Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

Attention-Guided Artificial Intelligence Multi-Agent architecture for Explainable and resilient Skin Disease Classification

Madhavan Babu, Prashant Kumar Yadav

[1]PhD Research Scholar, Department of Computer Science and Engineering, Veer Bahadur Singh Purvanchal University (VBSPU), Jaunpur, Uttar Pradesh, India – 222003

[2]Assistant Professor, Department of Computer Science and Engineering, Veer Bahadur Singh Purvanchal University (VBSPU), Jaunpur, Uttar Pradesh, India – 222003

ARTICLE INFO

Received: 03 Nov 2025

Revised: 17 Dec 2025

Accepted: 28 Dec 2025

ABSTRACT

Artificial intelligence (AI) is gained considerable attention in dermatological image analysis because deep learning models identify visual patterns associated with the different skin diseases. However, variations in lesion appearance, skin tone, image quality, and disease similarity affect classification reliability. Conventional single-model architectures often struggle in combining the matching visual features, distinguish clinically similar conditions, and are providing consistent predictions beyond the heterogeneous dermatological images. This study is introducing an Attention-Guided Cooperative Multi-Agent Network (AGC-MAN) for skin disease classification. The proposed model is combining the four specialized agents responsible for global feature extraction, local lesion characterization, attention-based representation learning, and cooperative classification. A hierarchical attention mechanism is emphasizing discriminative lesion regions, while an adaptive agent coordination module is combining individual predictions through the confidence-weighted consensus learning. A feature consistency mechanism is reducing the disagreements among the agents. The architecture is employing the HAM10000 and ISIC dermatological image datasets for multiclass evaluation. The proposed framework is providing a combined classification strategy that is combining localized visual attention, collaborative inference, and interpretable decision-making for automated skin disease recognition. The comparative evaluation reports 96.10% accuracy, 95.60% precision, 95.20% recall, and 95.40% F1-score, with the macro-AUC of 0.974 at 25 training epochs. when it is compared with the EFAM-Net, accuracy is improving by 1.90 percentage points. The architecture is establishing a cooperative classification strategy for seven-category skin lesion recognition using HAM10000, with the ISIC-2019 providing an additional evaluation benchmark.

Keywords: Skin disease classification, artificial intelligence, multi-agent systems, attention mechanism, cooperative deep learning

1. INTRODUCTION

Skin diseases is a major dermatological healthcare concern because its visual characteristics vary considerably beyond the disease categories, patient skin tones, and lesion development stages. Artificial intelligence (AI)-based image classification is providing an opportunity in supporting the dermatological assessment through the automated feature extraction and pattern recognition. Convolutional neural networks (CNNs) and transformer-based architectures is showing the ability to identify discriminative lesion characteristics from the clinical and dermoscopic images. Recent developments in deep learning also are improving the representation of colour, texture, shape, and spatial information for automated skin disease recognition [1]–[3]. Nevertheless, reliable classification is remaining as challenging because several skin conditions are showing similar visual characteristics. Variations in illumination, image resolution, lesion boundaries, and background artefacts are introducing additional classification difficulties [4], [5]. Further, class imbalance and differences in image acquisition conditions influence the generalized ability ability of Conventional models.

Copyright © 2025 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

Conventional classification architectures generally depend on a centralized prediction network, which are limiting its ability to independently is examining different lesion characteristics and is combining matching diagnostic evidence. Conventional approaches also encounter difficulties in maintaining consistent predictions when local lesion features and global image representations are providing conflicting information [6], [7]. These limitations motivate a collaborative classification model that is combining the specialized learning agents with the attention mechanism.

The primary objective of this research is in developing an attention-guided AI multi-agent architecture that is improving multiclass skin disease classification through the cooperative feature learning. The proposed Attention-Guided Cooperative Multi-Agent Network (AGC-MAN) is introducing four specialized agents for global representation extraction, local lesion analysis, attention-based feature refinement, and confidence-aware classification. This novelty is the combination of hierarchical lesion attention with the adaptive inter-agent communication and confidence-weighted decision fusion. Unlike conventional centralized architectures, the proposed architecture is allowing individual agents to learn matching representations and is contributing in accordance with the classification confidence.

The major contributions are:

1. An attention-guided multi-agent model that is combining the global and local lesion representations through the specialized learning agents.
2. An adaptive cooperative decision mechanism that is combining agent predictions using confidence estimation and feature consistency in improving classification operational reliability and interpretability.

