Journal of Trends in Computer Science and Smart Technology (ISSN: 2582-4104)
www.irojournals.com/tcsst/

Confidence-Weighted Ensemble Learning and Graph-Based Shapley Theory for Refined Emotion Prediction in Autistic Children

Sujatha Hanumantharayappa[1], Manjula R. Bharamagoudra[2]

School of Electronics and Communication Engineering, REVA University, Bengaluru, India.

E-mail: [1]sujag12@gmail.com, [2]manjula.rb@reva.edu.in

Orcid ID: [1]0009-0008-9949-275X, [2]0000-0002-6277-9385

Abstract

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition characterized by variations in brain function and structure that affect emotion perception, communication, and social interaction. Traditional machine learning (ML) algorithms suffer from bias-variance imbalance and redundant feature representations, leading to inaccurate emotion classification in autistic children due to heterogeneous facial cues and limited annotated data. To address these challenges, this research proposes a Mutual-Information-weighted Shapley Graph (MISG) and Confidence Weighted Aggregation with Ensemble ML (CWAEML) for emotion classification in autistic and non-autistic facial datasets. MISG performs structured graph-based feature refinement by modelling inter-feature dependencies using mutual information and estimating marginal contribution via Shapley theory. Deterministic feature selection removes semantic redundancy while preserving discriminative attributes, thereby improving classifier robustness. The EML integrates multiple base learners, including Multi-Support Vector Machine (MSVM), Naïve Bayes (NB), K-Nearest Neighbour (KNN), Decision Tree (DT), and Random Forest (RF), through confidence-weighted fusion to ensure adaptive learning across emotion domains. The proposed MISG-CWAEML achieves accuracies of 98.75%, 97.95%, and 68.72% on AffectNet, Autistic Children Emotions - Dr. Fatma M. Talaat, and Young AffectNet HQ, respectively, which are superior to conventional SVM and other benchmark models in cross-domain evaluations. The findings demonstrate the effectiveness of graph-based feature selection combined with adaptive ensemble learning in handling complex and imbalanced emotion datasets. The proposed MISG-CWAEML provides practical implications for developing reliable systems, thereby enabling enhanced emotion recognition for children with ASD.

Keywords: Ensemble Machine Learning, Emotion Prediction, Feature Refinement, Shapley Graph Theory, Autism Children.

1. Introduction

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition that affects individuals worldwide, with an estimated prevalence of 1 in 270 people. However, the

DOI: https://doi.org/10.36548/jtcsst.2026.3.016
Received: 13.05.2026, received in revised form: 16.06.2026, accepted: 08.07.2026, published: 20.07.2026

early diagnosis of ASD is limited due to the absence of definite pathological biosignatures and considerable healthcare inequities in underserved areas [1]. Emotion recognition systems are essential for supporting children with ASD who also have issues with Theory of Interaction (TOI), still current deep learning algorithms do not perform well on emotion classification tasks for faces with ASD because of high individual variability, noise, and feature overload, which fails to account for the important synchronized facial motion between features [2]. It is important to tackle internal components such as susceptibility, reasoning, self-efficacy, and ambition to enhance emotional understanding in ASD [3]. For affective computing, 7 basic classes of emotions are preferred (happiness, sadness, surprise, fear, disgust, anger, and contempt), which are globally recognized. However, boundaries between discrete classes are not always sharp and lead to information loss [4, 5]. Children with autism express emotions thorugh subtle facial cues, which differ significantly from typical expressions [6-8]. Because of these variations, it is important to analyse subtle changes in facial features to understand an individual's emotional state [9, 10]. This research focuses on facial image analysis as facial expressions are the main channel for non-verbal communication [11, 12]. A challenge in creating a tracking system is the high individuality in facial structures [13], which leads to redundant data during feature extraction [14]. By refining these facial feature inputs and focusing on how specific areas of the face interact, ML models can achieve stable classification of emotions in children with ASD [15]. To overcome these limitations and move towards a standardized diagnostic procedure, the present research suggests a computational framework based on three major innovations:

- An advanced feature extraction system fine-tuned to ensure high detail consistency when dealing with various patient structures.
- The novel structural feature selection framework, called the Mutual-Information-weighted Shapley Graph (MISG), estimates the marginal contribution and interaction of facial features to reduce their redundancy without falling into local optima.
- Ensemble Classification (CWAEML) which employs weighted probability fusion across multiple classifiers to increase the reliability of the decisions and their generalization. The proposed model transforms clinical observations into a computational approach for ASD emotion analysis.

2. Literature Review

The increasing complexity of non-typical facial expressions has prompted the development of emotion recognition systems for Autism Spectrum Disorder (ASD) using deep learning (DL). Recent methodologies are discussed, and potential areas of computational shortcoming are highlighted.

Talaat et al. [16] developed a real-time emotion recognition model based on kernel autoencoders and pre-trained deep Convolutional Network (CNN) models like ResNet, Xception, and MobileNet. To deal with data size restrictions, extraction and refinement were achieved using the autoencoder. The use of the model was, however, limited as it was observed that this model was unable to tackle class imbalance well during the classification phase, leading to an imbalanced reliability of the diagnostic outputs. Saranya and Anandan [17] suggested a Facial Action Coding System (FACS) combined with a hybrid DL model for autism detection. The FACS-trained Convolutional Neural Network (CNN) was developed to improve ASD detection rates among children. DEEPFACENET was applied to extract features, then the FACS-CNN and hybrid Long Short-Term Memory (LSTM) were implemented for classification. Savechenko et al. [18] suggested emotion classification and online learning engagement using a neural network for recognizing a single facial expression. The pre-trained model was developed with robust optimization techniques and was applied to video processing on mobile devices without requiring remote server support. Hosney et al. [19] proposed an attention-based YOLOv8 model that uses a Deep Neural Network (DNN) to quickly classify affective states. The researchers improved the model by adding an attention mechanism, which allows it to pay attention to key facial features while ignoring irrelevant background information. Despite these advancements, no specific feature optimization phase was included in the framework, so redundant data is processed, which reduces overall classification accuracy.

Talaat [20] developed an emotion recognition system that used the U-Net network for segmenting facial images, followed by a CNN for classification. A Genetic Algorithm (GA) was used to adjust the network's hyperparameters and increase its accuracy to optimize performance. However, the study revealed that the model's effectiveness was compromised in various clinical settings due to instances of misclassification caused by the evolution of the gradient during the training process. Cheng et al. [21] introduced a computer-aided ASD diagnosis model with behavior signal processing. The standardized platform enabled the gathering, simulation, modeling, interpretation and analysis of human behavioral data for computer-aided ASD diagnosis. The structured evaluation procedure automatically estimated multiple aspects of children's social interaction by capturing audio-visual data and generating final diagnostic outputs. However, noise created while training the model caused overfitting issues and reduced classification accuracy. Ciraolo et al. [22] presented a facial expression recognition system based on emotional AI for telerehabilitation. The model incorporated cloud/edge telerehabilitation, where the cloud interacted with remote rehabilitation by monitoring edge devices. A face mesh model generated facial mesh data from input images, and features extracted from this data were then used to train a classifier for recognizing various facial expressions.

Overall, the above-analysed review reveals that existing research exhibits significant improvements in emotion recognition in autistic children, but most rely on fixed CNN or regression-based designs, which do not consider multi-feature correlation and dynamically weighting classifiers. Class imbalance, noisy inputs, and poor choice of features are additional challenges that limit generalization and stability using different datasets. These restrictions emphasize the need for a dynamic and contextual framework. In this regard, the suggested MISG-CWAEML model presents a two-stage architecture that combines structured graph-based feature refinement and the context-weighted adaptive ensemble learning model, which ensures refined feature representation, balanced adaptive classifier, and enhanced accuracy and robustness to recognize emotion in ASD datasets.

3. Methodology

In this research, Mutual-Information-weighted Shapley Graph and Confidence Weighted Aggregation with Ensemble ML (MISG-CWAEML) is proposed for effective emotion classification in children with ASD. Initially, the facial images are pre-processed through min-max normalization, resizing, and data augmentation to ensure uniformity and address class imbalance. Then, deep features are extracted through ResNet50 and

EfficientNetB0 by capturing both spatial and semantic information. After that, the features are refined by eliminating redundant and non-informative attributes to improve feature quality and reduce computational complexity. Finally, emotion recognition is performed using an ensemble learning model that combines multiple classifiers such as MSVM, Navie Bayes, KNN, DT and RF through confidence-based aggregation for adaptive decision-making. The proposed MISG-CWAEML introduces a dual-phase architecture, which differs from existing emotion classification by incorporating structured graph-based feature selection with adaptive ensemble learning. Compared to traditional DL models that rely on higher-dimensional features without focusing on inter-feature relationships, the proposed model captures feature dependencies by mutual information and estimates feature importance by Shapley values. Moreover, the proposed confidence-weighted aggregation dynamically allocates weights to classifiers based on prediction reliability, thereby enhancing robustness and generalization across heterogeneous datasets. The overall explanation of the proposed methodology is illustrated in Figure 1.

