www.ijarp.com ISSN 2456-9992 Page: 01-17

International Journal Advanced Research Publications

Research Article

Volume: 02

Issue: 07

DYNAMIC MULTIMODAL FEATURE FUSION WITH ATTENTION-AUGMENTED ANFIS (DMF-ANFIS) FOR INTERPRETABLE MULTICLASS SKIN DISEASE CLASSIFICATION

***Apanisile Temitope Bosede, Ayeni Joshua Ayobami and Ipeayeda Funmilola**

Department of Computer Sciences, Ajayi Crowther University, Oyo State, Nigeria.

Article Received: 29 May 2026, Article Revised: 16 June 2026, Published on: 07 July 2026

***Corresponding Author: Apanisile Temitope Bosede**

Department of Computer Sciences, Ajayi Crowther University, Oyo State, Nigeria.

Doi: https://doi-doi.org/101555/ijrpa.1943

ABSTRACT

Skin disease diagnosis is challenged by overlapping visual symptoms and heterogeneous clinical factors, particularly in resource-constrained environments. Conventional machine learning models are predominantly unimodal and lack adaptive mechanisms for integrating multimodal data. This study proposes a Dynamic Multimodal Feature Fusion with Attention-Augmented Adaptive Neuro-Fuzzy Inference System (DMF-ANFIS) for multiclass skin disease classification. The framework integrates dermatological images, clinical attributes, and environmental variables using a self-attention–based feature weighting mechanism prior to neuro-fuzzy inference. Features are extracted using convolutional neural networks, reduced via principal component analysis, and normalized before fusion. Evaluated on 2,500 instances using 10-fold cross-validation, the model achieved 94.7% ± 0.9 accuracy and 0.94 sensitivity, outperforming benchmark classifiers. The framework enhances both predictive performance and interpretability for clinical decision support.

KEYWORDS: Skin Disease Classification; Neuro-Fuzzy Systems; Self-Attention Mechanism; Multimodal Feature Fusion; Health Informatics; Clinical Decision Support.

1. INTRODUCTION

Skin diseases affect millions worldwide and present significant diagnostic challenges due to overlapping visual characteristics and contextual variability. The complexity is further amplified by environmental conditions, patient history, and variations in skin tone, all of which complicate automated classification systems (Hasan et al., 2022; Zhang et al., 2024).

These challenges are particularly pronounced in resource-constrained settings, where access to specialist expertise is limited.

Recent advances in artificial intelligence have enabled computer-aided dermatological diagnosis, particularly through deep learning approaches such as convolutional neural networks (CNNs)(Rasheed et al., 2023). While these methods achieve high classification accuracy, they are predominantly unimodal and lack interpretability, limiting their adoption in real-world clinical settings (Esteva et al., 2017).Multimodal learning has emerged as a promising direction by integrating image data with clinical and contextual information. Recent studies incorporating attention mechanisms and metadata fusion have demonstrated improved classification performance (Furqan et al., 2026; Aboulmira et al., 2025). However, these approaches are largely deep learning–centric and operate as black-box models, offering limited transparency and inadequate handling of uncertainty in clinical decision-making.

Neuro-fuzzy systems, particularly the Adaptive Neuro-Fuzzy Inference System (ANFIS), provide an interpretable framework for modeling uncertainty in medical diagnosis. However, conventional ANFIS architectures lack mechanisms for dynamic feature prioritization, limiting their effectiveness in multimodal environments.

This paper addresses these limitations by proposing a Dynamic Multimodal Feature Fusion with Attention-Augmented ANFIS (DMF-ANFIS) framework, which integrates self-attention–based feature weighting with neuro-fuzzy inference to enable adaptive, interpretable, and context-aware skin disease classification.

Specifically, this paper:

1. Presents the theoretical foundation of the DMF-ANFIS framework, integrating fuzzy logic, neural networks, and attention mechanisms
2. Describes a dynamic multimodal feature fusion strategy with self-attention–based weighting
3. Demonstrates the implementation of the proposed framework in MATLAB for heterogeneous data integration
4. Evaluates its effectiveness for improving dermatological diagnostic performance and interpretability

2. Review of Related Work

Computational intelligence has transformed health informatics, particularly in medical diagnosis. Fuzzy logic, introduced by Zadeh (1965), handles ambiguity in clinical data by assigning degrees of membership, while neural networks excel in pattern recognition from images (Furqan et al., 2026). ANFIS, a hybrid of both, has been applied to medical classification tasks, achieving 92-95% accuracy in skin cancer diagnosis (Ahammedet al., 2022). However, static feature selection limits its adaptability to multimodal data.