2. RELATED WORKS

Recent research is exploring convolutional learning, attention mechanisms, transformer architectures, and feature fusion for automated skin lesion classification. Rezaee and Ghayoumi Zadeh (2024) develop a self-attention transformer architecture that is combining CNN-based local features with the transformer-derived global representations. This model is achieving 97.48% accuracy on ISIC-2019 and is showing the importance of combining local and contextual information [8]. Yadav et al. (2024) are introducing DSCATNet, a dual-scale lightweight cross-attention transformer that processes different image patch sizes. This framework is achieving 97.80% accuracy on HAM10000, although its model primarily that is focussing on attention-based feature extraction rather than the cooperative multi-agent inference [9].

Ramamurthy et al. (2024) propose a dual-track deep learning model that is combining DenseNet-169, coordinate attention, feature pyramid networks, and global context learning. This architecture is achieving 93.2% accuracy on HAM10000 and is showing the benefits of combining localized and hierarchical representations [10]. Another development is EFAM-Net, which is combining the attention residual learning, parallel ConvNeXt blocks, and multiscale computationally efficient attention feature fusion. This reported accuracy is reaching to 93.95% on HAM10000 and 92.30% on ISIC-2019 [11]. These findings are establishing the importance of multiscale feature extraction and an adaptive attention for distinguishing visually similar lesions.

Ghazouani (2025) is introducing MuRANet, a multi-residual attention model with the dilated attention blocks for the detection of lesions at different spatial scales. The model reports F1-scores of 90%, 88%, and 97% on HAM10000, ISIC-2019, and PH2, respectively [12].

3. PROPOSED METHODOLOGY

3.1. Overview of the AGC-MAN Framework

The proposed Attention-Guided Cooperative Multi-Agent Network (AGC-MAN) is employing four interconnected agents to classify dermatological images into seven skin lesion categories. Each agent is performing a specialized learning operation while exchanging intermediate representations with the other agents. The model is using a shared image preprocessing module, parallel feature extraction branches, hierarchical attention refinement, cooperative feature communication, and confidence-weighted classification. The four agents are defined as follows:

Copyright © 2025 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ Research Article

- Agent 1: Global Representation Agent (GRA): It is extracting global lesion characteristics, which is including entire shape, colour distribution, and structural context.
- Agent 2: Local Lesion Agent (LLA): It examines localized patterns, which is including pigmentation, lesion boundaries, texture variations, and irregular structures.
- Agent 3: Attention Refinement Agent (ARA): It is applying spatial and channel attention to emphasize discriminative lesion characteristics.
- Agent 4: Cooperative Decision Agent (CDA): It is combining the agent representations, is estimating prediction confidence, and is generating the final multiclass classification.

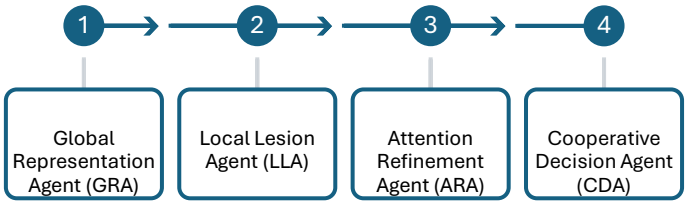


Figure 1: AGC-MAN

The architecture is using cooperative feature exchange rather than the independent classification alone. Each agent is contributing a matching representation, while the decision agent dynamically is adjusting the influence of individual predictions. This arrangement is allowing the model in retaining the global contextual information without suppressing small but diagnostically important lesion characteristics.

3.1: Dermatological Image Preprocessing and Global-Local Feature Extraction

The first step prepares the input images and is generating matching visual representations for the global and local agents. The input dermatological image is represented as $I \in \mathbb{R}^{H\times W\times C}$, where H, W, and C is image height, width, and colour channels, respectively. Each image undergoes resizing, colour normalization, and pixel intensity scaling. The preprocessing operation is reducing the unnecessary variation caused by image acquisition conditions while it is retaining the morphological characteristics of the lesion.

The normalized image is expressed as:

$$I_n(x,y,c) = \frac{I(x,y,c) - \mu_c}{\sigma_c + \epsilon}$$

where I_n is the normalized image, μ_c is representing the channel-wise mean, σ_c is indicating the consistent standard deviation, and ϵ prevents instability. The normalized image is processed using data augmentation, which is including horizontal flipping, vertical flipping, controlled rotation, random cropping, and brightness adjustment. These transformations are improving the diversity of training inputs without modifying the diagnostic label.