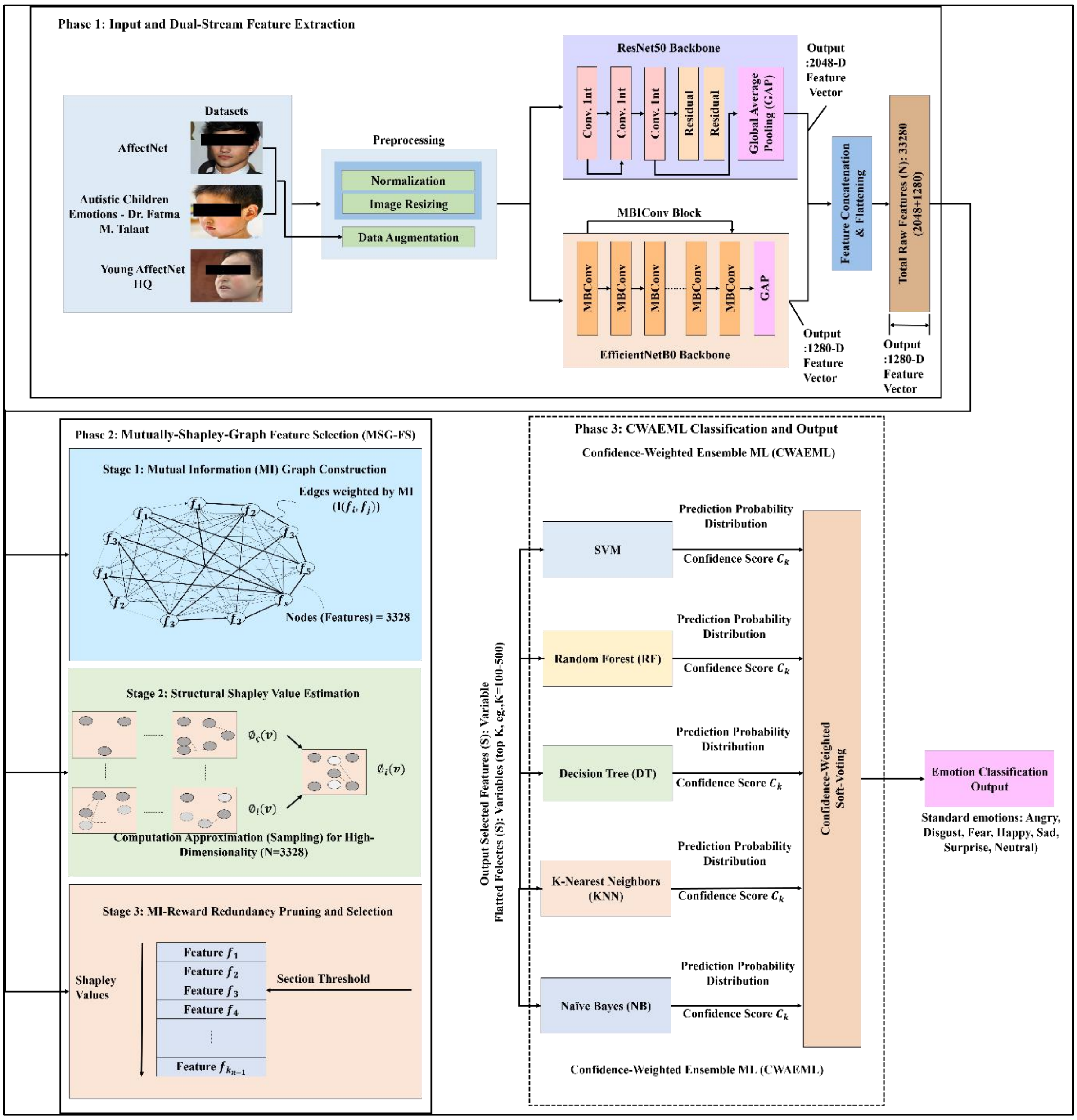


Figure 1. Workflow of the proposed MISG-CWAEML for emotion classification in autistic children

3.1 Dataset Acquisition

The study utilizes three major datasets that represent the wide range of emotions of autistic children: AffectNet [23], Autistic Children Emotions - Dr. Fatma M. Talaat [24] and Young AffectNet HQ [25]. The original data are combined into specific classification schemes and Figure 2 shows example images from the AffectNet, Dr. Fatma M. Talaat, and Young AffectNet HQ datasets, respectively, to demonstrate the visual variation and quality of the data that were used.

Figure 2. Sample facial emotion images from the 3 datasets illustrating varied facial expressions

To prevent train-test data leakage, the original datasets are partitioned into stratified 5-fold cross-validation subsets before applying augmentation. In each fold, 80% of the samples are used for training and 20% for testing. Data augmentation is performed only on the training folds, while the testing samples remain unseen during model development. Feature extraction, MISG-based feature selection, and classifier training are conducted independently within each training fold and evaluated on the respective test fold thereby ensuring unbiased performance and reducing overfitting. The datasets considered in this research are publicly available and inherently contain variations in ethnicity, age distribution, gender representation, clinical labeling procedures, and image acquisition conditions. These factors introduce demographic and domain-specific biases that influence emotion recognition performance. Since detailed demographic data are not available across all datasets, a comprehensive subgroup analysis is not presented in this research. The considered datasets differ in terms of sample size and emotion categories. Therefore, each dataset is processed and evaluated independently using its original class labels and distribution, ensuring consistency with the source datasets and enabling a fair assessment of the proposed framework across heterogeneous data settings.

3.2 Preprocessing

The pre-processing stage includes cleaning the AffectNet, Autistic Children Emotions, and Young AffectNet HQ datasets to remove noise and artifacts. All pixel values are then normalized using min-max normalization to ensure numerical stability within the range of [0, 1] as shown in Equation (1).

$$Y_i = \frac{[X_i - min(X)]}{[max(X) - min(X)]} \tag{1}$$

Here, X_i is a data point, min and max are the minimum and maximum values, respectively and Y_i is the converted result. Each image is then resized to a fixed 224*224*3, consistent with the network's requirements. Lastly, the training data is augmented with various techniques, such as mirroring and rotations, to enhance the model's ability to generalize across real-world scenarios. Augmentation is performed during training, generating multiple transformed samples from the original images while maintaining unchanged test sets. The distribution of original and augmented data is presented in Tables 1 and 2. Augmented sample images from these datasets are given in Figure 3.

Table 1. Dataset description for AffectNet and young AffectNet HQ dataset

Classes	Original data for AffectNet dataset	Original data for Young AffectNet HQ dataset	Augmented data for AffectNet dataset
Anger	1500	1822	4500
Contempt	1559	1833	4677
Disgust	1229	1740	3687
Fear	1512	1839	4536
Happy	2340	1862	7020
Neutral	2758	1880	8274
Sad	3091	1821	9273

Table 2. Dataset details and characteristics of Autistic Children Emotions - Dr. Fatma M. Talaat dataset

Classes	Original data	Augmented data
Anger	67	201
Fear	30	90
Joy	50	150
Neutral	48	144
Sadness	90	270
Surprise	63	189

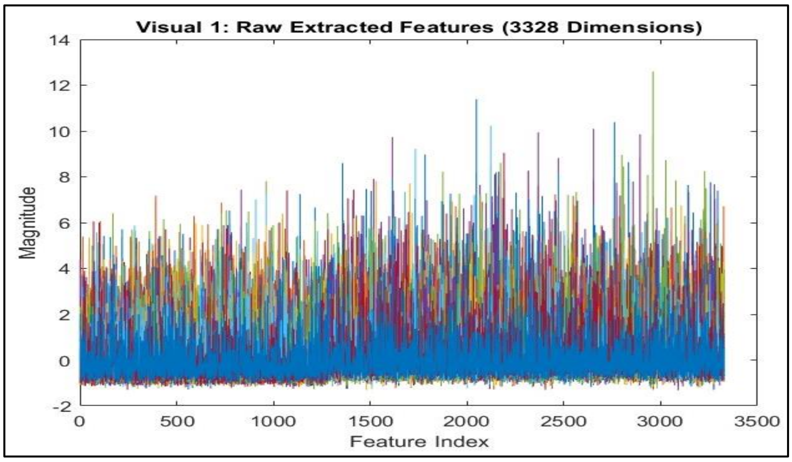
Figure 3. Augmented facial emotion samples demonstrating illumination variations

To ensure experimental validity, preprocessing, including normalization and resizing, is applied independently after dataset partitioning. Data augmentation is limited to the training folds only and is not applied to testing samples. This strategy prevents preprocessing leakage, augmented sample overlap, and information transfer between training and testing sets, thereby ensuring unbiased model evaluation.

