Pass feature map F through learnable weight layer W_1
- Apply ReLU activation
- Pass output through learnable weight layer W_2
- Apply sigmoid activation to generate channel attention vector A_c

Step 4: Attention-Enhanced Feature Representation
  Multiply spatial attention map A_s with feature map F
  Multiply channel attention vector A_c with feature map F
  Fuse both attention outputs to obtain attention-enhanced feature map F_a
Step 5: Multi-Scale Feature Aggregation
  For each scale i = 1 to n do
    Extract scale-specific feature map F_i
    Assign adaptive scale weight ω_i
    Compute weighted feature representation $\omega_i \times F_i$
  End For
  Aggregate all weighted feature maps to obtain multi-scale feature map F_m
Step 6: Region Probability Estimation
  Apply segmentation head on F_m
  Compute predicted region probability map P(x, y)
Step 7: Region-Aware Segmentation
  For each pixel (x, y) in P do
    If P(x, y) ≥ T then
      S(x, y) ← 1
    Else
      S(x, y) ← 0
    End If
  End For
Step 8: Region Feature Extraction
  Extract segmented facial regions such as eyes, nose, mouth and facial contours
  Generate region-specific facial feature representation R
Step 9: Output Generation
  Return S(x, y) and R
End
Output:
  S(x, y) → Final segmented facial region mask
  R → Region-specific facial feature representation

The IAM-FE segmentation algorithm is defined through the use of the IAM-FE pseudocode that defines the whole segmentation procedure, including creation of the spatial and channel attention maps through FaFPN and multiscale feature fusion, region probability estimation and finally, segmentation of regions such as the eyes, nose, mouth and face contour.

3.4 Transfer-Driven Deep Feature Extraction Using TDA-E

The Transfer-Driven Attention Extractor aim to extract discriminative deep facial phenotype features from the segmented regions produced by IAM-FE. It uses a pre-trained backbone network by which domain-invariant facial texture, morphology and structural patterns are captured. The attention mechanism highlights facial features that are particularly significant for diagnosis, including eye shape, nose structure, oral morphology and facial symmetry. Contextual feature fusion: It fuses both the local and global facial representations, thereby reducing the redundancy and enhancing the features. Finally, TDA-E produces high-dimensional attention enhanced embeddings for reliable classification using NeVRS.

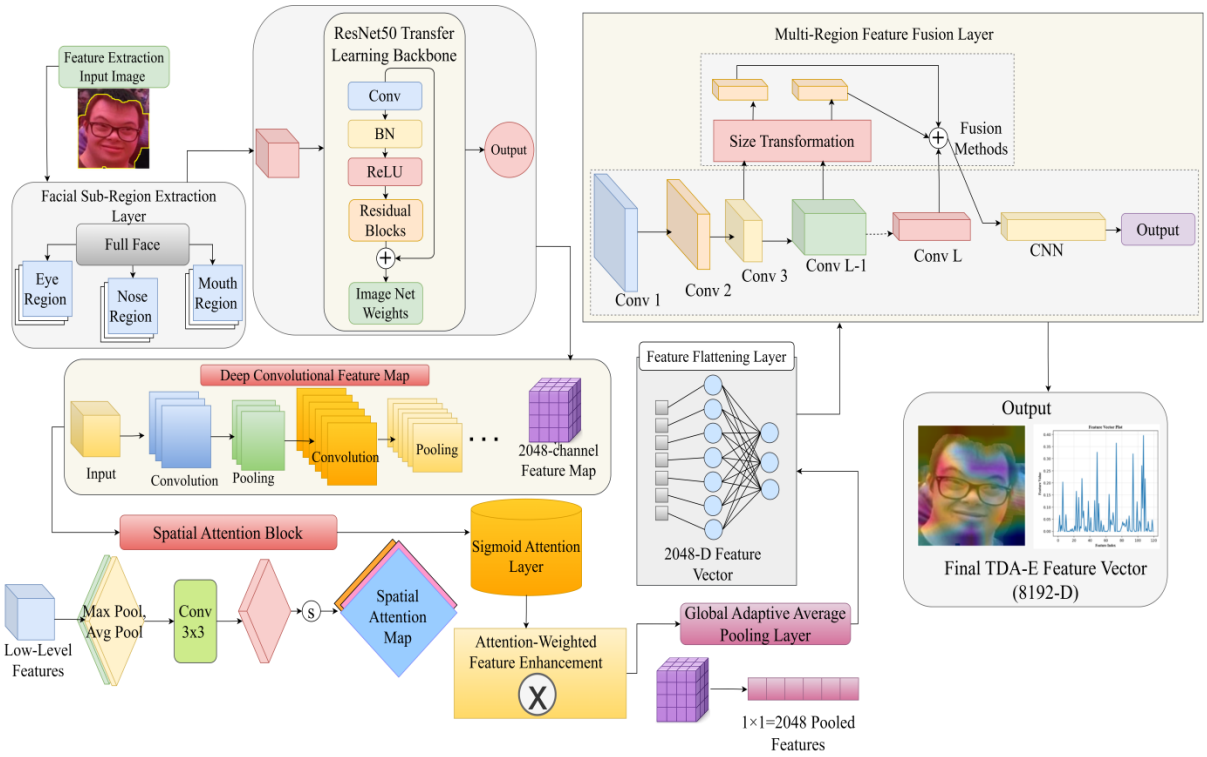


Figure 4: Architecture of TDA-E for Facial Phenotype Feature Extraction

Figure 4 shows the TDA-E model that processes segmented face, eyes, nose and mouth segments using the ResNet50 TL backbone to produce deep features with 2048 channels. The model uses spatial attention, global adaptive average pooling, feature flattening and multi-region fusion to produce an enhanced 8192-D feature representation using NeVRS model for Down syndrome detection.

$$F_t = \Phi(I_s; \theta_p) \tag{11}$$

In equation 11 represents that F_t is a transfer-learned feature representation, I_s denotes a segmented facial image from IAM-FE, Φ and θ_p refer to a pre-trained backbone network and model parameters, respectively. The following equation is used to extract the face-invariant features using a pre-trained DNN.

$$A_t = \sigma(W_f F_t + b_f) \tag{12}$$

In equation 12 represents that A_t denotes an attention weight vector, W_f is a learnable attention weight matrix, b_f indicates a bias term. This equation calculates feature importance scores to accentuate the face's character information that is diagnostic.

$$F_a = A_t \odot F_t \tag{13}$$

In equation 13 represents that F_a indicates an attention-refined feature representation. This equation emphasizes informative facial features and reduces redundant responses to features.

$$F_c = \gamma F_a + (1 - \gamma) F_g \tag{14}$$

In equation 14 represents that F_c is a contextually fused feature representation, F_g denotes a global contextual feature representation, γ indicates an adaptive fusion coefficient. This equation uses the context of the global face to enhance the robustness of features, while simultaneously incorporating local morphological details.

$$E = \psi(F_c) \tag{15}$$

In equation 15 represents that E is a final high-dimensional feature embedding, ψ denotes an embedding transformation function (fully connected layer). This equation produces the discriminative feature vector for NeVRS for the final classification.

Algorithm: Transfer-Driven Attention Extractor

Input:

I_s → Segmented facial image from IAM-FE

Φ → Pre-trained backbone network

θ_p → Pre-trained model parameters

W_f → Learnable attention weight matrix

b_f → Bias term

γ → Adaptive fusion coefficient

ψ → Embedding transformation function

Begin

Step 1: Transfer Feature Extraction

Load pre-trained backbone network Φ with parameters θ_p

Pass segmented facial image I_s into Φ

Extract transfer-learned feature representation F_t

Step 2: Attention Weight Computation

Compute attention score using learnable weight matrix W_f and bias b_f

$A_t \leftarrow \sigma(W_f F_t + b_f)$

Step 3: Attention-Guided Feature Refinement

Apply element-wise multiplication between A_t and F_t

$F_a \leftarrow A_t \odot F_t$

Step 4: Global Context Feature Extraction

Extract global contextual feature representation F_g from F_t

using global average pooling or contextual encoding

Step 5: Contextual Feature Fusion

Fuse local attention-refined features and global contextual features

$F_c \leftarrow \gamma F_a + (1 - \gamma) F_g$

Step 6: Embedding Generation

Pass fused feature representation F_c into embedding transformation function ψ

$E \leftarrow \psi(F_c)$

Step 7: Output Generation

Return E as the final discriminative feature embedding

End

Output:

E → Final high-dimensional feature embedding for NeVRS classification

The TDA-E algorithm provides the procedure for extracting features through feeding the segmented facial regions from IAM-FE using a pre-trained deep learning backbone to acquire transfer learning facial representations. The subsequent steps involve attention refinement, feature fusion and transformation into embeddings for NeVRS classification.