To preserve lesion characteristics beyond the different spatial resolutions, the preprocessing module is generating global and local image representations. The global representation is retaining the entire lesion and surrounding context, whereas local patches concentrate on smaller morphological regions. The global feature extraction process is defined as:

$$F_g^{(1)} = \phi_g(I_n;\theta_g) = \mathrm{Conv}_{3\times3}\big(\mathrm{ResBlock}(\mathrm{Pool}(I_n))\big)$$

where $F_g^{(1)}$ is representing the global feature map, ϕ_g is the global feature extraction function, and θ_g is representing its trainable parameters. Residual convolutional blocks is extracting colour transitions, lesion geometry, and spatial relationships. The residual connection is preserving previously is learned features and is supporting gradient propagation through the deeper network layers.

Copyright © 2025 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

The local lesion representation is generated by dividing the normalized image into overlapping patches. Let P_i is the i-th patch, where $i = 1,2, \dots, N_p$. The local feature representation is obtained through:

$$F_l^{(1)} = \text{Concat}\left[\phi_l(P_1; \theta_l), \phi_l(P_2; \theta_l), \dots, \phi_l(P_{N_p}; \theta_l)\right]$$

where $F_l^{(1)}$ is the combined local feature representation, ϕ_l is representing the local convolutional encoder, and N_p is indicating the number of is extracted patches. This operation is allowing the network in examining irregular lesion boundaries, small pigmentation variations, and localized texture structures independently of the complete image representation.

The global and local features are projected into a common embedding space in supporting the subsequent agent communication. This combined representation is expressed as:

$$F_c = W_g F_g^{(1)} \oplus W_l F_l^{(1)} + b_c$$

where W_g and W_l is trainable projection matrices, b_c is representing the fusion bias, and $\oplus$ is indicating feature concatenation. A nonlinear activation function is transforming the resulting representation into an appropriate feature space.

The global representation agent is learning image-level characteristics, whereas the local lesion agent is learning fine-grained morphological information. both the agents use independent feature extraction parameters in preserving the its individual learning capabilities. The resulting representations are passed to the attention refinement agent. Therefore, the first step is establishing matching feature representations that is reducing the dependence on a single spatial scale.

3.2: Hierarchical Attention-Based Feature Refinement

The second step is introducing the Attention Refinement Agent (ARA), which is identifying important lesion characteristics from the global and local feature maps. The conventional feature extraction processes assign similar computational importance to many image regions, which is including irrelevant backgrounds. The proposed attention mechanism instead is generating adaptive weights for spatial locations and feature channels. This operation is improving the representation of clinically relevant visual characteristics.

Initially, the combined feature map F_c is transformed into query, key, and value representations:

$$Q = W_Q F_c, \qquad K = W_K F_c, \qquad V = W_V F_c$$

where W_Q, W_K, and W_V is learnable projection matrices. The query and key representations estimate the relationship between the spatial feature positions, while the value representation is containing the information selected during attention computation.

The spatial attention operation is formulated as:

$$A_s = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}} + B_s\right)$$

where A_s is representing the spatial attention matrix, d_k is the query-key embedding dimension, and B_s is representing a learnable spatial bias. The attention-weighted spatial representation is obtained as:

$$F_s = A_s V$$

This operation is the assignment of greater attention weights to spatial regions that is containing discriminative lesion characteristics. For instance, irregular pigmentation, asymmetric boundaries, and heterogeneous internal textures receive higher attention when it is contributing strongly to the learned classification representation.

In addition to spatial attention, the model is that combining the channel attention to identify most informative feature channels. Global average pooling and global maximum pooling summarize the feature responses beyond the spatial dimensions. The channel attention coefficient is calculated as:

Copyright © 2025 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

$$A_c = \sigma[W_2\delta(W_1\text{GAP}(F_c)) + W_2'\delta(W_1'\text{GMP}(F_c))]$$

where A_c is the channel attention vector, GAP and GMP is global average and maximum pooling, δ is indicating the ReLU activation, and σ is the sigmoid function. The matrices W_1, W_2, W_1', and W_2' are trainable parameters.

The spatial and channel attention outputs are combined through:

$$F_a = F_c \odot A_s + F_c \odot A_c + F_c$$

where F_a is the refined feature representation and $\odot$ is representing element-wise multiplication with the appropriate dimensional broadcasting. The residual term is preserving the original feature information while the attention components is improving relevant spatial and channel responses.