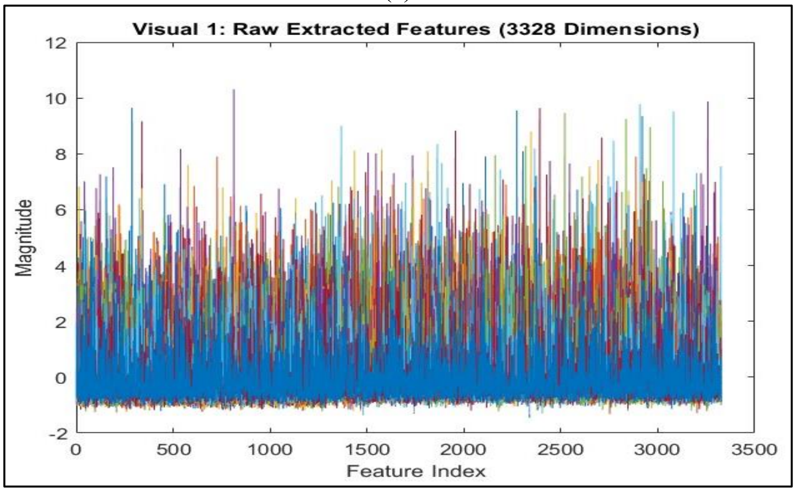
3.3 Feature Extraction

This study adopts a dual-architecture strategy with ResNet50 and EfficientNetB0 to extract a variety of emotional characteristics. ResNet50 captures macro-level spatial expressions and facial structures, while EfficientNetB0 detects micro-level expressions and movements. These models are combined to balance between the two types of facial movement: large-scale and fine-scale muscle movement. The number of extracted features is illustrated in Figure 4 and Table 3, respectively. The concatenation of both networks then produces a rich 3328-dimensional feature vector per image, as it is expressed in Equation (2).

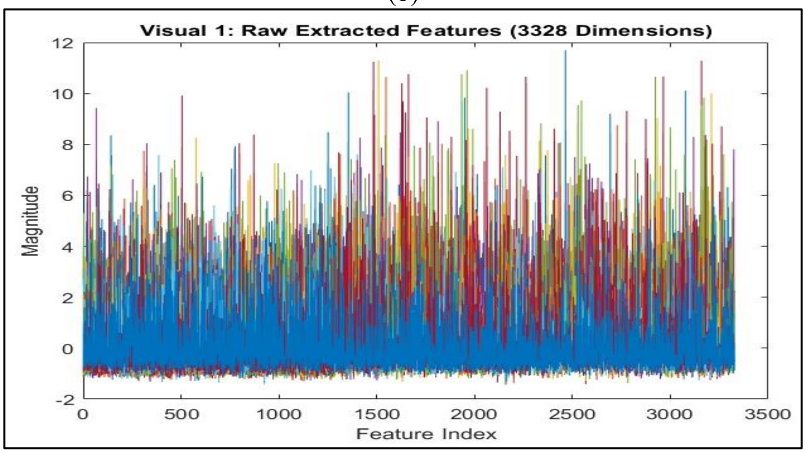
$$F = [F_{ResNet50} || F_{EfficientNetB0}] \tag{2}$$



(a)



(b)



(c)

Figure 4. Extracted features from (a) AffectNet, (b) Autistic Children Emotions – Dr. Fatma M Talaat, and (c) Young AffectNet HQ datasets

Table 3. Extracted features for various datasets

Datasets	ResNet50	EfficientNetB0	Total
AffectNet	2048	1280	3328
Autistic Children Emotions - Dr. Fatma M. Talaat	2048	1280	3328
Young AffectNet HQ	2048	1280	3328

3.4 Feature Selection Using MISG

In high-dimensional DL frameworks, feature selection plays an important role, mainly in recognizing emotional tasks. Redundant and correlated features increase complexity and reduce generalization. Moreover, traditional filter and wrapper methods struggle to capture inter-feature dependencies because of their frequent evaluation of features independently or their full dependence on stochastic processes. To resolve these limitations, a Mutual-Information-Weighted Shapley Graph (MISG) is proposed. This method offers deterministic and structured feature selection combined with graph theory, cooperative game theory, and information theory. MISG extracts deep features from ResNet50 and EfficientNet30 using three continuous stages: graph construction, structural Shapley estimation, and ML-based redundancy pruning. This process ensures discriminative selection and non-redundant features, as shown in Equation (3). Finally, the MISG framework reduces the high-dimensional feature space by retaining meaningful and discriminative attributes, which ensures enhanced classification performance and generalization for ASD emotion recognition, as shown in Figures 5 and 6. The number of discriminative selected features for each dataset is depicted in Figure 7 and Table 4, respectively. MISG provides feature selection as an optimization, which enhances the cumulative Shapley contribution and MI-based feature dependency while reducing the selected feature subset size. This allows the selection of structurally informative and non-redundant features, which enhances classification performance.

$$\max_{S\subseteq F} J(S) = \alpha \sum_{f_i \in S} \phi(f_i) + \beta \sum_{(f_i, f_j)\in S} MI\big(f_i, f_j\big) - \gamma |S| \tag{3}$$

Where S is a selected feature subset, $\phi(f_i)$ is a Shapley value of feature f_i, $MI(f_i, f_j)$ specifies the MI between feature pairs, $|S|$ is a number of selected features, α, β and γ are weighting coefficients balancing feature importance, dependency and subset compactness.

The MISG contains five stages which are discussed as following:

1. Computation of pairwise mutual information between features to quantify dependency relationships.

2. Graph construction where each feature is represented as a node and mutual information values define edge weights.

3. Estimation of feature importance using Shapley values to measure marginal contributions within the graph structure.

4. Ranking and selection of features based on the combined mutual-information and Shapley scores.

5. Pruning of features whose contribution scores are below a predefined threshold, thereby resultant in a discriminative feature subset.

The MISG models each feature as a graph node, while edge weights are estimated by pairwise mutual information to capture feature dependencies. The optimization objective enhances feature relevance and structural contribution while reducing redundancy. The graph construction requires $O(N^2)$ pairwise mutual-information evaluations for N features, whereas Shapley-based importance estimation is performed on the constructed graph to identify discriminative features. The iterative pruning process terminates when no further improvements in the objective function are attained or when the maximum iteration limit is reached, thereby ensuring convergence to a stable feature subset. Unlike heuristic feature ranking methods, MISG utilizes Shapley values derived from cooperative game theory to estimate the marginal contribution of every feature within the mutual-information graph. The feature selection is formulated to maximize cumulative Shapley contribution while reducing feature redundancy thereby enabling an evaluation of feature importance. The selected subset maintains structurally informative features, which contribute significantly to classification performance.

The MISG combines deterministic graph construction and Shapley-based feature significance estimation with a guided iterative optimization for feature subset refinement. The MI graph and Shapley values are estimated from the extracted features, while the subsequent iterative search provides candidate feature subsets by population updates and mutation operators to avoid local optima and attain a discriminative feature subset. Therefore, MISG is based on hybrid graph-based feature optimization compared to deterministic feature selection methods. The feature selection threshold is defined using training data within the 5-fold cross-validation framework. The MI threshold $\tau_{MI} = 0.10$ is selected to remove weak feature dependencies while maintaining informative feature interactions, whereas the Shapley score threshold $\tau_{SH} = 0.05$ is utilized to retain significant features with marginal contributions for classification performance. Multiple threshold combinations are estimated during the experiments, and the final values are selected based on the trade-off between classification accuracy and feature reduction. The same threshold values are utilized across all datasets to ensure fair evaluation. The visualizations presented in Figure 5 are generated from the proposed MISG feature-selection framework. After creating the MI graph, each feature is allocated a Shapley value representing its marginal contribution to emotion classification. The selected high-contribution features are mapped to their respective spatial locations in the final convolutional feature maps of the feature extraction process. The mapped feature significance value is normalized on the original facial images to provide facial regions retained by MISG. Figure 5 shows the structural feature importance identified by the graph-based feature selection rather than gradient-based network attention.