3.5 Multi-Scale Facial Expression Classification Using NeVRS

The Neuro-Versatile Resolution System is designed as a flexible multi-resolution classifier for detecting phenotypes associated with Down syndrome faces. The system uses the multi-resolution representation learned through TDA-E via multi-resolution feature learning, hierarchical neural mapping and adaptive feature weighing. NeVRS identifies both local features on the face and global features within the face structure to enhance the discrimination between the classes. Phenotype-specific decision fusion and softmax classification minimize

uncertainty and provide the final probability of the diagnosis. Its flexible architecture enables the reliable classification based on facial expressions under varied facial image conditions.

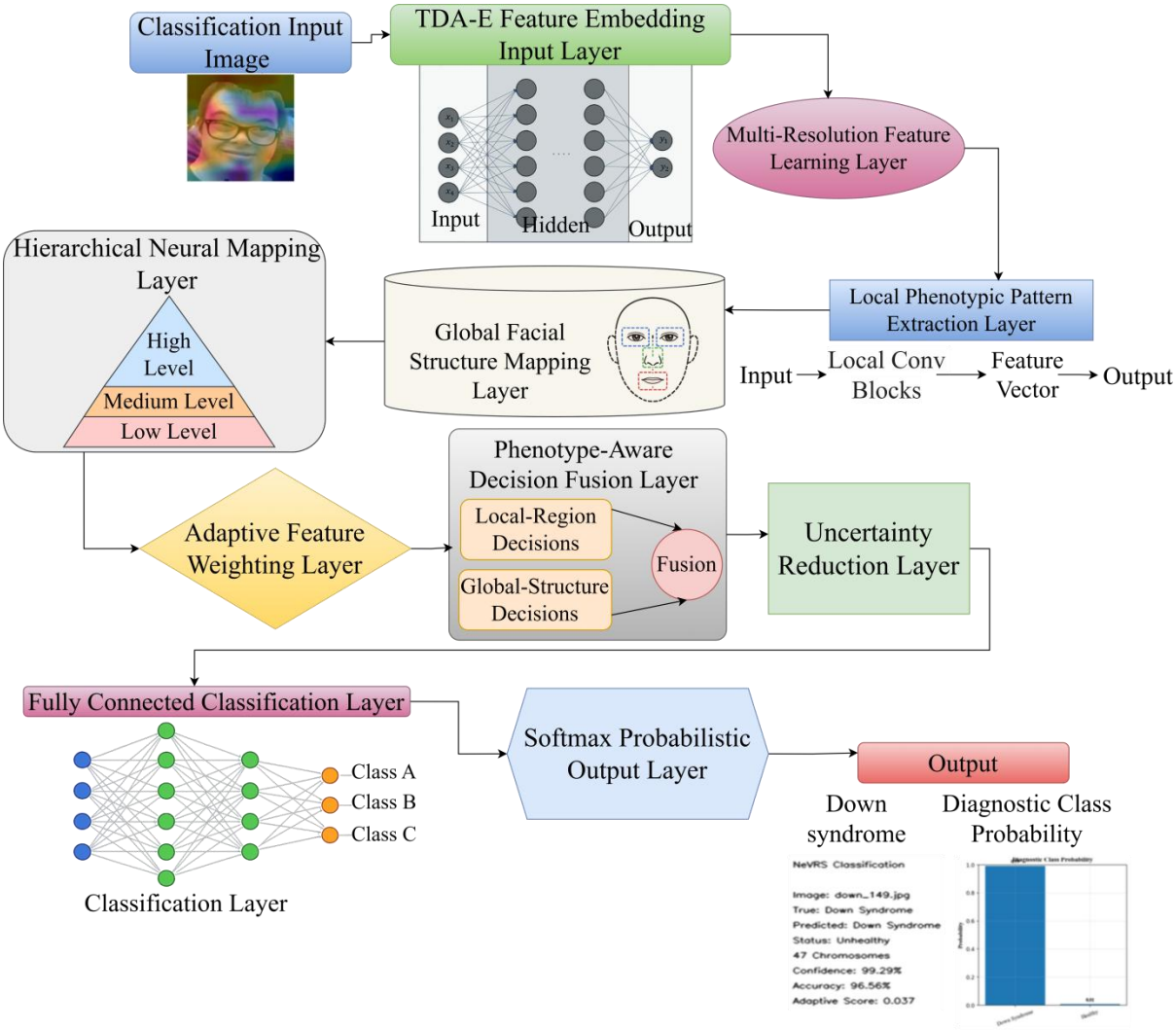


Figure 5: NeVRS Architecture for Multi-Resolution Facial Phenotype Classification

Figure 5 shows the proposed NeVRS architecture that utilizes TDA-E facial phenotype embeddings and classifies the using local pattern extraction of local patterns, global mappings, adaptive feature weightings and hierarchical neural learning. The phenotype-based decision fusion and softmax output layers produce the final classification result of Down syndrome with a confidence score.

$$R_i = \phi_i(E) \tag{16}$$

In equation 16 represents that R_i indicates a feature representation at resolution level i, E denotes a input feature embedding from TDA-E, ϕ_i is a resolution transformation function, i is a resolution scale index. This equation breaks apart the feature embedding in several resolution-specific representations for in-depth analysis.

$$W_i = \frac{\exp(R_i)}{\sum_{j=1}^{n} \exp(R_j)} \tag{17}$$

In equation 17 represents that W_i indicates a importance weight of resolution level i, n is a total number of resolution levels. The equation assigns different discriminative relevance scores to the different resolution features.

$$F_i = W_i \odot R_i \tag{18}$$

In equation 18 represents that $\odot$ indicates an element-wise multiplication. This equation highlights the information-relevant resolution features and minimizes the effect of less information-relevant representations.

$$F_m = \sum_{i=1}^{n} F_i \quad (19)$$

In equation 19 represents that F_m denotes a multi-resolution fused feature representation. This equation is a merger of fine-grained and coarse-grained facial information in a single representation.

$$D = \sigma(W_d F_m + b_d) \quad (20)$$

In equation 20 represents that D is a decision feature vector, W_d indicates a decision layer weight matrix, b_d denotes a bias vector. This equation converts fused facial features to discriminative decision representation for classification.

$$P(c_k) = \frac{\exp(D_k)}{\sum_{m=1}^{C} \exp(D_m)} \quad (21)$$

In equation 21 represents that $P(c_k)$ is a probability of class$k$, D_k is a decision score for class$k$, C indicates a total number of classes, D_m denotes a decision score of class$m$. The resulting probability distribution is used in this equation to calculate a final probability distribution and the facial image is attributed to the most likely diagnostic class.

Algorithm: Neuro-Versatile Resolution System

Input:
- E → High-dimensional feature embedding from TDA-E
- n → Number of resolution levels
- ϕ_i → Resolution transformation function at level i
- W_d → Decision layer weight matrix
- b_d → Decision layer bias
- C → Number of diagnostic classes

Begin

Step 1: Multi-Resolution Feature Decomposition
- For each resolution level i = 1 to n do
  - $R_i = \phi_i(E)$
- End For

Step 2: Resolution Importance Weight Computation
- For each resolution level i = 1 to n do
  - $W_i \leftarrow \exp(R_i) / \Sigma \exp(R_j)$, where j = 1 to n
- End For

Step 3: Weighted Feature Enhancement
- For each resolution level i = 1 to n do
  - $F_i \leftarrow W_i \odot R_i$
- End For

Step 4: Multi-Resolution Feature Fusion
- Initialize $F_m \leftarrow 0$
- For each resolution level i = 1 to n do
  - $F_m \leftarrow F_m + F_i$
- End For

Step 5: Neural Decision Fusion
- $D \leftarrow \sigma(W_d F_m + b_d)$

Step 6: Softmax Classification
- For each class k = 1 to C do
  - $P(c_k) \leftarrow \exp(D_k) / \Sigma \exp(D_m)$, where m = 1 to C
- End For

Step 7: Final Class Prediction
  Cpred ← class with maximum probability $P(c_k)$
Step 8: Output Generation
  Return C_{pred} and $P(c_k)$
End
Output:
  C_{pred} → Final predicted class label
  $P(c_k)$ → Class probability score

NeVRS Pseudocode explains the process of classification by breaking down the TDA-E feature embedding in terms of several resolution-based representation forms and using adaptive feature weighting. Neural Decision Fusion and Softmax Classification are used to obtain class probabilities for making the final diagnosis of Down syndrome or non-Down syndrome.

4. Results and Discussion

This section provides the experimental findings for the developed facial expression-based Down syndrome detection system using Python. The evaluation includes the assessment of denoising process, segmentation, feature extraction efficiency, training efficiency and finally diagnosis.

4.1 Experimental Setup

This section describes the dataset partitioning, training configuration, implementation setup, and evaluation strategy used to validate the proposed framework.

To test the effectiveness of the proposed framework, the data was divided into 80% training data and 20% testing data. The 80% data set was used for parameter tuning and feature extraction purposes while the rest of the data set was used only for validating the model. Data randomization technique was used to ensure unbiased evaluation of the proposed framework.