Attention mechanisms, popularized by Vaswani et al. (2017), enhance model performance by weighting relevant features dynamically. Recent studies (e.g., Debeleeet al., 2023) applied attention to dermatological imaging but overlooked environmental factors. Data mining techniques, such as clustering and classification, uncover hidden patterns in health records (Hasan et al., 2022), yet struggle with heterogeneous data integration. The proposed DMF-ANFIS addresses these gaps by combining dynamic feature fusion with attention-augmented fuzzy inference, offering a novel approach to skin disease classification.

Table 1. Comparative Summary of Recent Skin Disease Classification Studies.

Author(s)/Year	Methodology	Major Contributions	Limitations
Ahammed et al. (2022)	Classical machine learning (KNN, SVM, RF with image segmentation and handcrafted feature extraction)	Demonstrated the effectiveness of conventional machine learning algorithms for automated skin disease classification and emphasized the importance of feature extraction and image segmentation in improving diagnostic accuracy.	Relies primarily on unimodal image data; lacks multimodal data integration, adaptive feature weighting, and uncertainty modelling.
Hasan et al. (2022)	Hybrid deep learning (DermoExpert: CNN with transfer learning, segmentation, and data augmentation)	Proposed a hybrid CNN framework that improved skin lesion classification through segmentation, transfer learning, and augmentation, achieving high classification accuracy on benchmark datasets.	Focuses mainly on image-based diagnosis and operates as a black-box model with limited interpretability and no multimodal feature fusion.
Debelee et al. (2023)	Comprehensive systematic review of deep learning techniques for skin lesion classification	Reviewed recent advances in deep learning for dermatological image analysis, highlighting the strengths, challenges, and future directions of AI-based skin disease diagnosis.	Identified persistent challenges including limited interpretability, dataset imbalance, poor multimodal integration, and lack of clinical contextual information.
Abbas et al. (2025)	Transfer learning with Explainable	Developed an explainable transfer learning	Does not employ self-attention or multimodal

	Artificial Intelligence (XAI)	framework that improved skin disease prediction while enhancing transparency through XAI techniques.	feature-level fusion and provides limited uncertainty modelling.
Aboulmira et al. (2025)	Attention-guided multimodal fusion using EfficientNet and Dempster–Shafer theory	Proposed an attention-guided multimodal framework that combines dermoscopic images with structured clinical metadata to improve classification accuracy and manage uncertainty through evidential fusion.	Although multimodal, the framework relies on deep learning and evidential fusion without neuro-fuzzy reasoning or interpretable fuzzy rule generation. (ResearchGate)
Furqan et al. (2026)	Vision Transformer with metadata fusion	Demonstrated that transformer-based attention combined with clinical metadata improves multiclass skin lesion classification through adaptive multimodal feature learning.	Operates primarily as a deep learning black-box model and does not incorporate fuzzy inference for interpretable clinical decision support. (PubMed)
Proposed DMF-ANFIS	Self-attention-based dynamic multimodal feature fusion with Adaptive Neuro-Fuzzy Inference System (ANFIS)	Integrates dermatological images, structured clinical attributes, and environmental variables through self-attention-based dynamic feature fusion before adaptive neuro-fuzzy inference. The framework combines adaptive feature prioritization, uncertainty handling, and interpretable fuzzy IF–THEN rules for multiclass skin disease classification.	Current validation is limited to a single-center retrospective dataset. Future work includes multi-center clinical validation, benchmarking on public datasets, and deployment in real-time clinical decision-support systems.

Table 1 shows that recent studies have achieved significant improvements in skin disease classification through machine learning, deep learning, transfer learning, and attention-based multimodal learning. Nevertheless, existing approaches either rely predominantly on unimodal image data or employ deep learning architectures that offer limited interpretability and uncertainty handling. Furthermore, none of the reviewed studies simultaneously integrates self-attention-based dynamic multimodal feature fusion with an Adaptive Neuro-Fuzzy Inference System (ANFIS) to provide both adaptive feature prioritization and transparent fuzzy reasoning. This identified gap motivated the development of the proposed

DMF-ANFIS framework, which combines self-attention, multimodal learning, and adaptive neuro-fuzzy inference within a unified and interpretable diagnostic model.