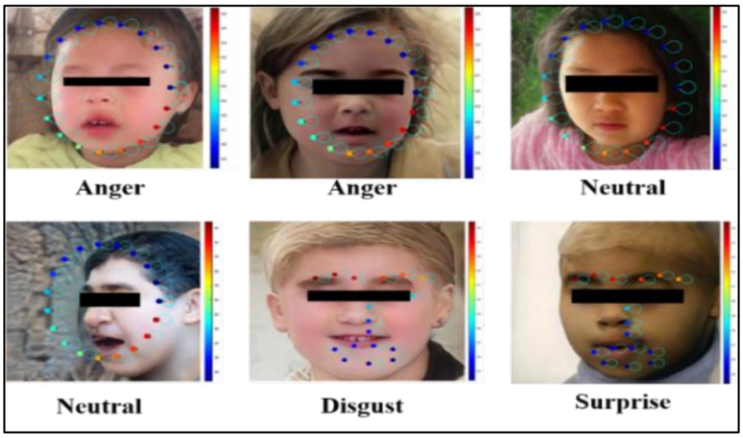


Figure 5. Visualization of MISG-selected discriminative facial features illustrating key emotional regions for classification

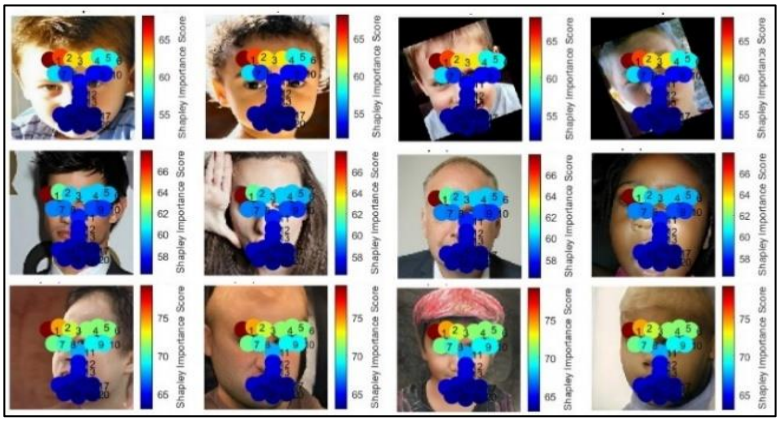
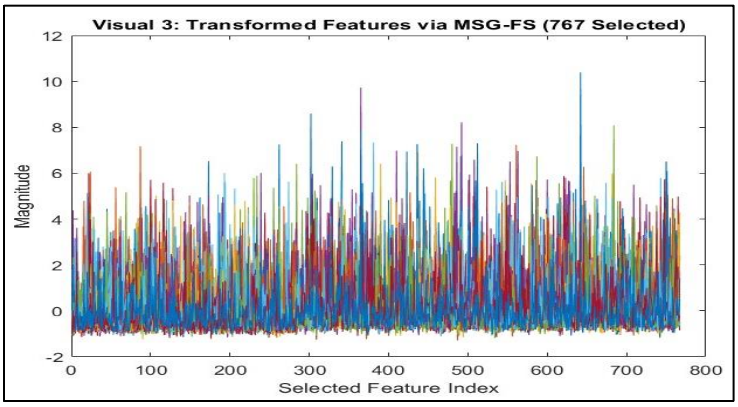
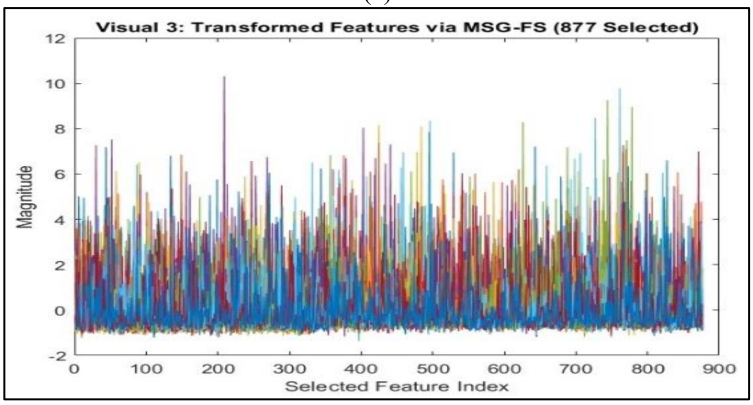


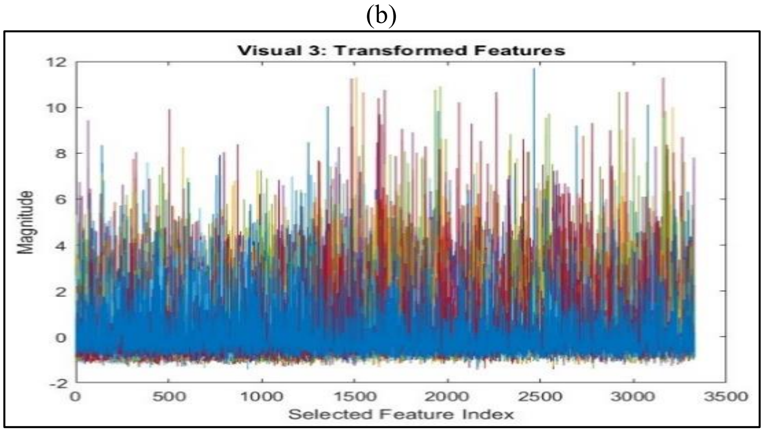
Figure 6. Shapley MIS facial score graphs for discriminative feature selection



(a)



(b)



(c)

Figure 7. Transformed features from (a) AffectNet, (b) Autistic Children Emotions – Dr. Fatma M Talaat, and (c) Young AffectNet HQ datasets

Table 4. Selected features for various datasets

Dataset	MISG-Selected Features
AffectNet	767
Autistic Children Emotions - Dr. Fatma M. Talaat	877
Young AffectNet HQ	882

3.5 Pseudocode for Proposed MISG Feature Selection

Figure 8 shows the flowchart of the MISG for feature refinement. Algorithm 1 summarizes the MISG procedure, including graph construction, Shapley-based importance estimation, redundancy pruning, objective-function evaluation, and convergence.

Algorithm 1: MISG-based feature selection

Input: $F = \{f_1, f_2, \ldots, f_n\}$ Extracted deep feature set obtained from ResNet50 and EfficientNetB0, τ_{MI} = Mutual Information threshold, τ_S = Shapley score threshold and Max_{Iter} Maximum number of optimization iterations

Output: FS Optimized discriminative feature subset

Step 1: Extract the raw deep feature vector (F) using the dual-stream backbone

Step 2: Compute the pairwise Mutual Information $MI(f_i, f_j)$ between every feature pair

Step 3: Construct the Feature Interaction Graph $G = (V, E)$, where:

Each feature represents a node $v_i \in V$

Edge weights correspond to the pairwise values $E_{i,j} = MI(f_i, f_j)$

Step 4: Estimate the Shapley value $\Phi(f_i)$ for each feature node to measure its marginal contribution

Step 5: Compute the combined feature importance score $Score\ (f_i) = MI_{avg}(f_i) + \Phi(f_i)$

Step 6: Rank all features according to $Score\ (f_i)$ in descending order

Step 7: Initialize a population of candidate feature subsets (X) (binary feature masks) $Initialize\ Population\ P = \{X_1, X_2, \ldots, X_n\}$

Step 8: Repeat for each iteration $t = 1$ to T

Evaluate the information-theoretic fitness objective function for each solution mask $Fitness\ (X_k) = Accuracy - \lambda \cdot MI_{redundancy}$

Identify and isolate redundant feature pairs where $MI(f_i, f_j) > \tau_{MI}$

Apply a redundant pruning operator to eliminate cloned metrics, retaining the node with the higher $Score\ (f_i)$

Execute a guided mutation operator to update the candidate feature population while explicitly preserving high-Shapley nodes

Step 9: Until convergence is reached or iteration limit $t = T$ is met

Step 10: Extract the final remaining features that satisfy $\begin{cases} MI(f_i, f_j) < \tau_{MI} \\ Score\ (f_i) > \tau_S \end{cases}$

Step 11: Generate the structural optimized feature subset FS

Step 12: Provide FS as the direct input to the CWAEML classifier and Return FS

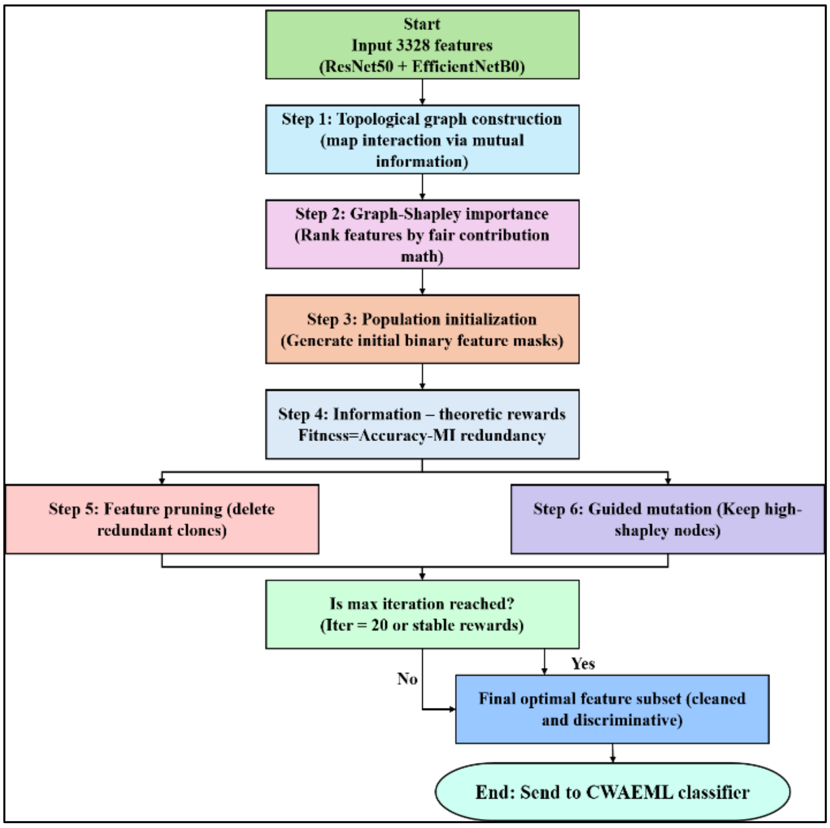


Figure 8. Flowchart of procedural steps of MISG for structured graph-based feature refinement