Table 2: Epoch-wise Training Loss

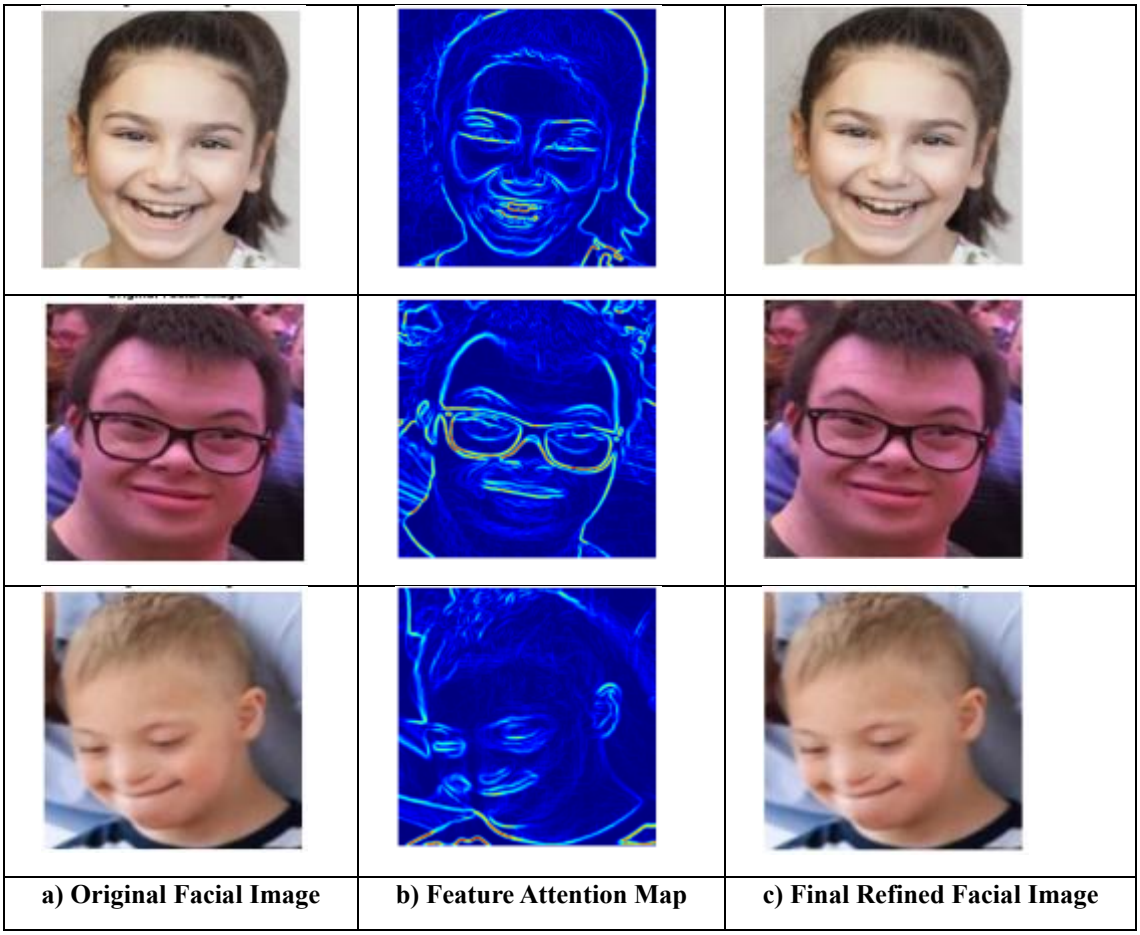
Epoch	Loss	Epoch	Loss	Epoch	Loss
1	1.1516	21	0.0875	41	0.0339
2	0.8254	22	0.0916	42	0.0335
3	0.6948	23	0.0777	43	0.0296
4	0.7311	24	0.0683	44	0.0564
5	0.6381	25	0.0817	45	0.0425
6	0.4113	26	0.0848	46	0.0284
7	0.3977	27	0.0584	47	0.0384
8	0.2904	28	0.1152	48	0.0414
9	0.3483	29	0.0840	49	0.0427
10	0.2526	30	0.0582	50	0.0392
11	0.1684	31	0.0519	51	0.0289
12	0.1270	32	0.0404	52	0.0285
13	0.1884	33	0.0489	53	0.0411
14	0.1160	34	0.0480	54	0.0347

15	0.1486	35	0.0602	55	0.0870
16	0.1077	36	0.1203	56	0.0384
17	0.2013	37	0.0537	57	0.0285
18	0.0992	38	0.0720	58	0.0235
19	0.1272	39	0.0440	59	0.0216
20	0.0903	40	0.0554	60	0.0299

Table 2 presents the epoch wise training loss which is showing a steady decline from 1.1516 in the first epoch to 0.0299 in the sixtieth epoch, TDA-E for deep phenotype feature extraction indicating stable convergence and effective optimization of the proposed model. This shows that model is capable enough to learn discriminative features of the faces of Down syndrome and non-Down syndrome people.

4.2 FaFPN-Based Denoising Performance Evaluation

This section examines the capability of FaFPN in terms of removing noise, illumination changes and other artifacts while keeping facial texture and structure intact.



a) Original Facial Image	b) Feature Attention Map	c) Final Refined Facial Image

Figure 6: Preprocessing Results of the FaFPN

Figure 6 shows the processed output from FaFPN Pre-processing, in which the raw images of the face are converted into refined images by reducing the background, illumination and noise artifacts. These refined images have significant information that works as a valuable input to the process of segmentation, feature extraction and classification.

Table 3: Denoising and Feature Preservation Analysis of the FaFPN

LAB Luminance Color Map	LAB Channel Separation	Illumination Corrected	Background Artifact Removed	Edge Preserved Output
				
				
				

Table 3 illustrates the process of noise reduction and feature preservation in FaFPN, which consists of luminance modification in LAB color space, illumination normalization, artifact detection and removal and edge-preserving enhancement. The improved images have intact facial landmark points and contours with textured surfaces.

Table 4: Denoising Performance Comparison of Existing Methods and the Proposed FaFPN

Method	MSE	PSNR (dB)	SSIM
DS-CNN [1]	0.0089	37.42	0.9115
GMS-ViT [8]	0.0067	39.85	0.9287
DNN [14]	0.0078	38.24	0.9204
Transfer Learning [17]	0.0059	41.16	0.9368
ResNet50 [23]	0.0048	42.78	0.9482
EfficientNet [30]	0.0039	44.03	0.9567
Proposed FaFPN	0.0016	46.21	0.9693

Table 4 below, the denoising performance of existing techniques versus the new FaFPN algorithm in relation to the MSE, PSNR and SSIM values is compared. From the table, it is clear that the FaFPN algorithm gives the lowest MSE value of 0.0016, highest PSNR value of 46.21dB and highest SSIM value of 0.9693.

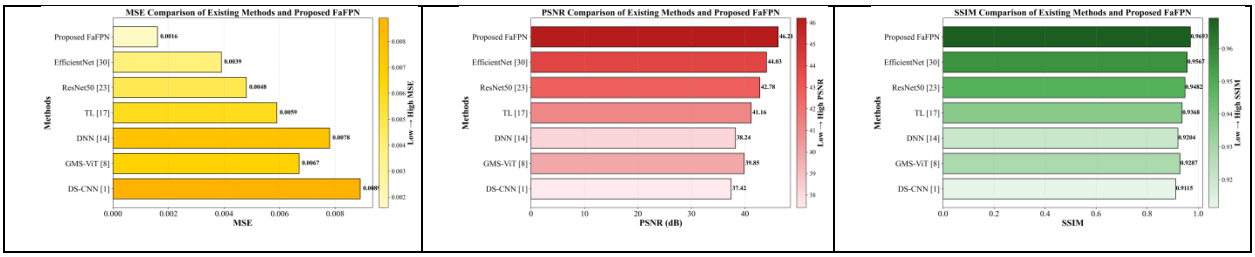


Figure 7: Comparative Denoising Performance Evaluation

The noise reduction capability of the FaFPN algorithm and some other algorithms is shown in Figure 7 using various metrics. It can be observed that the FaFPN algorithm has the lowest value for the metric MSE and highest values for the metrics PSNR and SSIM.

4.3 IAM-FE-Based Facial Region Segmentation Analysis

This section shows the performance of IAM-FE in identifying the diagnostically important regions of the face and facial contours.

a) Segmentation Input Image	b) Key Facial Region Heatmap	c) Segmentation Output Image

Figure 8: Facial Region Segmentation Using the IAM-FE

The figure 8 below depicts the output of the IAM-FE based facial region segmentation process. Here, the input images are transformed into the heat maps that show attention to certain areas of faces like the eyes, nose, mouth, forehead and face margins. This approach leads to improved representation of facial areas through the accurate segmentation output.