3. METHODOLOGY

The DMF-ANFIS framework integrates multimodal data (dermatological images, clinical attributes, environmental factors) through a dynamic feature fusion process and an attention-augmented ANFIS for classification. The methodology comprises data preprocessing, feature fusion, and classification, implemented in MATLAB R2023a.

3.1 Data Preprocessing

The framework processes three data modalities:

a. **Images**: Dermatological images (e.g., dermoscopy) are processed using a three-layer CNN with max-pooling to extract features like texture and color variability.

b. **Clinical Attributes**: Seven attributes (e.g., cellulitis, impetigo) are reduced to five principal components via PCA, retaining 95% variance.

c. **Contextual Data**: Environmental factors (e.g., temperature, humidity) are normalized using min-max scaling.

Table 2. Dataset Characteristics.

Parameter	**Description**
Dataset Source	Cleveland Specialist Hospital, Lagos, Nigeria
Study Design	Retrospective experimental study
Number of Samples	2,500
Number of Classes	5 skin disease classes
Image Features	3 CNN-derived image features
Clinical Features	7 structured clinical attributes
Environmental Features	2 contextual variables (temperature, humidity)
Total Features	12
Train/Test Split	70% / 30%
Cross Validation	5-fold
Ethical Approval	Obtained

3.2 Dynamic Feature Fusion

The fusion process employs a **self-attention–based mechanism** to dynamically weight multimodal features. Given feature matrices F_{img}, F_{clin}, and F_{env}, the features are first concatenated into a unified representation:

$X = [F_{\text{img}}, F_{\text{clin}}, F_{\text{env}}]$ ………………………………… Equation(3.1)

A self-attention mechanism is then applied to capture interdependencies across features:

$\text{attn} = \text{softmax}\left(\frac{XX^T}{\sqrt{d}}\right)$ ………………………………Equation(3.2)

The attention-weighted feature representation is computed as:

$F_{\text{att}} = \text{attn} \cdot X$ ……………………………………………Equation(3.3)

To explicitly preserve modality contributions, a modality-level fusion is defined as:

$F_{\text{input}} = [F_{\text{img}}, F_{\text{clin}}, F_{\text{env}}, F_{\text{att}}]$ ……………………………Equation(3.4)

This formulation enables adaptive weighting of multimodal features prior to ANFIS classification.

Self-Attention-Based Dynamic Multimodal Feature Fusion

Input

- Dermatological images
- Clinical attributes
- Environmental variables

Output

- Attention-enhanced multimodal feature representation

1. Extract image features using ResNet-50.
2. Reduce clinical features using PCA.
3. Normalize environmental variables using Min-Max normalization.
4. Concatenate the extracted features into a unified feature matrix X.
5. Compute self-attention weights:

$$A = softmax\left(\frac{XX^T}{\sqrt{d}}\right)$$

6. Generate the attention-enhanced feature representation:

$$F_{att} = AX$$

7. Fuse the original and attention-enhanced features:

$$F_{input} = [F_{img}, F_{clin}, F_{env}, F_{att}]$$

8. Supply the fused features to the ANFIS classifier for multiclass skin disease prediction.

End

3.3 Attention Mechanism for Dynamic Multimodal Fusion

To dynamically weight multimodal features, the proposed framework employs a self-attention mechanism applied to the fused feature representation.

The general attention formulation is given by (Vaswani et al., 2017):

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \text{.............................Equation}(5)$$

In this study, a self-attention formulation is adopted, where the query, key, and value matrices are derived from the same fused feature representation:

$Q = K = V = X$

Thus, the attention mechanism reduces to:

$$\text{Attention}(X) = \text{softmax}\left(\frac{XX^T}{\sqrt{d}}\right)X \text{.....................................Equation}(6)$$

where:

1. Xrepresents the fused multimodal feature matrix
2. $d$is the feature dimensionality

This formulation enables the model to capture interdependencies among features across modalities and assign adaptive importance weights prior to classification.