3.6 Feature Classification

Further, the optimized feature subsets are classified by using MSVM, Naïve Bayes (NB), KNN, DT, and RF classifiers. These heterogeneous learners capture complementary decision boundaries, probabilistic relationships, instance-level variations, and nonlinear feature interactions. Then, MSVM decision function for multiclass emotion discrimination is defined as Equation (4).

$$f(z) = arg\,\max_i(w_i.z + b_i) \tag{4}$$

3.7 Confidence Weighted Aggregation (CWAEML)

Then, a CWAEML strategy is adopted where the classifier weights are adaptively assigned based on validation confidence and prediction stability. The confidence weight assigned to each base classifier is determined from its validation performance. Classifiers achieving higher predictive reliability receive larger weights, thereby contributing more to the final ensemble decision. The weights are normalized and their sum equals one, thereby ensuring a balanced aggregation of classifier outputs. This adaptive weighting mechanism enables the ensemble to leverage high-confidence classifiers while reducing the impact of comparatively weaker models. The classifier outputs are fused using CWAEML and the weighted class probability is computed as in Equation (5).

$$p_{wa}^{j} = \sum_{i=1}^{N} W_i p_i^{j} \tag{5}$$

Where, p_i^j signifies the predicted probability of class j from classifier i and W_i is its confidence-based weight derived from validation stability. The final prediction is obtained through weighted aggregation of the individual classifier outputs using the normalized confidence weights. The flowchart of CWA is given in Figure 9.

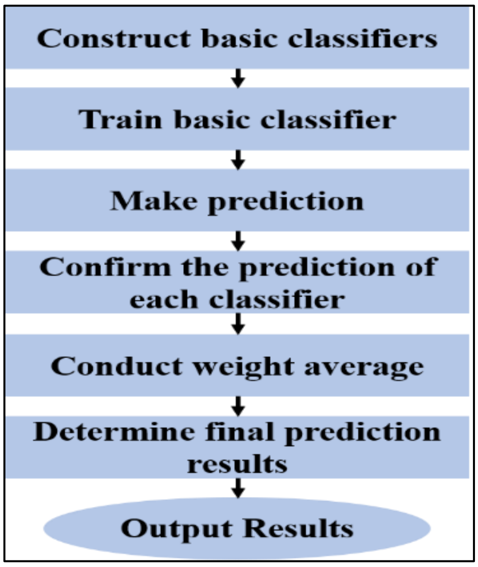


Figure 9. Flow chart of the procedure of CWA for adaptive ensemble-based decision fusion

Table 5 shows the hyperparameters considered for the proposed MISG-CWAEML model. The use of Stochastic Gradient Descent with Momentum (SGDM) with a learning rate of 0.01 and 20 epochs provides stable convergence with reduced training time. The small population size $N = 10$ and iterations $T = 20$ reduce complexity, while MI-Redundancy Penalty $\lambda = 0.5$ ensures a balanced trade-off between feature relevance and redundancy. Lastly, 5-fold cross-validation is selected to maintain fairness, avoid overfitting, and ensure reliable performance evaluation.

Table 5. Hyperparameter configurations considered for the proposed MISG-CWAEML model

Parameters	**Value**
Optimizer	SGDM
Initial Learning Rate	0.01
Maximum Epochs	20
Number of Solutions ($\boldsymbol{N}$)	10
Maximum Iterations ($\boldsymbol{T}$)	20
MI-Redundancy Penalty ($\boldsymbol{\lambda}$)	0.5
K-NN Evaluator ($\boldsymbol{k}$)	5
Number of Neighbors	5 to 10 (Dataset dependent)
Number of Trees	30 to 50 (Dataset dependent)
Kernel Function & Box Constraint	Linear Kernel, $C = 0.1$
Max Number of Splits	20
Cross-Validation Folds ($\boldsymbol{K}$)	5-fold, 80% training and 20% testing

3.8 Final Emotion Recognition

Finally, the unified MISG-CWAEML framework ensures optimized feature learning and adaptive ensemble decision fusion for achieving stable and accurate emotion recognition across heterogeneous ASD datasets by minimizing overfitting.

4. Results and Discussion

The proposed MISG-CWAEML is implemented in MATLAB R2020b. The software and hardware requirements include a windows 10 operating system, 16GB RAM, an Intel i7-7700k @4.20GHz processor, and an NVIDIA GeForce GTX 1060 6GB GPU. Experiments are conducted using stratified 5-fold cross-validation, where dataset partitioning is performed before augmentation. Data augmentation is applied to the training folds, while testing data remains unseen during model development. To ensure reproducibility, a fixed random seed of 42 is used throughout data partitioning and model training. Hyperparameters are selected empirically based on validation performance. Accuracy, sensitivity, specificity, precision, and F1-score are the performance metrics estimated in this research to demonstrate the performance of MISG-CWAEML. The mathematical expressions for these metrics are represented in Equations (6) - (10).

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negative} \times 100 \tag{6}$$

$$Sensitivity = \frac{True\ Positive}{True\ Positive + False\ Negative} \times 100 \tag{7}$$

$$Specificity = \frac{True\ Negative}{True\ Negative + False\ Positive} \times 100 \tag{8}$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \times 100 \tag{9}$$

$$F1 - score = \frac{2 \times Precision \times sensitivity}{Precision + sensitivity} \tag{10}$$

4.1 Performance Analysis

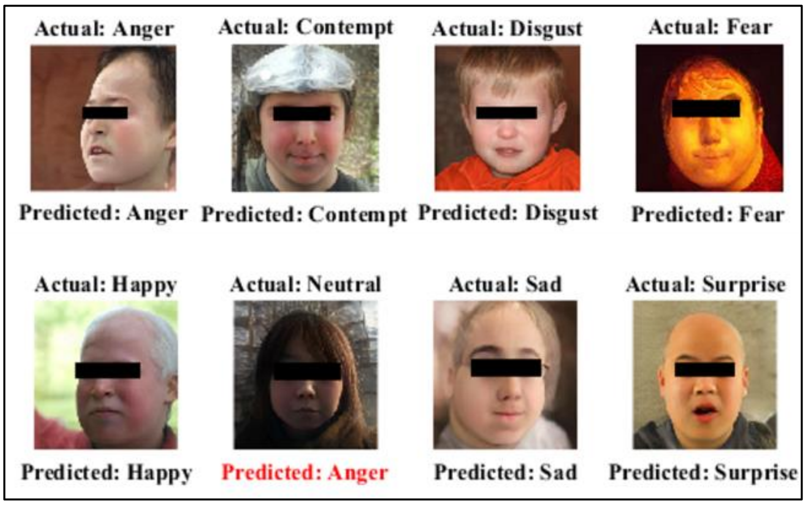


Figure 10. Emotion classification results

Figure 10 presents the overall classification results of the MISG-CWAEML by comparing actual and predicted emotion labels. Most samples are correctly categorized across various emotion categories, indicating high accuracy and resilience. Misclassification failures represent minor issues of inter-class similarity. Computational complexity is measured in floating-point operations (GFLOPs). The reported complexity of 40.1 GFLOPs corresponds to a single forward inference pass of the proposed MISG-CWAEML using an input image size of

224×224×3. This GFLOP value is computed from the network's computational complexity and represents the total number of floating-point operations required to process one input image. Runtime analysis evaluates the computational efficiency of the proposed MISG-CWAEML framework. The reported runtime of 385 seconds corresponds to the total model training time for one complete experimental run, including feature extraction, MISG-based feature selection, and CWAEML classifier training. The runtime is measured on the experimental platform used in this research and represents the average execution time across cross-validation runs. Samples are verified using the original dataset annotations and include correctly classified, visually similar, and representative misclassified emotion samples. These samples demonstrate the model's ability to differentiate subtle facial expressions while showcasing its performance on challenging emotion categories.