Table 5: Attention-Guided Facial Region Refinement Using IAM-FE

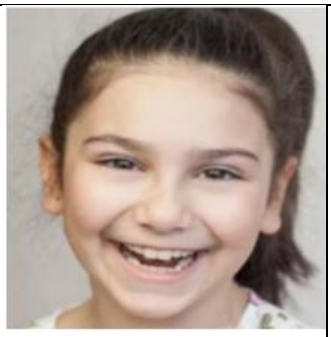
Refined Mask			
IAM-FE Attention Map			

Table 5 shows the face refinement process of IAM-FE using masks that have been refined to create attention maps by highlighting the distinctive regions such as the periocular region, nose region, mouth region and facial boundaries. The attention maps remove background noise and enhance localization of phenotypes associated with Down syndrome.

Table 6: Segmentation Performance Comparison of Existing Methods and Proposed IAM-FE

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
DS-CNN [1]	86.72	86.10	85.84	85.71
Grouped Multi-Scale Vision Transformer (GMS-ViT) [8]	91.63	91.05	90.88	90.74
DNN [14]	87.94	87.21	86.93	86.80
Transfer Learning [17]	90.18	89.65	89.42	89.28
ResNet50 [23]	91.04	90.53	90.31	90.12
EfficientNet [30]	92.46	92.01	91.78	91.62
Proposed IAM-FE	94.00	93.69	93.56	93.43

Table 6 highlights the segmentation performance of various methods and the IAM-FE method proposed by us. The IAM-FE gives an accuracy of 94.00%, a precision of 93.69%, recall of 93.56% and an F1 score of 93.43%, which indicates its excellent ability to localize the facial regions and perform attention-based segmentation.

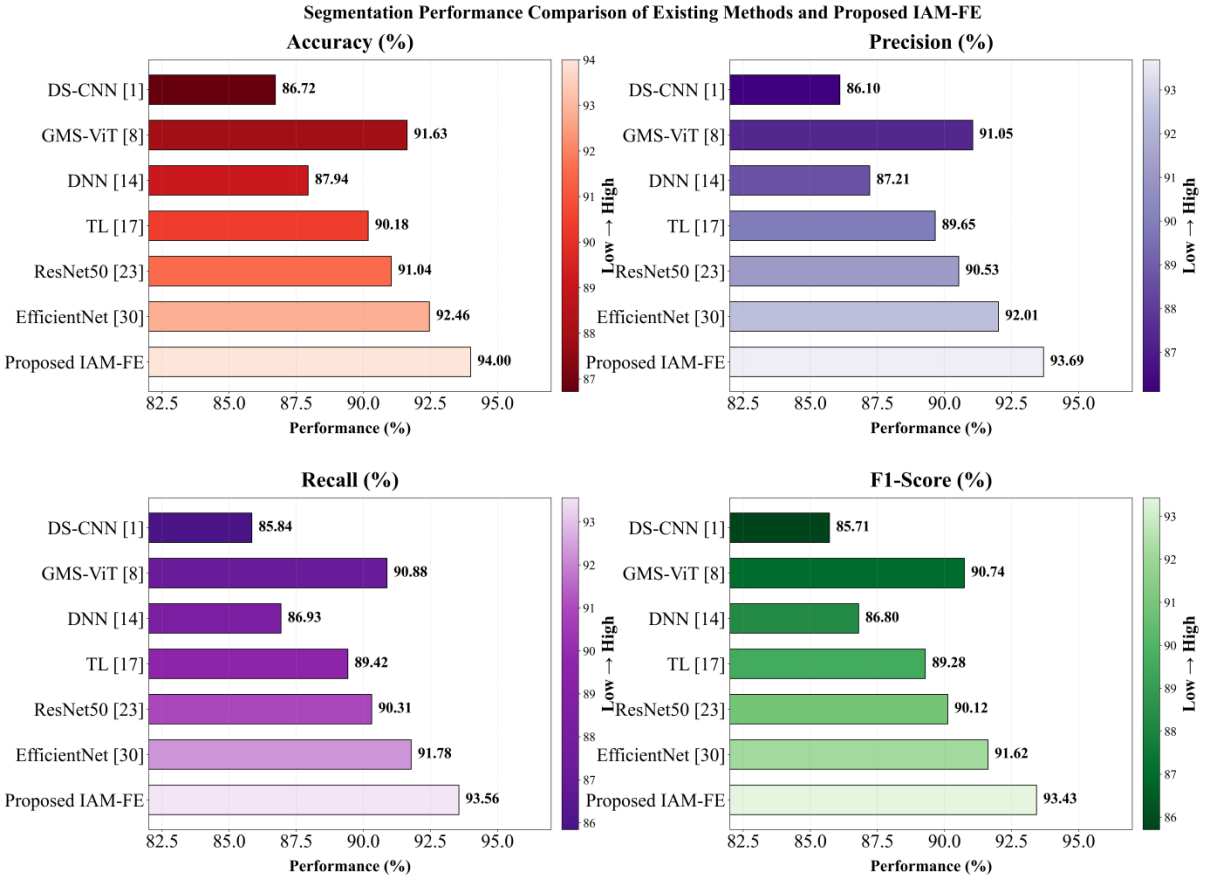
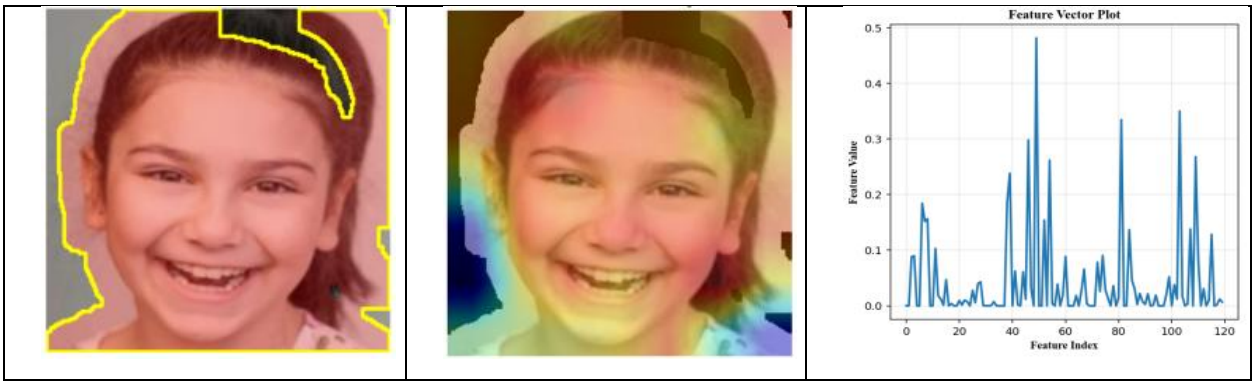


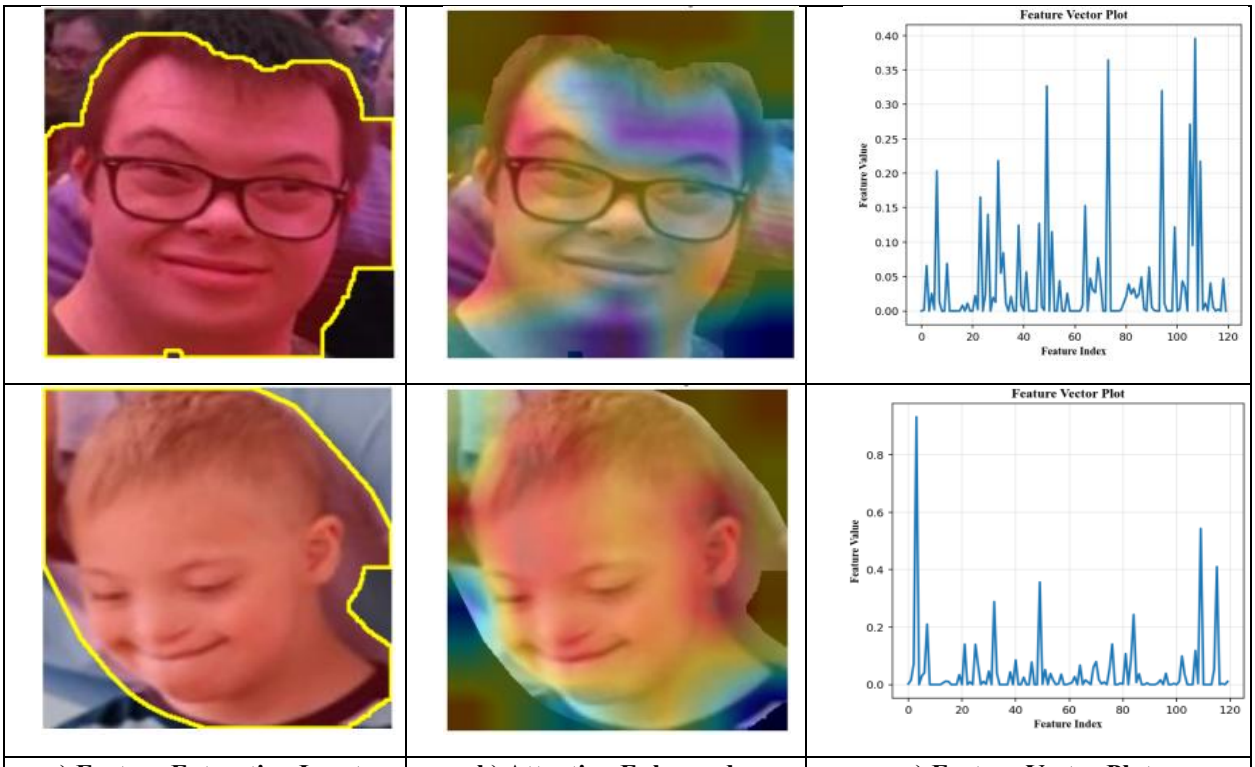
Figure 9: IAM-FE Segmentation Performance Comparison

Figure 9 depicts the performance comparison of the proposed method of segmentation with the existing methods through the use of some metrics. It can be observed that the IAM-FE is far more superior to others in all the parameters, demonstrating its robust nature.