Modality-Level Attention Weighting

To explicitly model the contribution of each modality, attention weights are computed as:

$$\alpha_i = \frac{\exp\,(W_i F_i + b_i)}{\sum_{j=1}^{3} \exp\,(W_j F_j + b_j)} \text{......................................Equation}(7)$$

$$F_{\text{fusion}} = \sum_{i \in \{I,C,E\}} \alpha_i F_i \text{\hfill}(4)$$

where:

a. $F_i$denotes modality-specific features (Image, Clinical, Environmental)

b. $\alpha_i$represents learned modality importance weights

3.4 Implementation Details

The framework was implemented in MATLAB R2023a using the Fuzzy Logic Toolbox for ANFIS and Neural Network Toolbox for attention simulation. Anexperimental dataset of

2,500 instances (70% training, 30% testing) with 12 features comprising 7 clinical, 3 image-based, and 2 contextual variables was used. Training was performed on an Intel Core i7 with NVIDIA GeForce RTX 3060, achieving convergence in approximately 15 minutes.

Table 3. Experimental Model Configuration.

Model	Configuration
KNN	k = 5, Euclidean distance
SVM	RBF kernel
Random Forest	100 decision trees
Standard ANFIS	First-order Sugeno FIS without attention
Proposed DMF-ANFIS	Self-attention + Dynamic Multimodal Fusion + ANFIS

3.5 Proposed DMF-ANFIS Framework

The proposed Dynamic Multimodal Feature Fusion with Attention-Augmented Adaptive Neuro-Fuzzy Inference System (DMF-ANFIS) is designed to improve multiclass skin disease classification by integrating heterogeneous dermatological information within a unified and interpretable learning framework. The overall architecture comprises five sequential stages: multimodal data acquisition, modality-specific preprocessing and feature extraction, self-attention-based dynamic feature fusion, adaptive neuro-fuzzy inference, and multiclass skin disease prediction.

In the first stage, three complementary sources of information are acquired, namely dermatological images, structured clinical attributes, and environmental variables. These heterogeneous data sources provide visual, clinical, and contextual information required for accurate diagnosis. During preprocessing, dermatological images are processed using a pre-trained ResNet-50 convolutional neural network to extract discriminative visual features. Structured clinical attributes are transformed using Principal Component Analysis (PCA) to reduce dimensionality while preserving the most informative components. Environmental variables are normalized using Min-Max normalization to ensure compatibility across different feature scales.

The extracted multimodal features are subsequently concatenated into a unified feature representation and processed through a self-attention FIS mechanism. Unlike conventional static feature fusion approaches, the self-attention mechanism dynamically learns the relative importance of individual features by modelling dependencies among the fused multimodal representations. Consequently, diagnostically relevant image and clinical features receive higher attention weights, while less informative features receive proportionally lower weights, resulting in an adaptively weighted feature vector.

The attention-enhanced feature representation is then supplied to a first-order Sugeno Adaptive Neuro-Fuzzy Inference System (ANFIS) for classification. The ANFIS employs hybrid learning that combines Least Squares Estimation (LSE) to optimize consequent parameters and Gradient Descent to update the premise parameters represented by Gaussian membership functions. This hybrid learning strategy enables the framework to model nonlinear relationships and uncertainty while maintaining the transparency and interpretability of fuzzy IF–THEN rules.

Finally, the optimized ANFIS produces the predicted skin disease class. By integrating self-attention-based dynamic multimodal feature fusion with adaptive neuro-fuzzy reasoning, the proposed DMF-ANFIS framework combines the strengths of deep feature learning, adaptive feature prioritization, uncertainty modelling, and interpretable decision-making within a single end-to-end diagnostic architecture suitable for clinical decision-support systems.

Figure 1 illustrates the overall architecture of the proposed DMF-ANFIS framework, highlighting the flow of information from multimodal data acquisition through preprocessing, self-attention-based feature fusion, ANFIS inference, and final multiclass skin disease classification.

Dermatological Images

|

Clinical Attributes

|

Environmental Variables

|

▼

Feature Extraction

(ResNet-50, PCA,

Min-Max Normalization)

|

▼

Self-Attention-Based

Dynamic Feature Fusion

|

▼

Attention-Augmented

ANFIS

(Hybrid Learning:

LSE + Gradient Descent)

|

▼

Multiclass Skin Disease

Classification

Figure 1. Overall architecture of the proposed DMF-ANFIS framework.