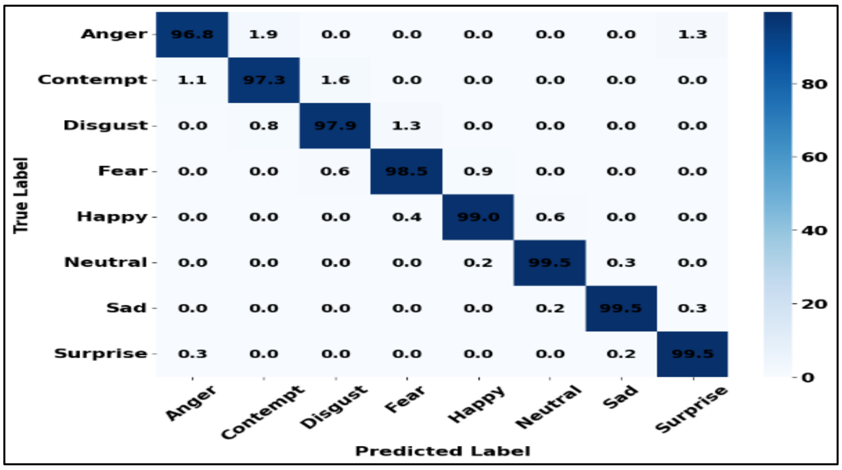
The proposed MISG-CWAEML demonstrates computational efficiency, requiring 40.1 GFLOPs, 425 MB of memory consumption, an average training runtime of 385 s, and an inference latency of 24 ms per image. The MISG-based feature refinement reduces feature redundancy before classification, thereby enabling emotion prediction while maintaining high recognition performance. The performance of the classifiers trained with optimized features is evaluated on benchmark datasets. The comparison of these models with the proposed model is defined in Table 6. ViT achieved comparatively lower performance because transformer-based models require large-scale training data to learn robust feature representations. However, the proposed MISG-CWAEML utilizes limited dataset samples through graph-based feature refinement and confidence-weighted ensemble learning, thereby resulting in better emotion prediction performance across all datasets.

Table 6. Classifier results with optimized features for 3 various datasets

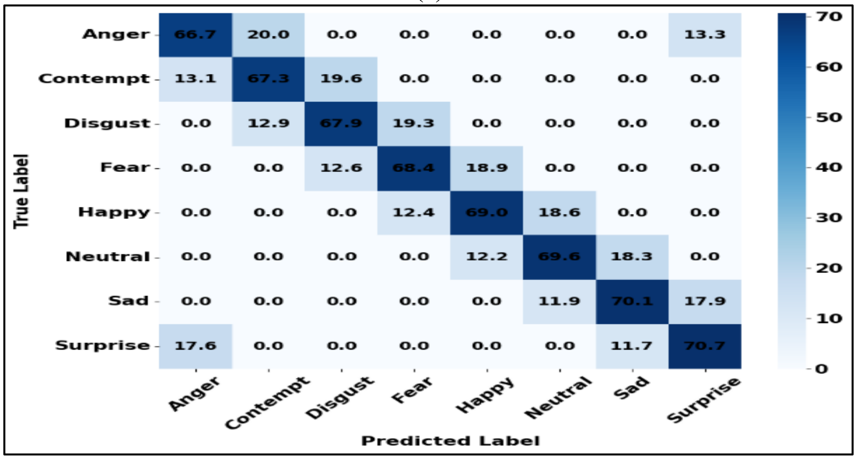
Methods	Sensitivity (%)	Accuracy (%)	Precision (%)	Specificity (%)	F1-score (%)
AffectNet dataset					
KNN	95.8	96.15	96.45	95.22	96.12
MSVM	96.88	97.42	97.7	96.2	97.28
NB	93.88	94.5	94.12	93.1	93.55
RF	95.9	96.88	96.8	95.1	96.39
DT	94.88	95.33	95.8	94.1	95.34
ViT	95.65	96.42	96.58	95.38	96.21
Proposed CWAEML	98.12	98.75	98.9	97.88	98.51
Autistic Children Emotions-Dr. Fatma M Talaat dataset					
KNN	94.66	94.36	95.22	94.08	94.77
MSVM	96.12	95.39	96.55	94.68	95.27
NB	92.55	93.17	93.1	93.8	91.88
RF	94.9	94.81	95.8	95.4	94.1
DT	93.8	93.96	94.45	94.12	93.55
ViT	94.85	95.12	95.48	94.36	95.03
Proposed CWAEML	97.45	97.95	98.15	96.88	97.80
Young AffectNet HQ dataset					
KNN	63.68	63.99	63.78	63.83	64.08
MSVM	64.01	64.95	65.11	64.47	65.48
NB	60.55	62.15	60.1	61.34	61.22
RF	64.22	65.2	63.45	64.54	65.15
DT	64	62.87	63.08	63.42	63.87
ViT	61.92	62.84	62.15	61.75	62.48
Proposed CWAEML	67.55	68.72	69.42	66.88	68.47

4.2 Confusion Matrix Analysis

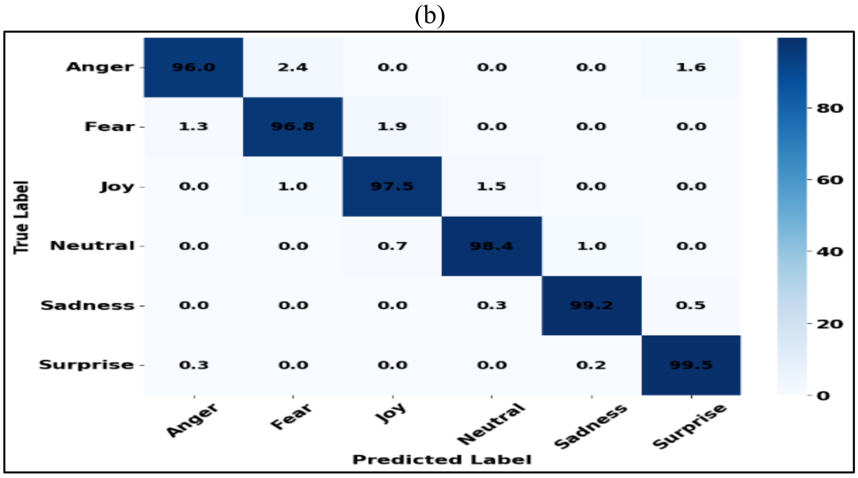
The confusion matrix for each dataset is given in Figures 11 (a), (b) and (c). Analysing this matrix across various classes enables one to understand the robustness of the model, how it identifies misclassification errors, and how it provides better accuracy in the classification process of autistic children’s emotions.



(a)



(b)



(c)

Figure 11. Confusion matrix for (a) AffectNet, (b) Autistic Children Emotions – Dr. Fatma M Talaat, and (c) Young AffectNet HQ datasets

4.3 Statistical Analysis

The statistical significance and reliability of the MISG-CWAEML with baseline models are estimated by P-value and Confidence Interval (CI) across three datasets, as presented in Table 7. The MISG-CWAEML outperforms the baseline ResNet50, hybrid ResNet + EffNet, and SHAP-based ensemble models in all three datasets. The MISG-CWAEML provides a better CI, thereby validating its robustness in complex facial emotion recognition. To evaluate the significance of the performance differences among the compared methods, paired t-test is conducted using repeated experimental runs. The null hypothesis (H_0) assumes that there is no significant difference between the performance of the proposed MISG-CWAEML framework and the baseline methods, while the alternative hypothesis (H_1) assumes a significant performance difference. The mean and standard deviation of the evaluation metrics are computed across multiple runs. Statistical significance is determined at a confidence level of 95% ($p < 0.05$), and confidence intervals are calculated based on the observed metric distributions.

To evaluate the statistical reliability of the proposed MISG-CWAEML framework, a statistical significance test is conducted using results obtained from the 5-fold cross-validation. Each fold is considered an independent experimental run, resulting in five observations per dataset. The mean and standard deviation of the evaluation metrics are computed across all folds. Furthermore, a paired t-test is performed between the proposed MISG-CWAEML framework and the baseline method using fold-wise accuracy. The obtained p-values were below the significance level $\alpha = 0.05$, indicating that the performance improvements achieved by the proposed framework are statistically significant and did not occurr by chance.

Table 7. Statistical analysis results of proposed MISG-CWAEML for AffectNet, Autistic Children Emotions – Dr. Fatma M Talaat, and Young AffectNet HQ datasets

Datasets	Methods	P-Value	95% Confidence Interval (CI)	Margin of Error (±)
AffectNet	ResNet50 (Baseline)	0.038	91.95% – 93.05%	0.55
	Hybrid (ResNet + EffNet)	0.029	94.62% – 95.62%	0.50
	SHAP + Ensemble	0.022	96.90% – 97.76%	0.43
	Proposed MISG-CWAEML	0.012	98.42% – 99.08%	0.33
Autistic Children Emotions – Dr. Fatma M Talaat	ResNet50 (Baseline)	0.034	90.72% – 91.72%	0.50
	Hybrid (ResNet + EffNet)	0.026	93.95% – 94.95%	0.50
	SHAP + Ensemble	0.020	95.65% – 96.55%	0.45
	Proposed MISG-CWAEML	0.010	97.60% – 98.30%	0.35
Young AffectNet HQ	ResNet50 (Baseline)	0.042	58.82% – 60.06%	0.62
	Hybrid (ResNet + EffNet)	0.031	62.15% – 63.45%	0.65
	SHAP + Ensemble	0.025	64.72% – 65.94%	0.61
	Proposed MISG-CWAEML	0.015	67.92% – 69.52%	0.80

4.4 K-fold Cross-Validation Analysis

K-fold cross-validation is accomplished in this research using K-fold validation and threshold-based feature selection for three datasets, as shown in Table 8. Various K-fold values (K=2, 3, 5, 6, and 7) are calculated for different evaluation metrics. In each dataset, K=5 performed better than other K values (K=2, 3, 6, and 7). At K=5, MISG-CWAEML achieves accuracies of 98.75%, 97.95% and 68.72%, in the AffectNet, Autistic Children Emotions – Dr. Fatma M Talaat dataset, and Young AffectNet datasets. K-fold validation minimizes overfitting risk by dividing images into K-subsets.