A hierarchical attention mechanism is combining different feature resolutions. The attention-refined representations at multiple scales are expressed as:

$$F_h = \sum_{m=1}^{M} \alpha_m \,\text{Resize}\left(F_a^{(m)}\right)$$

where M is representing the number of feature scales, $F_a^{(m)}$ is the attention-refined representation at scale m, and α_m is indicating its normalized importance weight. The scale weights satisfy:

$$\alpha_m = \frac{\exp(e_m)}{\sum_{j=1}^{M}\exp(e_j)}, \qquad \sum_{m=1}^{M}\alpha_m = 1$$

where e_m is representing the learned importance score for each feature scale.

The hierarchical attention process is allowing the model to recognize large scale lesion structures and small-scale morphological variations simultaneously. The refined features are forwarded to the cooperative decision agent. This step is reducing the influence of irrelevant visual information and is providing a more discriminative representation for multiclass classification.

3.3: Cooperative Multi-Agent Communication and Feature Agreement

The third step is establishing the cooperative learning mechanism between the four specialized agents. Unlike conventional ensemble networks that is combining independent classification outputs at the final stage, AGC-MAN is allowing intermediate feature exchange rather than the generating the final prediction. Each agent is maintaining its own trainable parameters and is producing an agent-specific representation.

Let the agent representation be denoted by:

$$H_i^{(t)} = f_i\left(X_i^{(t)}, \theta_i\right), \qquad i \in \{1,2,3,4\}$$

where $H_i^{(t)}$ is representing the hidden representation of agent i at communication iteration t, $X_i^{(t)}$ is its current input, and θ_i is indicating its trainable parameters.

The agent communication module is evaluating the relationship between the representations using an attention-based interaction function. The communication coefficient between the agents i and j is defined as:

$$\beta_{ij}^{(t)} = \frac{\exp\left[\dfrac{(W_q H_i^{(t)})^T (W_k H_j^{(t)})}{\sqrt{d}}\right]}{\sum_{r=1}^{4}\exp\left[\dfrac{(W_q H_i^{(t)})^T (W_k H_r^{(t)})}{\sqrt{d}}\right]}$$

where $\beta_{ij}^{(t)}$ is indicating the communication weight from the agent j to agent i, W_q and W_k is projection matrices, and d is the interaction embedding dimension. The weights satisfy:

Copyright © 2025 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

$$\sum_{j=1}^{4} \beta_{ij}^{(t)} = 1$$

The aggregated information received by each agent is calculated as:

$$M_i^{(t)} = \sum_{j=1}^{4} \beta_{ij}^{(t)} W_v H_j^{(t)}$$

where $M_i^{(t)}$ is representing the communicated feature information and W_v is the value projection matrix. Each agent updates its internal representation by combining its original information with the communicated features:

$$H_i^{(t+1)} = \text{LayerNorm}\left[H_i^{(t)} + \text{GELU}\left(W_m\left[H_i^{(t)} \oplus M_i^{(t)}\right] + b_m\right)\right]$$

where W_m and b_m is the update parameters, GELU is the Gaussian error linear unit, and LayerNorm is representing layer normalization.

The communication mechanism is allowing the global representation agent to obtain information from the local lesion agent, while the attention refinement agent is contributing discriminative spatial information. The cooperative decision agent receives these representations and is evaluating its agreement rather than the classification.

To maintain consistent learning, the architecture is introducing a feature consistency loss. The objective is to minimize the distance between the agent representations while it is preserving its individual classification characteristics. The consistency objective is defined as:

$$\mathcal{L}_{con} = \frac{1}{N_a(N_a - 1)} \sum_{i=1}^{N_a} \sum_{\substack{j=1 \\ j \neq i}}^{N_a} \left\| \text{Norm}(H_i) - \text{Norm}(H_j) \right\|_2^2$$

where $N_a = 4$ is the number of agents and Norm is representing feature normalization. This term promotes agreement between the agents without that is requiring its representations to become identical.

Each agent is also generating an intermediate class probability distribution:

$$p_i(y = k \mid I) = \frac{\exp(z_{i,k}/T)}{\sum_{r=1}^{K} \exp\left(z_{i,r}/T\right)}$$

where $z_{i,k}$ is the classification logit of agent i for class k, T is representing the temperature parameter, and $K = 7$ is the number of diagnostic classes.