4.4 TDA-E-Based Deep Feature Extraction Evaluation

In this section, discuss the effectiveness of TDA-E in extracting attention-enhanced, high-dimensional facial phenotype features from segmented facial regions.



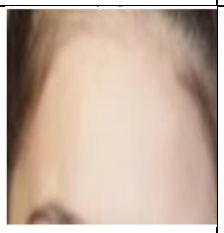
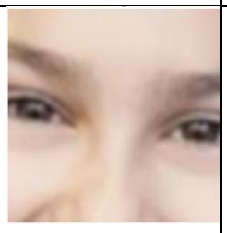
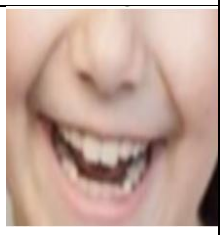
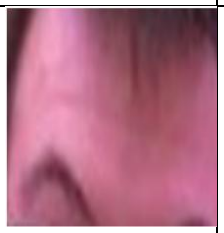
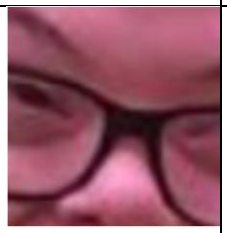


a) Feature Extraction Input Image	b) Attention Enhanced Output	c) Feature Vector Plot

Figure 10: Feature Extraction Using the TDA-E

Figure 10 depicts the TDA-E based feature extraction process, which extracts discriminative phenotypic representations from segmented face regions through transfer learning and attention mechanism. It is clear from the attention-augmented feature vector and feature map that the model learns spatial, contextual and morphological features effectively for accurate classification.

Table 7: Region-Specific Feature Extraction Using the TDA-E

Region-Masked Feature Image	Edge-Aware Feature Map Mask	Eye Region	Nose Region	Mouth Region
				
				

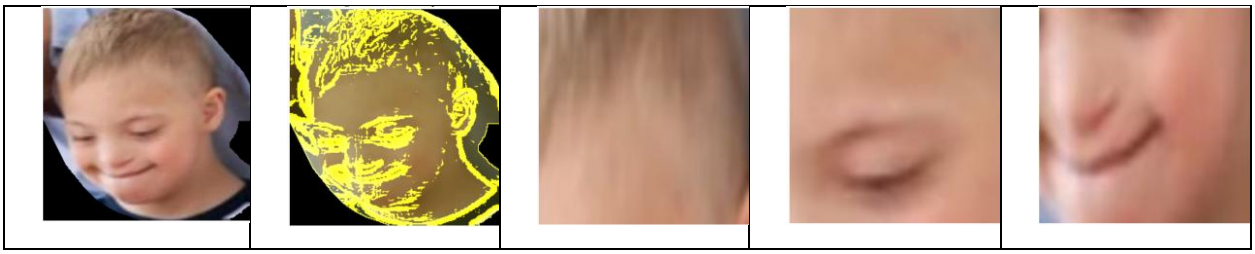


Table 7 depicts the procedure involved in the extraction of region-specific features based on the use of TDA-E, where region-masked facial images are transformed into edge-aware feature maps that highlight key morphological patterns. The independent extraction of the features of the eyes, nose and mouth ensures discrimination of phenotypic features.

Table 8: TDA-E Feature Extraction Performance Comparison

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
DS-CNN [1]	88.26	87.92	87.10	87.50
Grouped Multi-Scale Vision Transformer (GMS-ViT) [8]	92.84	92.15	91.70	91.92
DNN [14]	89.45	88.80	88.35	88.57
Transfer Learning [17]	91.72	91.20	90.76	90.98
ResNet50 [23]	92.63	92.04	91.45	91.74
EfficientNet [30]	93.74	93.12	92.70	92.91
Proposed TDA-E	95.60	94.79	94.03	94.54

Table 8 shows the comparative analysis of the performance of the current approaches and the proposed TDA-E in terms of accuracy, precision, recall and F1 score. The TDA-E method provides the maximum accuracy of 95.60%, precision of 94.79%, recall of 94.03% and F1 score of 94.54% justifying its superior capability in extracting attention-aware and context-dependent features of the facial phenotype.

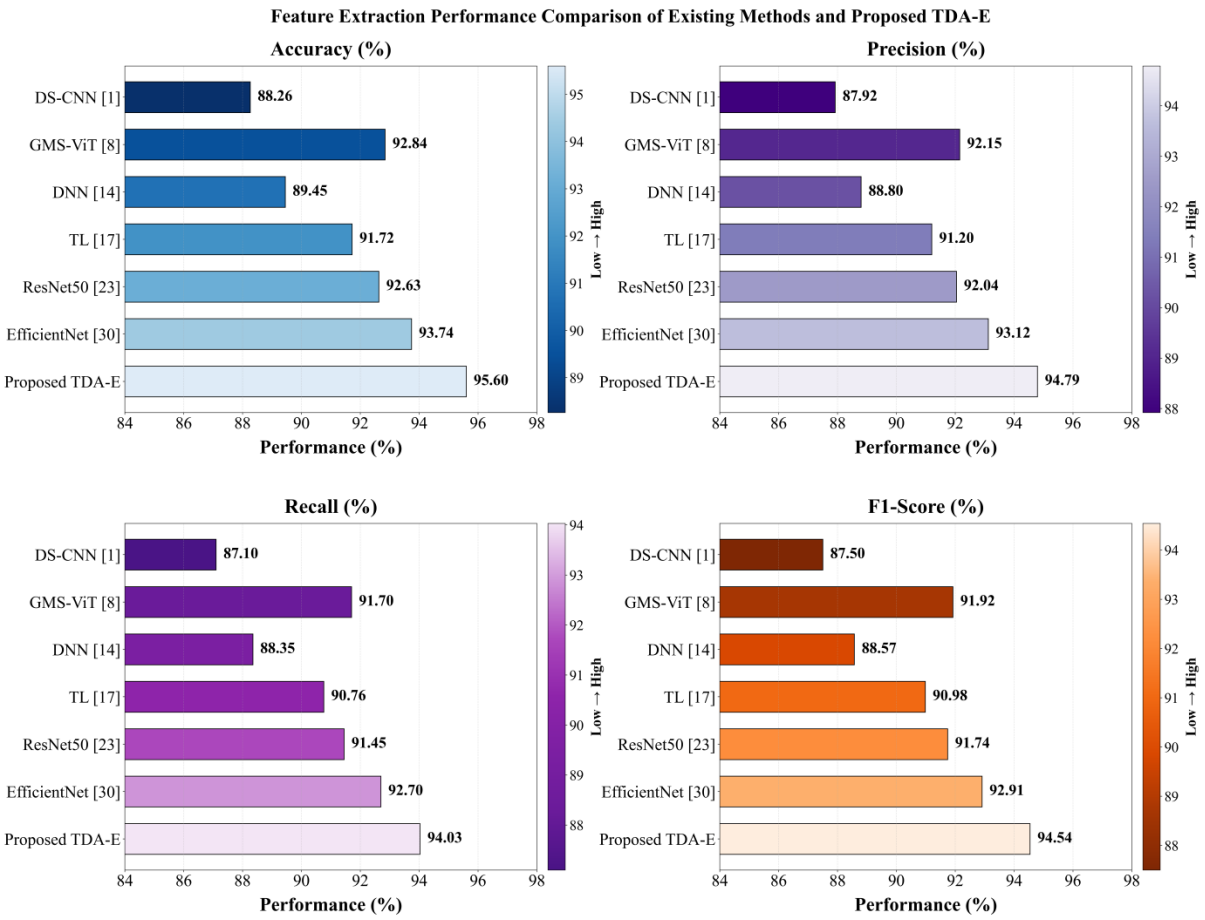


Figure 11: Feature Extraction Performance Comparison of Existing Methods and Proposed TDA-E