Table 8. K-fold cross-validation analysis of the proposed MISG-CWAEML on AffectNet, Autistic Children Emotions – Dr. Fatma M Talaat dataset, and Young AffectNet datasets

Dataset	K-Values	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1-score (%)	Selected features
AffectNet	K=2	92.15	91.80	91.10	92.45	92.12	720
	K=3	95.42	94.88	94.22	95.70	95.28	790
	K=5	98.75	98.12	97.88	98.90	98.51	767
	K=6	97.90	97.22	97.05	98.15	97.68	835
	K=7	97.22	96.55	96.40	97.42	96.98	890
Autistic Children Emotions – Dr. Fatma M Talaat	K=2	91.33	90.88	90.15	92.10	91.49	680
	K=3	94.80	94.12	93.55	95.22	94.66	760
	K=5	97.95	97.45	96.88	98.15	97.80	877
	K=6	97.10	96.55	96.12	97.40	96.97	920
	K=7	96.55	95.90	95.88	96.88	96.39	975
Young AffectNet HQ	K=2	61.45	60.88	60.12	62.10	61.48	810
	K=3	65.22	64.10	63.88	65.90	64.99	860
	K=5	68.72	67.55	66.88	69.42	68.47	882
	K=6	66.88	65.90	65.12	67.22	66.55	940
	K=7	65.45	64.12	63.90	66.15	65.12	1015

4.5 Comparative Analysis of Emotion Classification

Figure 12 shows the proposed MISG model's performance which highlights the optimization capability by comparatively analysing against existing methods including FACS-CNN [17], SVM [21], Xception [16], AutYOLO-ATT [19] and SVM [22] respectively.

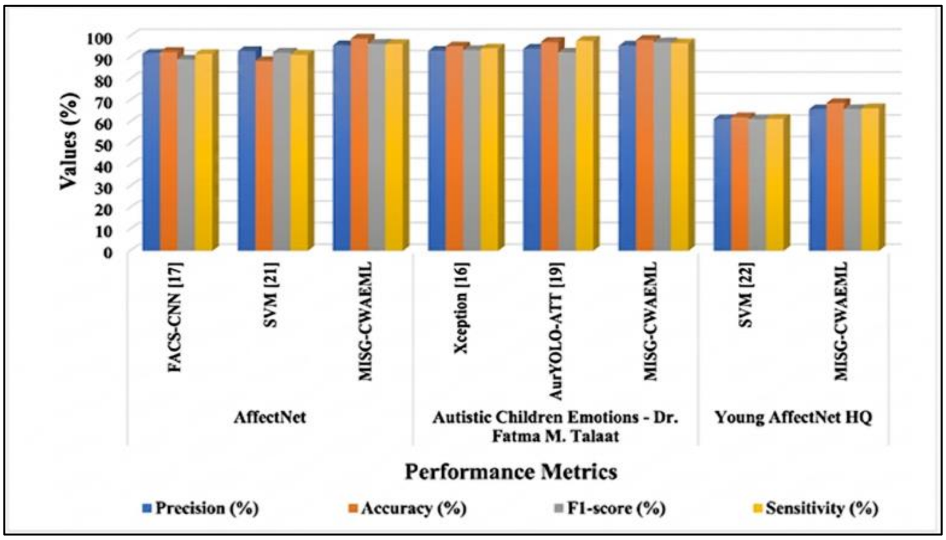


Figure 12. Comparative analysis of proposed MISG-CWAEML model

The ablation study shows the performance of each model baseline, hybrid, SHAP, and proposed MISG-CWAEML across AffectNet, Autistic Children Emotions - Dr. Fatma M. Talaat, and Young AffectNet HQ datasets, as presented in Table 9. Hybrid feature extraction, using models such as ResNet50 and EfficientNetB0, enhances all metrics compared to a single model, thereby ensuring that deep features improve emotional representation. Integrating SHAP-based selection with ensemble learning further improves discriminative ability by eliminating irrelevant features and enhancing class separation. The performance improvement from Hybrid + Standard SHAP + Ensemble to MISG-CWAEML demonstrates the contribution of the proposed graph-based feature refinement mechanism over conventional SHAP-based feature selection.

Table 9. Performance analysis of baseline, hybrid, SHAP and proposed MISG-CWAEML models across AffectNet, Autistic Children Emotions - Dr. Fatma M. Talaat and Young AffectNet HQ datasets

Datasets	Description	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1-Score (%)
AffectNet	ResNet50 + Standard SVM (No Hybrid, No Selection)	92.50	91.88	92.10	91.45	91.99
	Hybrid (ResNet + EffNet) + Standard SVM (No Selection)	95.12	94.55	94.80	94.10	94.67
	Hybrid + Standard SHAP Selection + Ensemble	97.33	96.80	97.12	96.44	96.96
	Proposed MISG-CWAEML (Full model)	98.75	98.90	98.12	97.88	98.51
Autistic Children Emotions – Dr. Fatma M Talaat	ResNet50 + Standard SVM (No Hybrid, No Selection)	91.22	90.55	90.80	90.12	90.67
	Hybrid (ResNet + EffNet) + Standard SVM (No Selection)	94.45	93.80	94.12	93.55	93.96
	Hybrid + Standard SHAP Selection + Ensemble	96.10	95.44	95.88	95.12	95.66
	Proposed MISG-CWAEML (Full model)	97.95	97.84	97.45	96.87	97.80
Young AffectNet HQ	ResNet50 + Standard SVM (No Hybrid, No Selection)	59.44	58.12	58.55	57.88	58.33
	Hybrid (ResNet + EffNet) + Standard SVM (No Selection)	62.80	61.55	61.90	60.12	61.72
	Hybrid + Standard SHAP Selection + Ensemble	65.33	64.22	64.88	63.45	64.55
	Proposed MISG-CWAEML (Full model)	68.72	65.65	67.55	66.79	65.51

4.6 Discussion

The MISG-CWAEML framework consists of three key stages: Feature Extraction (using ResNet50 and EfficientNetB0), Feature Selection (using MISG), and Classification (using CWAEML). The two extraction methods together provide a detailed representation of emotional expressions of autistic children, which increases overall robustness. The model outperforms KNN, MSVM, RF, DT and NB, recording the following superior accuracy scores: 98.75% on AffectNet, 97.95% on Autistic Children Emotions - Dr. Fatma M. Talaat and 68.72% on Young AffectNet HQ. The high accuracies achieved on AffectNet and Autistic Children Emotions - Dr. Fatma M. Talaat, datasets are attributed to the discriminative feature refinement capability of MISG and the adaptive fusion mechanism of CWAEML. The leakage-prevention protocol and fold-wise evaluation ensure that the reported results are not influenced by train-test leakage. The lower accuracy on Young AffectNet HQ indicates the presence of significant domain shifts, including variations in age, facial morphology, expression intensity, and image characteristics indicating cross-domain generalization challenges rather than model generalization. The lower accuracy achieved on Young AffectNet HQ dataset is due to the domain-shift effects between datasets. Young AffectNet HQ contains age-specific facial characteristics, variations in expression intensity, image quality differences and diverse acquisition conditions, which alter the underlying feature distribution. These demographic and visual discrepancies minimize the transferability of features learned from other datasets, thereby increasing classification difficulty. The observed performance degradation highlights the challenges of cross-domain emotion recognition and motivates future investigation of domain adaptation and transfer learning strategies to improve generalization. MISG module successfully performs the optimization of classifiers by removing the irrelevant and redundant features, reducing the initial 3,328 features to 767, 877 and 882 features for the respective datasets. The model captures feature correlations through graph-based modelling, which contributes to improved classification performance from activation maps to specific face-action regions and offers a bridge between computational analysis and real-world behaviour evaluation. Finally, this dual-phase structure is confirmed by statistical significance analysis: the standard deviations and p-values are below 0.05 for all folds. This systematic consistency ensure that using graph-based structural refinement along with CWAEML provides reliable architectural stability.