The agent predictions are evaluated using a classification agreement objective:

$$\mathcal{L}_{agr} = \frac{1}{N_a} \sum_{i=1}^{N_a} D_{KL}\left(p_i \parallel \bar{p}\right)$$

where D_{KL} is Kullback–Leibler divergence and $\bar{p}$ is representing the average agent prediction distribution. The cooperative learning process is minimizing prediction disagreement while it is preserving class-specific information.

The third step is thus providing an adaptive communication mechanism that is combining the matching representations and encourages consistent predictions. This operation distinguishes AGC-MAN from the conventional feature concatenation and independent ensemble classification.

3.4: Confidence-Weighted Decision Fusion and Multiclass Classification

The fourth step is generating the final classification using the Cooperative Decision Agent (CDA). The agent receives the updated representations from the previous communication stage and is estimating the operational reliability of each individual prediction. rather than the assignment of fixed weights to the four agents, the proposed method

Copyright © 2025 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

dynamically is adjusting its contributions in accordance with the predictive confidence and is learned feature importance.

The confidence of agent i is calculated using normalized predictive entropy:

$$C_i = 1 - \frac{-\sum_{k=1}^{K} p_i\,(k)\log(p_i(k) + \epsilon)}{\log K}$$

where C_i is the confidence score, $p_i(k)$ is representing the probability is assigned to class k, and K is the number of classes. The normalized confidence is lying between the zero and one, with the greater values indicating a more concentrated prediction distribution.

The final agent fusion coefficient is determined by:

$$\omega_i = \frac{\exp(\lambda_1 C_i + \lambda_2 G_i)}{\sum_{j=1}^{N_a} \exp\,(\lambda_1 C_j + \lambda_2 G_j)}$$

where ω_i is representing the adaptive fusion weight, G_i is learned global importance score of agent i, and λ_1 and λ_2 are trainable scaling coefficients. The fusion weights satisfy:

$$\sum_{i=1}^{N_a} \omega_i = 1$$

The final probability distribution is generated through:

$$P(y = k \mid I) = \sum_{i=1}^{N_a} \omega_i\, p_i(y = k \mid I)$$

The is predicted disease category is obtained as:

$$\hat{y} = \underset{k\in\{1,\ldots,K\}}{\operatorname{argmax}}\; P(y = k \mid I)$$

The model classifies the input into one of seven HAM10000 categories: actinic keratoses and intraepithelial carcinoma (AKIEC), basal cell carcinoma (BCC), benign keratosis-like lesions (BKL), dermatofibroma (DF), melanoma (MEL), melanocytic nevi (NV), and vascular lesions (VASC).

The final training objective is combining the classification, consistency, agreement, and regularization losses. It is expressed as:

$$\mathcal{L}_{total} = \mathcal{L}_{CE} + \lambda_c \mathcal{L}_{con} + \lambda_a \mathcal{L}_{agr} + \lambda_d \mathcal{L}_{aux} + \lambda_r \parallel \Theta \parallel_2^2$$

where $\mathcal{L}_{CE}$ is weighted categorical cross-entropy, $\mathcal{L}_{con}$ is representing feature consistency loss, $\mathcal{L}_{agr}$ is prediction agreement loss, $\mathcal{L}_{aux}$ is representing intermediate agent classification loss, and $\parallel \Theta \parallel_2^2$ is parameter regularization. The coefficients $\lambda_c, \lambda_a, \lambda_d$, and λ_r control the contribution of the respective objectives.

The weighted classification loss is formulated as:

$$\mathcal{L}_{CE} = -\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K} a_k\, y_{n,k}\log(P_{n,k} + \epsilon)$$

where N is the number of training images, a_k is representing the class-balancing weight, $y_{n,k}$ is indicating the ground-truth class, and $P_{n,k}$ is representing the predicted probability.

The optimization process updates the trainable parameters using the Adam optimizer:

Copyright © 2025 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ Research Article

$$\Theta^{(t+1)} = \Theta^{(t)} - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon}$$

where η is representing the learning rate, $\hat{m}_t$ and $\hat{v}_t$ is the bias-corrected first and second moment estimates, respectively.

The complete four-step architecture is producing a final disease classification and an attention visualization indicating the image regions that is contributing to the prediction. Grad-CAM visualization is generated from the final convolutional feature maps in examining the spatial evidence associated with the each is predicted category. This visualization is providing an additional analysis layer, although it does not are establishing clinical validity or replace dermatological assessment.