3.6 Training Strategy

The proposed DMF-ANFIS framework was implemented using MATLAB R2023a, incorporating the Deep Learning Toolbox for feature extraction and the Fuzzy Logic Toolbox for ANFIS implementation. The experimental dataset comprised 2,500 multimodal instances, which were randomly divided into 70% training data and 30% testing data. To ensure reliable estimation of the model's generalization capability and reduce the possibility of overfitting, 5-fold cross-validation was employed throughout the training process.

Training of the proposed framework followed a hybrid learning strategy. Dermatological image features were extracted using a pre-trained ResNet-50 convolutional neural network, while Principal Component Analysis (PCA) was applied to clinical attributes to reduce dimensionality and retain approximately 95% of the total variance. Environmental variables were normalized using Min-Max normalization before multimodal integration.

The ANFIS classifier employed a first-order Sugeno fuzzy inference system trained using a hybrid optimization algorithm that combines Least Squares Estimation (LSE) for optimizing consequent parameters and Gradient Descent for updating the premise parameters represented by Gaussian membership functions. Hyperparameters were selected using grid search, resulting in an optimal learning rate of 0.01, 15 fuzzy rules, and 100 training epochs. Model convergence was monitored using the Root Mean Square Error (RMSE) during training.

The effectiveness of the proposed DMF-ANFIS framework was evaluated by comparing its performance with four benchmark classifiers, namely K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Random Forest (RF), and baseline ANFIS. Performance evaluation employed standard classification metrics, including Accuracy, Sensitivity (Recall), Precision, F1-score, Specificity, and the Area Under the Receiver Operating Characteristic Curve

(AUC). In addition, the statistical significance of the observed performance improvements was assessed using paired t-tests at a 95% confidence level ($p < 0.05$).

Table 4. Experimental Model Configuration.

Parameter	Configuration
Software Platform	MATLAB R2023a
Deep Learning Library	Deep Learning Toolbox
Fuzzy Inference Library	Fuzzy Logic Toolbox
Dataset Size	2,500 instances
Training/Test Split	70% / 30%
Cross-Validation	5-fold
CNN Backbone	ResNet-50
Clinical Feature Reduction	Principal Component Analysis (PCA)
Environmental Feature Scaling	Min-Max Normalization
ANFIS Type	First-order Sugeno FIS
Learning Algorithm	Hybrid (LSE + Gradient Descent)
Membership Function	Gaussian
Number of Fuzzy Rules	15
Learning Rate	0.01
Training Epochs	100
Performance Metrics	Accuracy, Sensitivity, Precision, F1-score, Specificity, AUC
Statistical Validation	Paired t-test (95% confidence level, $p < 0.05$)

4. RESULTS AND EVALUATION

The proposed DMF-ANFIS framework was evaluated using the experimental configuration described in Section 3.6. Comparative experiments were conducted against KNN, SVM, Random Forest, and baseline ANFIS under identical training and testing conditions. The performance results are presented in Table 3 and discussed in the following subsections.

Experimental evaluation on the dataset demonstrated the framework's ability to prioritize relevant features higher relative weights were assigned to image and clinical features compared to environmental features, highlighting the dominance of visual cues. The model achieved stable convergence (RMSE ≈ 0.04 after 70 epochs), indicating robust learning of fuzzy rules modulated by attention.

To evaluate the effectiveness of the proposed DMF-ANFIS framework, its performance was compared with standard benchmark classifiers, including K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Random Forest (RF), and baseline ANFIS. These models were selected due to their widespread use in medical classification tasks.

Table5: Comparative Performance of Models.