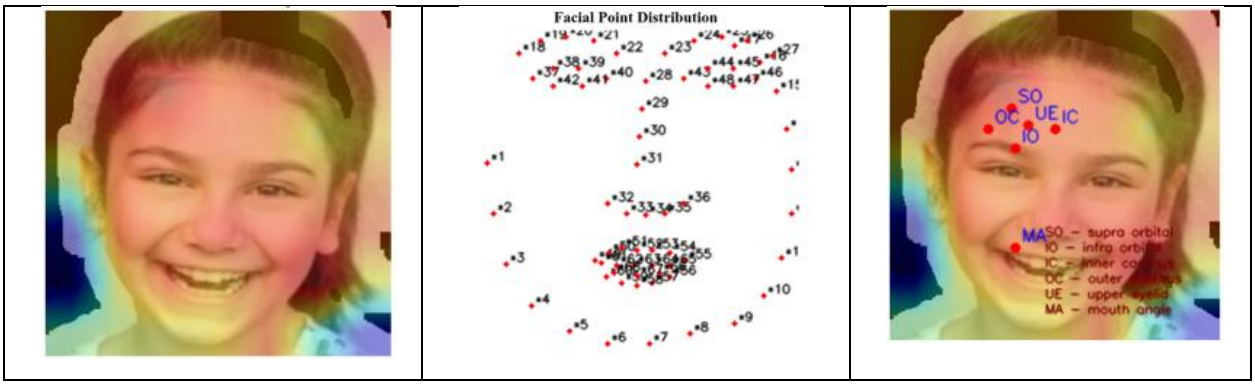
Figure 11 shows the comparison of feature extraction performance between the proposed TDA-E and the existing approaches in terms of metrics. The proposed TDA-E proves to have achieved better performance in all metrics, validating its efficacy in learning discriminative and attention-oriented facial phenotypes.

4.5 NeVRS-Based Classification Performance Analysis

This section analyzes the diagnostic classification performance of NeVRS through metrics and probability of class estimation.

4.5.1 Non-Down Syndrome Classification Output

This section provides the classification results for non-Down syndrome faces and explains how the proposed NeVRS model recognizes healthy face phenotypes with high confidence of diagnosis.



a) NeVRS Classification Input	b) Facial Point Distribution	c) Localized Facial Landmarks

Figure 12: Classification and Facial Landmark Analysis Using the NeVRS

Figure 12 depicts the NeVRS-based classification and facial landmark identification procedure, wherein enhanced attention features of the facial representation are used to identify crucial craniofacial landmarks as well as establish the relative geometric relationships. These landmarks based features provide reliable geometric as well as morphological information about the facial pattern.

Table 9: NeVRS-Based Diagnostic Classification and Decision Generation

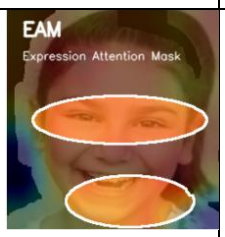
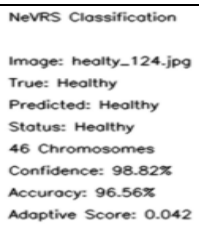
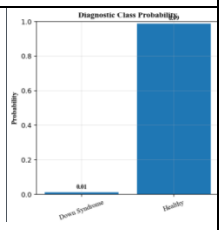
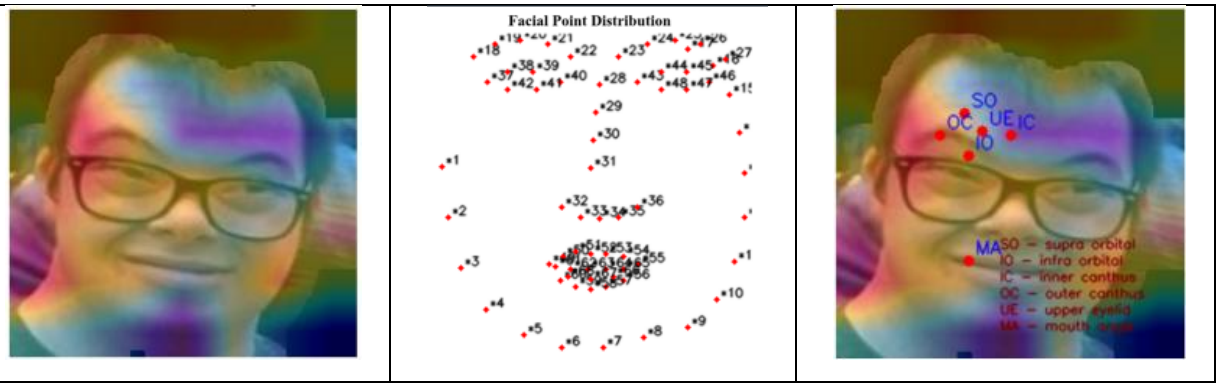
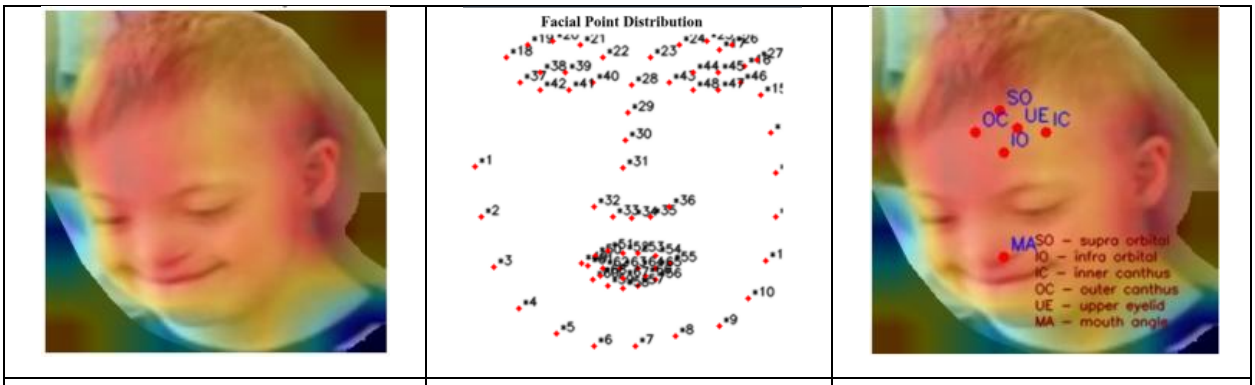
Binary Feature Mask (BFM)	Expression Attention Mask (EAM)	Trisomy Pattern Classification Mask (TPCM)	Healthy / Unhealthy Decision	Diagnostic Class Probability
				

Table 9 demonstrates the NeVRS-based diagnostic classification procedure, in which Binary Feature Mask, Expression Attention Mask and Trisomy Pattern Classification Mask are combined to construct a holistic facial feature. The resultant combination facilitates the healthy/unhealthy decision generation along with diagnostic class probability, thereby providing an awareness of the Down syndrome classification process.

4.5.2Down Syndrome Classification Output

This section discusses the classification results for Down syndrome facial images and the ability of the proposed method to recognize the phenotype variation for syndromes through attention-enhanced learning of features.





a) NeVRS Classification Input	b) Facial Point Distribution	c) Localized Facial Landmarks

Figure 13: Classification and Craniofacial Landmark Analysis Using the NeVRS

Figure 13 shows the procedure for classifying using NeVRS and analyzing the craniofacial landmarks, whereby the attention-focused facial images are used to detect and analyze the distribution of the landmarks on the face. These landmarks contribute towards providing descriptive geometric and morphological features that help determine the phenotype.

Table 10: Multi-Mask Diagnostic Classification Using the NeVRS

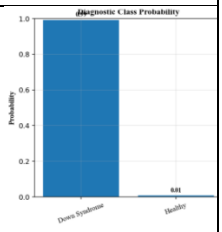
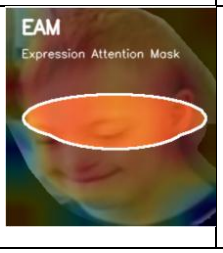
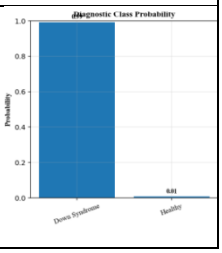
Binary Feature Mask (BFM)	Expression Attention Mask (EAM)	Trisomy Pattern Classification Mask (TPCM)	Healthy / Unhealthy Decision	Diagnostic Class Probability
			NeVRS Classification Image: down_149.jpg True: Down Syndrome Predicted: Down Syndrome Status: Unhealthy 47 Chromosomes Confidence: 99.29% Accuracy: 96.56% Adaptive Score: 0.037	
			NeVRS Classification Image: down_745.jpg True: Down Syndrome Predicted: Down Syndrome Status: Unhealthy 47 Chromosomes Confidence: 99.29% Accuracy: 96.56% Adaptive Score: 0.042	

Table 10 describes the process of creating a combined multi-mask based diagnosis classification using the NeVRS algorithm, where Binary Feature Mask, Expression Attention Mask and Trisomy Pattern Classification Mask are fused to create the facial image representation. The mask feature values give the diagnostic class probability and the ultimate decision for being healthy or unhealthy.