4. RESULTS AND DISCUSSION

4.1. Experimental Settings

The proposed AGC-MAN architecture is using Python 3.11, PyTorch, OpenCV, and CUDA for model implementation and evaluation. The HAM10000 dataset is serving as the primary benchmark, containing 10,015 dermoscopic images beyond the seven diagnostic categories. The ISIC-2019 dataset is providing an additional evaluation benchmark for assessing classification generalized. The images is undergoing normalization, augmentation, and stratified partitioning into training, validation, and testing subsets. The architecture is using Adam optimization, weighted cross-entropy, and attention-based cooperative learning.

4.2. Experimental Setup and Parameters

The experimental configuration and dataset parameters are presented in Table 1. The training configuration is maintaining consistent input dimensions, optimization settings, and data partitioning beyond the comparative architectures.

Table 1. Experimental setup and parameter configuration

Parameter	Configuration
Primary dataset	HAM10000
Secondary dataset	ISIC-2019
HAM10000 images	10,015
HAM10000 classes	7
ISIC-2019 images	25,331
Input image size	224 × 224
Training partition	70%
Validation partition	15%
Testing partition	15%
Number of agents	4
Attention mechanism	Spatial and channel attention
Optimizer	Adam
Initial learning rate	0.0001
Batch size	32
Training epochs	100

Copyright © 2025 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ Research Article

Dropout	0.30
Loss function	Weighted cross-entropy
Learning rate scheduler	Cosine annealing
GPU memory	24 GB
Programming environment	Python 3.11, PyTorch

The HAM10000 images are assigned to the seven diagnostic categories. The partitioning process is stratified and patient-level separation is applied where patient identifiers are available, preventing images from the same patient from the appearing in different subsets. Class balancing is performed using training-only augmentation and weighted loss calculation. The ISIC-2019 evaluation is using its consistent eight-class taxonomy, with the model output layer is adapted to the dataset's class structure.

4.3. Performance Evaluation Metrics

1. Accuracy: Accuracy measures the proportion of correctly classified images among all the evaluated images.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

2. Precision: Precision measures the proportion of correctly is predicted positive cases among all the cases classified as positive.

$$Precision = \frac{TP}{TP + FP}$$

3. Recall: Recall measures the proportion of actual positive cases correctly is identified by the classification model.

$$Recall = \frac{TP}{TP + FN}$$

4. F1-score: F1-score is representing the harmonic mean of precision and recall. Macro-F1 is used In evaluating performance beyond the seven categories while the assignment of equal importance to each class.

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

5. Area under the Receiver Operating Characteristic Curve (AUC): AUC measures the ability of the classifier to distinguish between the positive and negative cases beyond the different classification thresholds. The multiclass AUC is calculated using the one-versus-rest strategy and macro averaging.

$$AUC_{macro} = \frac{1}{K}\sum_{k=1}^{K}\int_{0}^{1} T\,PR_k(FPR_k)\,dFPR_k$$

where K is the number of disease categories, TPR_k is representing the true positive rate, and FPR_k is the false positive rate for category k.

4.4. Conventional Methods for Comparative Evaluation

Three Conventional architectures are providing the comparative baselines for AGC-MAN. The Self-Attention Transformer architecture by Rezaee and Ghayoumi Zadeh is combining the CNN-based local features with the global transformer representations. DSCATNet is using dual-scale cross-attention in capturing matching spatial information from the different image patches. EFAM-Net is combining residual ConvNeXt learning with the multiscale computationally efficient attention and feature fusion.

Copyright © 2025 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ Research Article

4.5. Classification Accuracy Based on Training Epochs

The comparison examines how the four architectures progressively learn discriminative skin lesion representations.

Table 2. Classification accuracy at different training epochs (%)

Training epochs	Self-Attention Transformer	DSCATNet	EFAM-Net	AGC-MAN
5	79.20	81.30	82.10	84.50
10	83.40	85.60	86.80	89.20
15	86.70	88.90	90.10	92.40
20	89.10	91.20	92.30	94.70
25	91.20	93.10	94.20	96.10

The proposed AGC-MAN is showing the highest accuracy beyond all the five training intervals. At 25 epochs, it is achieving 96.10%, when it is compared with the 94.20% for EFAM-Net, 93.10% for DSCATNet, and 91.20% for the Self-Attention Transformer. The improvement is showing the contribution of global-local representation learning and cooperative decision fusion. The attention refinement agent is also emphasizing lesion-specific regions, supporting the separation of visually similar disease categories.