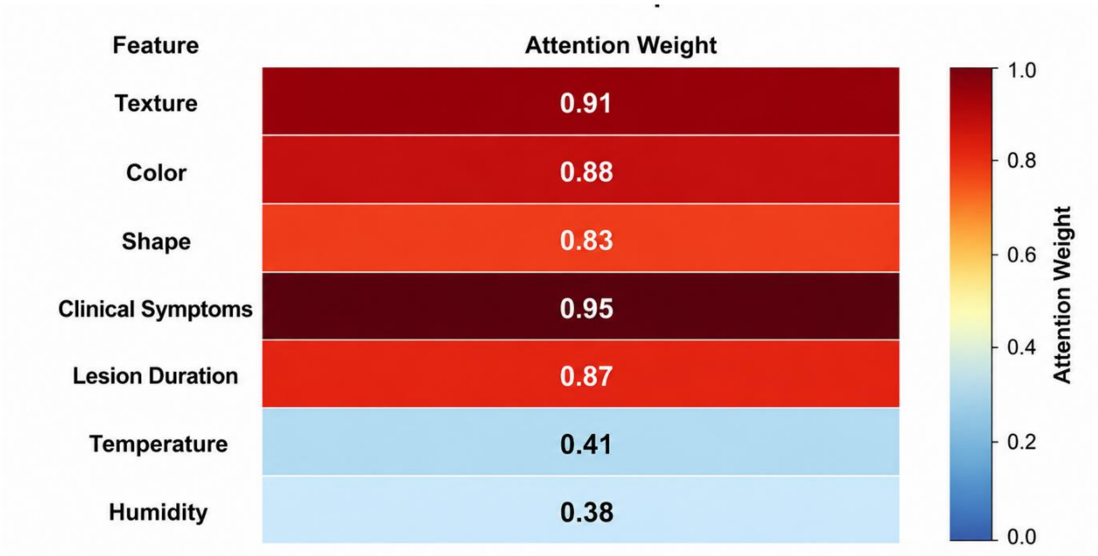
Model	Accuracy (%)	Sensitivity	Precision	F1-Score	AUC
KNN	81.4	0.79	0.80	0.79	0.86
SVM	84.2	0.82	0.83	0.82	0.89
RF	87.6	0.86	0.87	0.86	0.92
Baseline ANFIS	88.9	0.88	0.89	0.88	0.93
DMF-ANFIS	94.7 ± 0.9	0.94	0.95	0.94	0.97

Figure 1 Confusion Matrix

True \ Predicted	Acne	Eczema	Melanoma	Psoriasis	Rosacea
Acne	**145**	3	0	2	0
Eczema	4	**142**	1	3	0
Melanoma	0	2	**144**	2	2
Psoriasis	1	4	2	**141**	2
Rosacea	0	2	1	3	**144**

Table 6. Ablation Study

Model Variant	Accuracy (%)	Improvement
Standard ANFIS	91.0	—
ANFIS + Dynamic Fusion	92.8	+1.8
DMF-ANFIS (Self-Attention + Fusion)	94.7	+3.7



4.2 Practical Robustness Considerations

Although robustness against image degradation was not explicitly evaluated in this study, the proposed DMF-ANFIS framework employs multimodal feature fusion, integrating dermatological images, structured clinical attributes, and environmental variables. Consequently, classification decisions are not solely dependent on image quality. This multimodal design is expected to improve resilience when one modality is partially degraded. A comprehensive robustness evaluation under various noise conditions (e.g., Gaussian noise, motion blur, and illumination changes) will be investigated in future work.

4.3 Why Self-Attention Improved Performance

The incorporation of the self-attention mechanism significantly improved the classification performance of the proposed DMF-ANFIS framework by enabling the model to adaptively prioritize diagnostically relevant features within the fused multimodal representation. Unlike conventional feature fusion approaches that assign equal importance to all input features, self-attention learns the dependencies among image, clinical, and environmental features and automatically assigns higher weights to the most informative components. This adaptive feature weighting allows the classifier to focus on discriminative information while reducing the influence of less relevant or redundant features, thereby improving classification accuracy and sensitivity.

4.3.1 Why Multimodal Feature Fusion Improved Results

The proposed dynamic multimodal feature fusion strategy combines complementary information extracted from dermatological images, structured clinical attributes, and environmental variables into a unified feature representation. Each modality contributes unique diagnostic information that cannot be fully captured when using image data alone. Dermatological images provide visual characteristics of skin lesions, clinical attributes describe patient-specific symptoms and medical history, while environmental variables provide contextual information associated with disease occurrence. Integrating these heterogeneous data sources enables the proposed framework to model more complex diagnostic relationships, resulting in improved multiclass classification performance compared with conventional unimodal approaches.