Table 11: Classification Performance Comparison of Existing Methods and Proposed NeVRS

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
DS-CNN [1]	89.18	88.64	88.20	88.42

Grouped Multi-Scale Vision Transformer (GMS-ViT) [8]	93.41	92.86	92.47	92.66
DNN [14]	90.36	89.75	89.31	89.53
Transfer Learning [17]	92.28	91.74	91.36	91.55
ResNet50 [23]	93.16	92.58	92.20	92.39
EfficientNet [30]	94.62	94.08	93.75	93.91
Proposed NeVRS	96.56	95.72	95.62	95.67

Table 11 below shows a comparative analysis of the classification effectiveness of current DL models, transformer networks and transfer learning approaches, compared with the proposed NeVRS algorithm. The comparisons reveal the capability of NeVRS to improve the accuracy of classification using an adaptive neural mapping and multiresolution feature extraction approach.

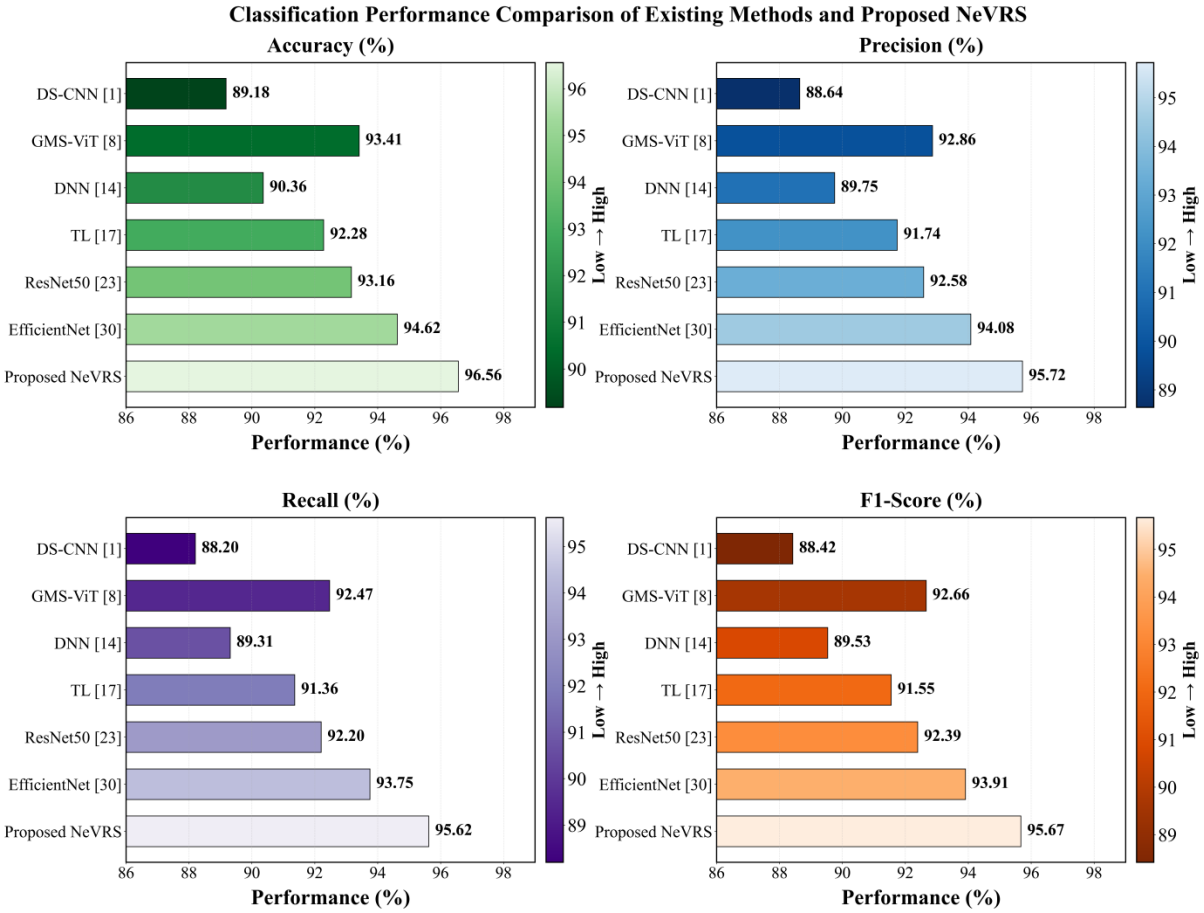


Figure 14: Classification Performance Comparison of Existing Methods and Proposed NeVRS

Figure 14 shows the comparison of the classification performance of the proposed NeVRS method and the existing approach using the metrics. It is evident that the proposed NeVRS outperforms in all metrics, thereby indicating its effectiveness in reducing errors in classification.

4.6 Ablation Study

In this section, the individual contributions of the four modules mentioned are evaluated by excluding each one individually and analyzing its effect on overall classification performance.

Table 12: Ablation Study of the Proposed FaFPN–IAM-FE–TDA-E–NeVRS Framework

Model Variant	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)
Without FaFPN Denoising	91.24	90.86	90.42	90.64	92.10
Without IAM-FE Segmentation	92.38	91.95	91.63	91.79	93.27
Without TDA-E Attention Feature Extraction	93.46	93.02	92.71	92.86	94.15
Without NeVRS Multi-Resolution Classification	94.21	93.84	93.56	93.70	95.04
Proposed FaFPN–IAM-FE–TDA-E–NeVRS	96.72	96.38	96.11	96.24	97.85

Table 12 reveals how the ablation study provides an assessment of the effect of each individual module, including FaFPN, IAM-FE, TDA-E and NeVRS, by eliminating one of the modules and examining the performance in comparison to the entire model. It is observed that the full-fledged framework of FaFPN-IAM-FE-TDA-E-NeVRS framework performs better.

4.7 Statistical Analysis

In this section, analyze the statistical stability and reliability of the proposed framework through mean performance, standard deviation, confidence interval and significance level analysis.

Table 13: Statistical Analysis of the Proposed FaFPN–IAM-FE–TDA-E–NeVRS Framework

Model Variant	Accuracy (%) Mean ± SD	Precision (%) Mean ± SD	Recall (%) Mean ± SD	F1-Score (%) Mean ± SD	95% CI for Accuracy	p-value
Without FaFPN Denoising	91.24 ± 0.48	90.86 ± 0.52	90.42 ± 0.57	90.64 ± 0.54	90.82–91.66	<0.01
Without IAM-FE Segmentation	92.38 ± 0.44	91.95 ± 0.49	91.63 ± 0.51	91.79 ± 0.46	91.99–92.77	<0.01
Without TDA-E Feature Extraction	93.46 ± 0.39	93.02 ± 0.43	92.71 ± 0.45	92.86 ± 0.41	93.12–93.80	<0.01
Without NeVRS Classification	94.21 ± 0.36	93.84 ± 0.40	93.56 ± 0.42	93.70 ± 0.38	93.89–94.53	<0.05
Proposed FaFPN–IAM-FE–TDA-E–NeVRS	96.72 ± 0.28	96.38 ± 0.31	96.11 ± 0.34	96.24 ± 0.32	96.47–96.97	—

Table 13, the statistical evaluation tests the stability and robustness of the proposed framework based on mean, standard deviation, confidence interval and significance level comparison of the results from different experiments. The low variation and statistical significance prove that the proposed framework gives stable and reliable performance for the diagnosis of Down syndrome.

4.8 Limitations of the Proposed Framework

- FaFPN could exhibit lower performance in noise removal in case of highly unbalanced lighting conditions, blur, or low resolution.
- The IAM-FE segmentation algorithm could perform poorly in cases of extensive occlusions, significant variation or partial visibility of critical regions of the face.
- TDA-E relies on transfer-learned backbone representations, which means that it might be sensitive to the differences between the pre-trained and target domains.
- NeVRS could have high computational requirements because of multi-resolution feature learning and requires large datasets to generalize better.

5. Conclusion

The research introduced the FaFPN-IAM-FE-TDA-E-NeVRS framework for detecting Down syndrome using analyzing facial expressions and facial phenotype analysis. This research framework included adaptive denoising, attention-guided facial regions segmentation, deep transfer learning-based features extraction and multi-resolution classification. In particular, FaFPN kept useful information about textures and edges, IAM-FE was responsible for locating important regions of faces, TDA-E created attention enhanced phenotype embeddings and NeVRS generated the final output for classification. The results of experiments proved the performance of the proposed framework in terms of an accuracy of 96.72%. The results of ablation and statistical analyses demonstrated the contribution of every module in increasing robustness, stability and reproducibility. Thus, this framework can be used for automatic and non-invasive screening support for assessing Down syndrome facial phenotypes. Future research work will focus on validation of the model on large multi-center clinical data. Moreover, future studies will integrate explainable AI into this framework and real-time screening deployment.