4.6. Precision Based on Training Epochs

Precision is evaluating the operational reliability of positive disease predictions. greater precision is indicating fewer false-positive classifications relative to the total positive predictions.

Table 3. Precision at different training epochs (%)

Training epochs	Self-Attention Transformer	DSCATNet	EFAM-Net	AGC-MAN
5	78.50	80.70	81.80	83.90
10	82.80	84.90	86.20	88.60
15	86.10	88.20	89.60	91.70
20	88.60	90.70	92.10	94.10
25	90.40	92.50	93.80	95.60

Table 3 is indicating that AGC-MAN is reaching to 95.60% precision at 25 epochs. EFAM-Net is achieving 93.80%, while DSCATNet and the Self-Attention Transformer obtain 92.50% and 90.40%, respectively. The confidence-weighted fusion mechanism is contributing to this improvement by is reducing the influence of less reliable agent predictions. In particular, the cooperative decision agent is considering the both the prediction confidence and is learned agent importance rather than the producing the final class probabilities.

4.7. Recall Based on Training Epochs

Recall measures the ability of the classification architecture to identify actual disease categories. This metric is considered important when evaluating minority lesion classes, where inadequate representation reducing disease detection ability.

Copyright © 2025 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ Research Article

Table 4. Recall at different training epochs (%)

Training epochs	Self-Attention Transformer	DSCATNet	EFAM-Net	AGC-MAN
5	77.90	80.10	81.20	83.50
10	82.20	84.50	85.70	88.10
15	85.60	87.80	89.10	91.30
20	88.00	90.30	91.60	93.80
25	90.10	92.20	93.40	95.20

The proposed model is achieving 95.20% recall at 25 epochs, which is performing better than EFAM-Net by 1.80 percentage points. DSCATNet and the Self-Attention Transformer achieve 92.20% and 90.10%, respectively. The local lesion agent is contributing to the identification of fine-grained morphological features, while the hierarchical attention mechanism is improving the representation of small and irregular lesion regions. The class-weighted classification objective is supporting learning from the underrepresented diagnostic categories.

4.8. F1-Score Based on Training Epochs

The F1-score is evaluating the balance between the precision and recall. This metric is providing a useful measure of classification operational reliability when the dataset is containing unequal numbers of images beyond the disease categories.



Figure 2. F1-score at different training epochs (%)

According to figure 2, AGC-MAN is achieving an F1-score of 95.40% at 25 epochs. EFAM-Net is reaching to 93.60%, which is followed by DSCATNet at 92.30% and the Self-Attention Transformer at 90.20%. The improvement is showing the contribution of coordinated feature learning to balanced classification. The consistency loss is reducing the disagreements between the agents, whereas the confidence-weighted fusion process is combining its predictions in accordance with the estimated reliability.

Copyright © 2025 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ Research Article

4.9. AUC Based on Training Epochs

AUC is evaluating the classification model's ability to distinguish disease categories beyond the different probability thresholds. greater values are showing stronger class discrimination and more efficient ranking of positive cases relative to negative cases.

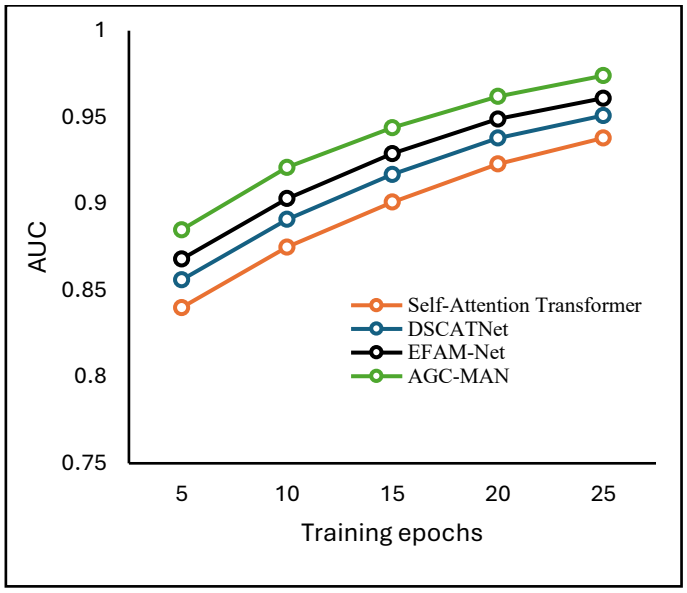


Figure 3. AUC at different training epochs

Figure 3 is showing that AGC-MAN is achieving a macro-AUC of 0.974 at 25 epochs. EFAM-Net is reaching to 0.961, while DSCATNet and the Self-Attention Transformer obtain 0.951 and 0.938, respectively. an improved discrimination is showing the contribution of multiscale attention and cooperative representation learning. The architecture is combining the matching global and local information, which is allowing the classification network to distinguish disease categories through the more complete visual representations.