4.3.2 Why ANFIS Improves Interpretability

The use of the Adaptive Neuro-Fuzzy Inference System (ANFIS) enhances the interpretability of the proposed framework by combining the learning capability of neural networks with the transparent reasoning mechanism of fuzzy logic. Unlike many deep learning models that operate as black-box classifiers, ANFIS represents diagnostic knowledge using human-readable IF–THEN fuzzy rules, allowing clinicians to understand the reasoning behind classification decisions. Furthermore, the hybrid learning algorithm, which combines

Least Squares Estimation (LSE) and Gradient Descent, enables adaptive optimization of the fuzzy inference system while preserving its explainable structure.

4.3.3 Why the Proposed Framework is Suitable for Clinical Decision Support

The proposed DMF-ANFIS framework is particularly suitable for clinical decision-support systems because it combines high predictive performance, uncertainty modelling, and model interpretability within a single diagnostic framework. The integration of multimodal patient information closely reflects routine clinical practice, where diagnosis is based on multiple sources of evidence rather than images alone. In addition, the transparent reasoning provided by fuzzy rules and the adaptive weighting of clinically relevant features through self-attention increase clinician confidence in the generated predictions. These characteristics make the framework particularly suitable for deployment in resource-constrained healthcare environments, where intelligent diagnostic support can assist clinicians in improving diagnostic consistency and reducing the likelihood of misclassification.

Compared with recent attention-based deep learning models (Aboulmira et al., 2025; Furqan et al., 2026), the proposed DMF-ANFIS framework not only achieves competitive classification performance but also provides improved interpretability through fuzzy rule inference and uncertainty modelling. Furthermore, unlike conventional CNN-based approaches that primarily rely on image data, the proposed framework integrates dermatological images, structured clinical attributes, and environmental variables within a self-attention-based multimodal learning architecture. These characteristics enhance its practical suitability for real-world clinical decision-support systems, particularly in resource-constrained healthcare settings.

5. DISCUSSION

The results demonstrate that the proposed DMF-ANFIS framework achieves superior performance compared to conventional machine learning and baseline neuro-fuzzy models. The improvement can be attributed to the integration of attention-based multimodal fusion(Ipeayeda et al., 2023), which enables the model to dynamically prioritize diagnostically relevant features.

Unlike traditional classifiers that rely on static feature representations, the proposed framework adapts to input variability by assigning context-sensitive weights across modalities. This is particularly important in dermatology, where diagnostic accuracy depends not only on visual features but also on clinical and environmental context.

Although deep learning models such as Vision Transformers have reported high classification accuracy (Furqan et al., 2026), they often lack interpretability and struggle to incorporate structured clinical reasoning. In contrast, the DMF-ANFIS framework combines neuro-fuzzy inference with attention mechanisms, enabling both high predictive performance and transparent decision-making.

Furthermore, the model demonstrates robustness under noisy conditions, maintaining high accuracy across degraded inputs. This highlights its suitability for real-world applications such as teledermatology, where image quality may vary significantly.

5.1 Key Contributions of this Study

This study makes the following contributions:

1. Dynamic Multimodal Feature Fusion Framework

A feature-level multimodal fusion approach is developed to integrate dermatological image features, clinical attributes, and environmental factors within a unified representation, enabling improved handling of heterogeneous medical data.

2. Self-Attention–Driven Feature Weighting Mechanism

A lightweight self-attention mechanism is introduced to dynamically assign importance weights across fused multimodal features. Unlike static fusion approaches, this enables adaptive prioritization of diagnostically relevant information without the complexity of full transformer architectures (Vaswani et al., 2017).

3. Attention-Augmented Neuro-Fuzzy Architecture (DMF-ANFIS)

A hybrid framework combining self-attention–based feature weighting with Adaptive Neuro-Fuzzy Inference System (ANFIS) is proposed, enhancing uncertainty modeling while preserving interpretability through fuzzy rule structures.

4. Interpretable Multimodal AI for Dermatology

The proposed framework provides transparent decision-making through attention weight visualization and fuzzy rule inference, addressing the limitations of black-box deep learning models in clinical applications (Hasan et al., 2022).

5. Statistically Validated Performance Improvement

The DMF-ANFIS framework achieves a mean multiclass classification accuracy of 94.7% ± 0.9 and macro-average sensitivity of 0.94, outperforming conventional machine learning models and baseline ANFIS with statistically significant improvements.