Reference

1. Moreno Escobar, J. J., Morales Matamoros, O., Aguilar del Villar, E. Y., Quintana Espinosa, H., & Chanona Hernández, L. (2023, August). DS-CNN: deep convolutional neural networks for facial emotion detection in children with down syndrome during Dolphin-Assisted therapy. In Healthcare (Vol. 11, No. 16, p. 2295). MDPI.https://doi.org/10.3390/healthcare11162295
2. Sherif, Fayroz F., Tawfik, Nahed, Mousa, Doaa, Abdallah, Mohamed S., & Cho, Young-Im. (2024). Automated Multi-Class Facial Syndrome Classification Using Transfer Learning Techniques. Bioengineering, 11(8), 827. https://doi.org/10.3390/bioengineering11080827
3. Kumaran, Yogesh S., Jospin Jeya, J., Mahesh, T. R., Bhatia Khan, Surbhi, Alzahrani, Saeed, & Alojail, Mohammed. (2024). Explainable lung cancer classification with ensemble transfer learning of VGG16, Resnet50 and InceptionV3 using grad-cam. BMC Medical Imaging, 24(1), 195. https://doi.org/10.1186/s12880-024-01345-x
4. An, Fengping, Li, Xiaowei, & Ma, Xingmin. (2021). Medical Image Classification Algorithm Based on Visual Attention Mechanism-MCNN. Mathematical Problems in Engineering, 2021, 6280690. https://doi.org/10.1155/2021/6280690
5. Kim, Taehyung, Mok, Jiwon, & Lee, Euichul. (2021). Detecting Facial Region and Landmarks at Once via Deep Network. Sensors, 21(16), 5360. https://doi.org/10.3390/s21165360
6. Geremek, Marek, & Szklanny, Krzysztof. (2021). Deep learning-based analysis of face images as a screening tool for genetic syndromes. Sensors, 21(19), 6595. https://doi.org/10.3390/s21195595
7. Zhang, J., Li, F., Zhang, X., Wang, H., & Hei, X. (2024). Automatic medical image segmentation with vision transformer. Applied Sciences, 14(7), 2741.https://doi.org/10.3390/app14072741
8. Ji, Z., Chen, Z., & Ma, X. (2025). Grouped multi-scale vision transformer for medical image segmentation. Scientific Reports, 15(1), 11122. https://doi.org/10.1038/s41598-025-95361-8
9. Khan, A. R., & Khan, A. (2025). Multi-axis vision transformer for medical image segmentation. Engineering Applications of Artificial Intelligence, 158, 111251. https://doi.org/10.1016/j.engappai.2025.111251
10. Ferdous, M., Mahmud, S., & Shimul, M. E. Z. (2025). MedNet: a lightweight attention-augmented CNN for medical image classification. Scientific Reports, 15(1), 41936. https://doi.org/10.1038/s41598-025-25857-w
11. Zubair, M., Owais, M., Hassan, T., Bendechache, M., Hussain, M., Hussain, I., & Werghi, N. (2025). An interpretable framework for gastric cancer classification using multi-channel attention mechanisms and transfer learning approach on histopathology images. Scientific Reports, 15(1), 13087.https://doi.org/10.1038/s41598-025-97256-0

12. Muksimova, Shaxnoza, Umirzakova, Surayo, Iskhakova, Nozima, Khaitov, Alisher, & Cho, Young-Im. (2025). Advanced convolutional neural network with attention mechanism for Alzheimer's disease classification using MRI. Computers in Biology and Medicine, 190, 110095. https://doi.org/10.1016/j.compbiomed.2025.110095
13. Allogmani, A. S., Mohamed, R. M., Al-Shibly, N. M., & Ragab, M. (2024). Enhanced cervical precancerous lesions detection and classification using Archimedes Optimization Algorithm with transfer learning. Scientific Reports, 14(1), 12076. https://doi.org/10.1038/s41598-024-62773-x
14. Salehi, A., & Khedmati, M. (2025). Hybrid clustering strategies for effective oversampling and undersampling in multiclass classification. Scientific Reports, 15(1), 3460. https://doi.org/10.1038/s41598-024-84786-2
15. Malik, P., Singh, J., Ali, F., Sehra, S. S., & Kwak, D. (2025). Action unit based micro-expression recognition framework for driver emotional state detection. Scientific Reports, 15(1), 27824.https://doi.org/10.1038/s41598-025-12245-7
16. Kroczek, L. O., Lingnau, A., Schwind, V., Wolff, C., & Mühlberger, A. (2024). Observers predict actions from facial emotional expressions during real-time social interactions. Behavioural Brain Research, 471, 115126.https://doi.org/10.1016/j.bbr.2024.115126
17. Shaikh, M. A., Al-Rawashdeh, H. S., & Sait, A. R. W. (2025). Deep Learning-Powered Down Syndrome Detection Using Facial Images. Life, 15(9), 1361.https://doi.org/10.3390/life15091361
18. Shi, L. (2025). Enhancing medical explainability in deep learning for age-related macular degeneration diagnosis. Scientific Reports, 15(1), 16975.https://doi.org/10.1038/s41598-025-01496-z
19. Zhou, S., Wang, J., Xu, Z., Wang, S., Brauer, D., Welton, L., ... & Zhang, R. (2025). Uncertainty-aware large language models for explainable disease diagnosis. npj Digital Medicine, 8(1), 690.https://doi.org/10.1038/s41746-025-02071-6
20. Liu, H., Mo, Z. H., Yang, H., Zhang, Z. F., Hong, D., Wen, L., ... & Wang, S. S. (2021). Automatic facial recognition of Williams-Beuren syndrome based on deep convolutional neural networks. Frontiers in Pediatrics, 9, 648255.https://doi.org/10.3389/fped.2021.648255
21. Rajawat, A. S., Bedi, P., Goyal, S. B., Bhaladhare, P., Aggarwal, A., & Singhal, R. S. (2023). Fusion fuzzy logic and deep learning for depression detection using facial expressions. Procedia Computer Science, 218, 2795-2805.https://doi.org/10.1016/j.procs.2023.01.251
22. Shahzad, T., Iqbal, K., Khan, M. A., & Iqbal, N. (2023). Role of zoning in facial expression using deep learning. IEEE Access, 11, 16493-16508.DOI: 10.1109/ACCESS.2023.3243850
23. Bisogni, C., Cimmino, L., De Marsico, M., Hao, F., & Narducci, F. (2023). Emotion recognition at a distance: The robustness of machine learning based on hand-crafted facial features vs deep learning models. Image and vision computing, 136, 104724.https://doi.org/10.1016/j.imavis.2023.104724
24. Islam, M. M., Nooruddin, S., Karray, F., & Muhammad, G. (2024). Enhanced multimodal emotion recognition in healthcare analytics: A deep learning based model-level fusion approach. Biomedical Signal Processing and Control, 94, 106241.https://doi.org/10.1016/j.bspc.2024.106241
25. Essahraui, S., Lamaakal, I., El Hamly, I., Maleh, Y., Ouahbi, I., El Makkaoui, K., ... & Abd El-Latif, A. A. (2025). Real-time driver drowsiness detection using facial analysis and machine learning techniques. sensors, 25(3), 812.https://doi.org/10.3390/s25030812
26. Simić, N., Suzić, S., Milošević, N., Stanojev, V., Nosek, T., Popović, B., & Bajović, D. (2024). Enhancing emotion recognition through federated learning: A multimodal approach with convolutional neural networks. Applied sciences, 14(4), 1325.https://doi.org/10.3390/app14041325
27. Mamieva, D., Abdusalomov, A. B., Kutlimuratov, A., Muminov, B., & Whangbo, T. K. (2023). Multimodal emotion detection via attention-based fusion of extracted facial and speech features. Sensors, 23(12), 5475.https://doi.org/10.3390/s23125475
28. Mutawa, A. M., & Hassouneh, A. (2024). Multimodal real-time patient emotion recognition system using facial expressions and brain EEG signals based on machine learning and log-sync methods. Biomedical Signal Processing and Control, 91, 105942.https://doi.org/10.1016/j.bspc.2023.105942
29. Das, S., Pratihar, S., Pradhan, B., Jhaveri, R. H., & Benedetto, F. (2024). IoT-assisted automatic driver drowsiness detection through facial movement analysis using deep learning and a U-Net-based architecture. Information, 15(1), 30.https://doi.org/10.3390/info15010030
30. Gaya-Morey, F. X., Ramis, S., Buades-Rubio, J. M., & Manresa-Yee, C. (2026). Assessing the efficacy of deep learning approaches for facial expression recognition in individuals with intellectual disabilities. Multimedia Systems, 32(2), 156.https://doi.org/10.1007/s00530-026-02233-w