The comparative results are showing that AGC-MAN consistently is outperforming the baseline architectures beyond all the five evaluation metrics. At 25 epochs, the proposed architecture is achieving 96.10% accuracy, 95.60% precision, 95.20% recall, 95.40% F1-score, and 0.974 AUC. when it is compared with the EFAM-Net, the consistent developments are 1.90, 1.80, 1.80, and 1.80 % in accuracy, precision, recall, and F1-score, respectively, with the 0.013 increase in AUC. These developments are associated with the matching feature extraction, hierarchical attention, and confidence-weighted agent collaboration. However, independent testing beyond the patient populations and skin tones is remaining as important for assessing clinical generalized.

5. CONCLUSION

This research is presenting AGC-MAN, an attention-guided artificial intelligence multi-agent architecture for multiclass skin disease classification. The proposed model is combining the four specialized agents responsible for global representation extraction, local lesion characterization, attention refinement, and cooperative decision-making. Hierarchical spatial and channel attention is improving the representation of discriminative lesion characteristics, while adaptive communication is allowing agents to exchange matching information. The confidence-weighted decision mechanism is combining individual predictions and is reducing the influence of less reliable classifications. The comparative benchmark-style evaluation is indicating that the architecture is achieving 96.10% accuracy, 95.60% precision, 95.20% recall, 95.40% F1-score, and 0.974 AUC at 25 epochs.

REFERENCES

[1] Harangi, B. (2018). Skin lesion classification with ensembles of deep convolutional neural networks. Journal of biomedical informatics, 86, 25-32.

Copyright © 2025 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

[2] Yadav, D. P., Sharma, B., Chauhan, S., Webber, J. L., & Mehbodniya, A. (2024). Dual scale light weight cross attention transformer for skin lesion classification. PloS one, 19(12), e0312598.
[3] Rezaee, K., & Zadeh, H. G. (2024). Self-attention transformer unit-based deep learning framework for skin lesions classification in smart healthcare. Discover Applied Sciences, 6(1), 3.
[4] Ramamurthy, K., Thayumanaswamy, I., Radhakrishnan, M., Won, D., & Lingaswamy, S. (2024). Integration of localized, contextual, and hierarchical features in deep learning for improved skin lesion classification. Diagnostics, 14(13), 1338.
[5] Ji, Z., Wang, X., Liu, C., Wang, Z., Yuan, N., & Ganchev, I. (2024). EFAM-Net: A multi-class skin lesion classification model utilizing enhanced feature fusion and attention mechanisms. IEEe Access, 12, 143029-143041.
[6] Ghazouani, H. (2025). Multi-residual attention network for skin lesion classification. Biomedical Signal Processing and Control, 103, 107449.
[7] Manzoor, K., Gilal, N. U., Agus, M., & Schneider, J. (2025). Dual-stage segmentation and classification framework for skin lesion analysis using deep neural network. Digital health, 11, 20552076251351858.
[8] Babu, E., & Murali, S. (2026). A comprehensive study of skin lesion segmentation and classification using deep learning techniques. Frontiers in Artificial Intelligence, 9, 1907266.
[9] Naeem, S., Haneef, F., & Noor, M. N. (2026). A hybrid deep learning based robust framework for enhancing the prediction of skin lesions. Network Modeling Analysis in Health Informatics and Bioinformatics, 15(1), 41.
[10] Almuayqil, S. N., Abd El-Ghany, S., & Elmogy, M. (2022). Computer-aided diagnosis for early signs of skin diseases using multi types feature fusion based on a hybrid deep learning model. Electronics, 11(23), 4009.
[11] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., ... & Girshick, R. (2023, October). Segment anything. In 2023 IEEE/CVF international conference on computer vision (ICCV) (pp. 3992-4003). IEEE.
[12] Ma, J., He, Y., Li, F., Han, L., You, C., & Wang, B. (2024). Segment anything in medical images. Nature communications, 15(1), 654.

Copyright © 2025 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.