5. Conclusion

The study proposed the MISG-CWAEML approach for classifying the emotions of children with ASD. ResNet50 and EfficientNetB0 were used to extract deep features that have complementary high-level semantic and low-level spatial information to enhance the training feature representation. MISG performs graph-based structural feature refinement by modelling feature correlations and estimating marginal contributions using Shapley theory, resulting in the selection of informative and non-redundant features. This also improved the robustness of the classifier and reduced redundant computation. The CWAEML dynamically integrated multiple base learners including MSVM, NB, KNN, DT, and RF with confidence weighting, thereby improving predictability across different datasets. The proposed MISG-CWAEML achieved classification accuracies of 98.75%, 97.95%, and 68.72% on the AffectNet, Autistic Children Emotions – Dr. Fatma M Talaat, and Young AffectNet HQ datasets, respectively, while reducing the original 3328-dimensional feature space to 767, 877, and 882 discriminative features. These findings demonstrate the model's potential for integration into clinical and education systems for emotion assessment and behavioural analysis of children with ASD. Future work will focus on integrating multimodal emotion analysis using facial, audio, neurophysiological, and neuroimaging data, along with cross-domain feature interactions, to enhance the interpretability, clinical applicability, and personalization of ASD emotion recognition and diagnostic support systems.

Acknowledgements

The authors would like to thank REVA University and the REVA AICTE IDEA Lab for continuous support, infrastructure, and encouragement in carrying out the research, fostering a collaborative academic environment.

Funding Information

This research received no external funding.

Conflict of Interest

The authors declare no conflict of interest.

Ethical Approval

The datasets used in this research such as AffectNet, Autistic Children Emotions – Dr. Fatma M. Talaat, and Young AffectNet HQ are publicly available and were collected with prior consent by the original dataset providers. No additional human data were collected in this study. All experiments adhere to the ethical standards for research involving publicly available data, particularly for studies involving children.

Data Availability

The study utilizes publicly available data from the Kaggle repository, specifically the AffectNet [23], Autistic Children Dr. Fatma M. Talaat [24], and Young AffectNet HQ [25] datasets as listed in the references.

References

[1] Pavez, Rodolfo, Jaime Diaz, Jeferson Arango-Lopez, Danay Ahumada, Carolina Mendez-Sandoval, and Fernando Moreira. "Emo-Mirror: A Proposal to Support Emotion Recognition in Children with Autism Spectrum Disorders." Neural Computing and Applications 35, no. 11 (2023): 7913-7924.

[2] Deveau, Nicholas, Peter Washington, Emilie Leblanc, Arman Husic, Kaitlyn Dunlap, Yordan Penev, Aaron Kline, Onur Cezmi Mutlu, and Dennis P. Wall. "Machine Learning Models Using Mobile Game Play Accurately Classify Children with Autism." Intelligence-Based Medicine 6 (2022): 100057.

[3] Liao, Mengyi, Hengyao Duan, and Guangshuai Wang. "Application of Machine Learning Techniques to Detect the Children with Autism Spectrum Disorder." Journal of Healthcare Engineering 2022, no. 1 (2022): 9340027.

[4] Han, Yu, Donna M. Rizzo, John P. Hanley, Emily L. Coderre, and Patricia A. Prelock. "Identifying Neuroanatomical and Behavioral Features for Autism Spectrum Disorder Diagnosis in Children Using Machine Learning." PLoS One 17, no. 7 (2022): e0269773.

[5] Priyadarshini, Ishaani. "Autism Screening in Toddlers and Adults Using Deep Learning and Fair AI Techniques." Future Internet 15, no. 9 (2023): 292.

[6] Lin, Ping-I., Mohammad Ali Moni, Susan Shur-Fen Gau, and Valsamma Eapen. "Identifying Subgroups of Patients with Autism by Gene Expression Profiles Using Machine Learning Algorithms." Frontiers in Psychiatry 12 (2021): 637022.

[7] Alhakbani, N. "Facial Emotion Recognition-Based Engagement Detection in Autism Spectrum Disorder." International Journal of Advanced Computer Science and Applications 15, no. 3 (2024): 959–967.

[8] Sundaresan, Avirath, Brian Penchina, Sean Cheong, Victoria Grace, Antoni Valero-Cabré, and Adrien Martel. "Evaluating Deep Learning EEG-Based Mental Stress Classification in Adolescents with Autism for Breathing Entrainment BCI." Brain Informatics 8, no. 1 (2021): 13.

[9] Farooq, Muhammad Shoaib, Rabia Tehseen, Maidah Sabir, and Zabihullah Atal. "Detection of Autism Spectrum Disorder (ASD) in Children and Adults Using Machine Learning." scientific reports 13, no. 1 (2023): 9605.

[10] Vakadkar, Kaushik, Diya Purkayastha, and Deepa Krishnan. "Detection of Autism Spectrum Disorder in Children Using Machine Learning Techniques." SN computer science 2, no. 5 (2021): 386.

[11] Khudhur, Dhuha Dheyaa, and Saja Dheyaa Khudhur. "The Classification of Autism Spectrum Disorder by Machine Learning Methods on Multiple Datasets for Four Age Groups." Measurement: Sensors 27 (2023): 100774.

[12] Aina, Joseph, Oluwatunmise Akinniyi, Md Mahmudur Rahman, Valerie Odero-Marah, and Fahmi Khalifa. "A Hybrid Learning-Architecture for Mental Disorder Detection Using Emotion Recognition." IEEE Access 12 (2024): 91410-91425.

[13] Salhi, Intissar, Mohammed Qbadou, Soukaina Gouraguine, Khalifa Mansouri, Chris Lytridis, and Vassilis Kaburlasos. "RETRACTED: Towards Robot-Assisted Therapy for Children with Autism—The Ontological Knowledge Models and Reinforcement Learning-Based Algorithms." Frontiers in Robotics and AI 9 (2022): 713964.

[14] Bala, Mousumi, Mohammad Hanif Ali, Md Shahriare Satu, Khondokar Fida Hasan, and Mohammad Ali Moni. "Efficient Machine Learning Models for Early Stage Detection of Autism Spectrum Disorder." Algorithms 15, no. 5 (2022): 166.

[15] Song, Chao, Zhong-Quan Jiang, Li-Fei Hu, Wen-Hao Li, Xiao-Lin Liu, Yan-Yan Wang, Wen-Yuan Jin, and Zhi-Wei Zhu. "A Machine Learning-Based Diagnostic Model for Children with Autism Spectrum Disorders Complicated with Intellectual Disability." Frontiers in psychiatry 13 (2022): 993077.

[16] Talaat, Fatma M., Zainab H. Ali, Reham R. Mostafa, and Nora El-Rashidy. "RETRACTED ARTICLE: Real-Time Facial Emotion Recognition Model Based on Kernel Autoencoder and Convolutional Neural Network for Autism Children: Fatma M. Talaat et al." Soft Computing 28, no. 9 (2024): 6695-6708.

[17] Saranya, A., Anandan, R. “Facial Action Coding and Hybrid Deep Learning Architectures for Autism Detection.” Intelligent Automation and Soft Computing 33, no. 2 (2022): 1167-1182.

[18] Savchenko, Andrey V., Lyudmila V. Savchenko, and Ilya Makarov. "Classifying Emotions and Engagement in Online Learning Based on a Single Facial Expression Recognition Neural Network." IEEE transactions on affective computing 13, no. 4 (2022): 2132-2143.

[19] Hosney, Reham, Fatma M. Talaat, Eman M. El-Gendy, and Mahmoud M. Saafan. "AutYOLO-ATT: An Attention-Based YOLOv8 Algorithm for Early Autism Diagnosis Through Facial Expression Recognition." Neural Computing and Applications 36, no. 27 (2024): 17199-17219.

[20] Talaat, Fatma M. "Real-Time Facial Emotion Recognition System Among Children with Autism Based on Deep Learning and IoT." Neural Computing and Applications 35, no. 17 (2023): 12717-12728.

[21] Cheng, Ming, Yingying Zhang, Yixiang Xie, Yueran Pan, Xiao Li, Wenxing Liu, Chengyan Yu et al. "Computer-Aided Autism Spectrum Disorder Diagnosis with

Behavior Signal Processing." IEEE Transactions on Affective Computing 14, no. 4 (2023): 2982-3000.

[22] Ciraolo, Davide, Maria Fazio, Rocco Salvatore Calabrò, Massimo Villari, and Antonio Celesti. "Facial Expression Recognition Based on Emotional Artificial Intelligence for Tele-Rehabilitation." Biomedical Signal Processing and Control 92 (2024): 106096.

[23] AffectNet Dataset: https://www.kaggle.com/datasets/mstjebashazida/affectnet (accessed on July 2026).

[24] Autistic Children Emotions - Dr. Fatma M. Talaat dataset: https://www.kaggle.com/datasets/fatmamtalaat/autistic-children-emotions-dr-fatma-m-talaat/data (accessed on July 2026).

[25] Young AffectNet HQ: https://www.kaggle.com/datasets/vfomenko/young-affectnet-hq/data (accessed on July 2026).