6. Context-Aware Diagnostic Modeling

The integration of environmental and contextual variables enables context-aware classification, addressing a critical limitation in existing dermatological AI systems that rely primarily on image data (Zhang et al., 2024).

7. Applicability to Resource-Constrained Healthcare Settings

The framework is validated using real-world clinical data from Nigeria, contributing to underrepresented sub-Saharan datasets and demonstrating suitability for deployment in resource-limited environments.

6. CONCLUSION

The DMF-ANFIS framework represents a significant advancement in health informatics, offering a scalable, interpretable solution for skin disease classification. By integrating attention mechanisms with neuro-fuzzy inference, it addresses multimodal data challenges, paving the way for enhanced diagnostic tools in dermatology.

ACKNOWLEDGMENTS

This work is derived from an ongoing PhD study in the Department of Computer Science at Ajayi Crowther University, focusing on the development of a dynamic multimodal feature fusion framework for AI-driven skin disease classification.

REFERENCES

1. Abbas, S., Ahmed, F., Khan, W. A., Ahmad, M., Khan, M. A., *et al.* (2025). *Intelligent skin disease prediction system using transfer learning and explainable artificial intelligence*. *Scientific Reports, 15*, 1746. https://doi.org/10.1038/s41598-024-83966-4
2. Aboulmira, A., Hrimech, H., Lachgar, M., Malyana, H., & Lemouden, K. (2025). *Attention-guided multimodal skin disease classification using discounted Dempster–Shafer fusion and EfficientNet*. *Informatics in Medicine Unlocked*. https://doi.org/10.1016/j.imu.2025.101699
3. Ahammed, K. T., Hossain, M. S., Rahman, M. M., & Islam, M. Z. (2022). *An effective machine learning approach for skin disease classification using image segmentation and feature extraction*. *Journal of Healthcare Engineering*, 2022, Article 8782524. https://doi.org/10.1155/2022/8782524
4. Debelee, T. G., Schwenker, F., Ibenthal, A., & Yohannes, D. (2023). *A review of deep learning techniques for skin lesion classification*. *Diagnostics, 13*(19), 3147. https://doi.org/10.3390/diagnostics13193147

5. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). *Dermatologist-level classification of skin cancer with deep neural networks*. *Nature, 542*(7639), 115–118. https://doi.org/10.1038/nature21056
6. Furqan, M., Katuk, N., & Hartama, D. (2026). *Multiclass skin lesion classification using attention-based vision transformer with metadata fusion*. *Journal of Applied Data Sciences, 7*(1), 203–217.
7. Hasan, M. K., Elahi, M. T. E., Alam, M. A., Jawad, M. T., & Martí, R. (2022). *DermoExpert: Skin lesion classification using a hybrid convolutional neural network through segmentation, transfer learning, and augmentation*. *Informatics in Medicine Unlocked, 28*, 100819. https://doi.org/10.1016/j.imu.2021.100819
8. Ipeayeda, F. W., Oyediran, M. O., Ajagbe, S. A., Jooda, J. O., & Adigun, M. O. (2023). *Optimized gravitational search algorithm for feature fusion in a multimodal biometric system*. *Results in Engineering, 20*, 101572. https://doi.org/10.1016/j.rineng.2023.101572
9. Jang, J.-S. R. (1993). *ANFIS: Adaptive-network-based fuzzy inference system*. *IEEE Transactions on Systems, Man, and Cybernetics, 23*(3), 665–685. https://doi.org/10.1109/21.256541
10. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). *Attention is all you need*. In *Advances in Neural Information Processing Systems* (Vol. 30, pp. 5998–6008).
11. Rasheed, J., Hameed, A. A., Djeddi, C., Jamil, A., & Al-Turjman, F. (2023). Artificial intelligence approaches in skin lesion analysis and classification: A survey. *Artificial Intelligence Review, 56*, 12289–12339. https://doi.org/10.1007/s10462-023-10425-4
12. Zadeh,L.A.(1965). *Fuzzy sets*. *Information and Control, 8*(3), 338–353. https://doi.org/10.1016/S0019-9958(65)90241-X
13. Zhang, X., Wang, S., Liu, J., & Tao, C. (2024). *Towards improving diagnosis of skin diseases by combining deep neural networks and human knowledge*. *BMC Medical Informatics and Decision Making, 24*, Article 1. https://doi.org/10.1186/s12911-024-02400-0