

TITLE:

Development and External Validation of a Fuzzy Machine Learning Framework for Dermatological Disease Classification Using Clinical Data

AUTHORS AND AFFILIATIONS:

Ubaida Fatima^{1*}

Department of Mathematics, NED University of Engineering & Technology, Karachi, Pakistan.

***CORRESPONDING AUTHOR:**

Ubaida Fatima (ubaida@neduet.edu.pk)

KEYWORDS:

Fuzzy machine learning; Dermatological disease classification; Fuzzy logic; Multi-class classification; Clinical decision support; External validation.

SUMMARY:

This protocol presents a reproducible fuzzy machine learning workflow for multi-class dermatological disease classification using symptom-based clinical data. The protocol includes data preprocessing, fuzzification, fuzzy predictor construction, rule-based inference, machine learning classification, and external validation to support interpretable and reproducible clinical decision-making.

ABSTRACT:

Dermatological diseases frequently present with overlapping clinical manifestations, making accurate diagnosis challenging due to uncertainty and variability in symptom interpretation. This study aimed to develop and externally validate a fuzzy machine learning framework for multi-class dermatological disease classification using symptom-based clinical and histopathological data. The proposed protocol integrates fuzzy logic with machine learning via a reproducible workflow that includes data acquisition, preprocessing, exploratory data analysis, four-valued fuzzification, fuzzy predictor construction, rule-based inference, model training, and external validation. The dermatology dataset, containing 366 patient records and six disease classes, was transformed from 34 original features into eight interpretable fuzzy predictors ($P_1 - P_8$). Random Forest (RF), Support Vector Machine (SVM), and Extreme Gradient Boosting (XGBoost) classifiers were evaluated using accuracy, precision, recall, F1-score, Cohen's kappa coefficient, 10-fold cross-validation accuracy, and training time. The proposed fuzzy framework achieved a maximum classification accuracy of 97.30%, an F1-score of 97.24%, and a Cohen's kappa coefficient of 0.966 using the RF classifier, while SVM demonstrated the shortest training time (0.038 s). External validation was performed using an independent dataset of 200 dermatology patients collected from Fatima Medical Centre, where RF and SVM each achieved 99.00% classification accuracy using the original clinical features and maintained strong predictive performance after fuzzy transformation. The proposed framework effectively reduced feature dimensionality while preserving clinically relevant information and demonstrated robust performance on independent clinical data, supporting its applicability for

interpretable dermatological disease classification and clinical decision-support systems.

INTRODUCTION:

Dermatological diseases are among the most prevalent chronic health conditions worldwide and impose a substantial clinical and socioeconomic burden. Diseases such as psoriasis, eczema, seborrheic dermatitis, lichen planus, pityriasis rosea, chronic dermatitis, and pityriasis rubra pilaris frequently exhibit overlapping clinical manifestations, including erythema, scaling, itching, papules, and inflammatory skin lesions, making accurate diagnosis challenging^{1,2}. Diagnosis typically relies on clinical examination and histopathological assessment, which may involve uncertainty, subjectivity, and inter-observer variability. Conventional machine learning (ML) approaches generally utilize crisp input values and often fail to adequately represent the uncertainty inherent in symptom-based clinical data³. Consequently, there is an increasing need for computational frameworks that can manage uncertainty while maintaining interpretability and clinical relevance.

Recent advances in artificial intelligence have significantly improved automated diagnosis of dermatological diseases through feature selection, deep learning, and ensemble learning techniques. Javed et al. demonstrated the importance of feature selection in improving skin lesion detection, whereas Khan et al. proposed a unified deep learning framework that integrates feature fusion and selection for skin lesion segmentation and classification^{4,5}. Karimi et al. comprehensively reviewed feature selection strategies for large-scale medical databases, highlighting the importance of dimensionality reduction for robust predictive modeling⁶. Furthermore, Aliyu et al. reported enhanced feature selection using a Chaotic Salp Swarm Algorithm for imbalanced medical datasets, while Nasir et al. presented a CNN-enabled framework for skin cancer detection^{7,8}. Although these studies have achieved promising results, they primarily focus on image-based diagnosis or conventional feature selection techniques and place limited emphasis on interpretable fuzzy machine learning models that use symptom-based clinical and histopathological data.

Previous fuzzy approaches have demonstrated the potential of fuzzy logic for dermatological diagnosis and broader healthcare decision support. Laskaris et al. (2010) developed a fuzzy perception-based framework for describing skin lesion images by assigning semantic attributes such as color, symmetry, border irregularity, and texture using machine learning trained on human assessments. Their work focused on semantic image description rather than interpretable fuzzy clinical decision support or disease classification⁹. Similarly, Ayo et al. (2020) proposed a fuzzy rule-based expert system for diagnosing severe acne skin disease, where fuzzy rules were employed to manage uncertainty in symptom interpretation and estimate disease severity¹⁰. However, their framework was designed specifically for acne diagnosis and did not incorporate feature reduction, integration of supervised machine learning, or external clinical validation. Beyond dermatology, Fatima and Khushal (2025) introduced an explainable fuzzy machine learning framework integrating fuzzy logic with machine learning and deep learning algorithms for epidemiological disease prediction under uncertain clinical conditions, while Khushal and Fatima (2024) developed a fuzzy logic-based machine learning framework for diabetes risk prediction using modifiable clinical risk factors and further reviewed the

expanding role of fuzzy computing in healthcare decision support, disease diagnosis, and intelligent clinical systems^{11–13}. In contrast, the proposed framework introduces a reproducible four-valued fuzzy representation that transforms clinically related variables into interpretable fuzzy predictors, integrates fuzzy reasoning with multiple supervised machine learning classifiers, and validates the methodology using an independent real-world clinical dataset, thereby providing a generalized and clinically interpretable framework for symptom-based disease classification.

Previous fuzzy dermatology studies have primarily employed conventional fuzzy inference systems, manually defined linguistic rules, neuro-fuzzy classifiers, or fuzzy preprocessing followed by standard machine-learning models. These approaches generally operate directly on the original clinical variables or use membership functions only as a preprocessing mechanism, without constructing a compact intermediate representation that preserves clinically related feature groups. In contrast, the proposed framework introduces a structured fuzzy predictor architecture in which 34 clinical and histopathological variables are first transformed using a reproducible four-valued membership scheme and then aggregated into eight clinically interpretable predictors. The methodological contribution, therefore, extends beyond dimensionality reduction. It provides a two-stage representation-learning mechanism that combines clinically guided feature grouping with fuzzy transformation, rule-based reasoning, and supervised classification. In addition, the same predictor-construction procedure is adapted to an independent clinical dataset, allowing the framework itself, rather than only the final classifiers, to be externally evaluated.

Existing fuzzy dermatological classification studies generally employ conventional fuzzy inference systems, fuzzy rule-based classifiers, or fuzzy preprocessing techniques without simultaneously addressing dimensionality reduction, model interpretability, reproducibility, and independent clinical validation. In contrast, the proposed framework introduces an interpretable eight-predictor ($P_1 - P_8$) representation that transforms 34 clinical and histopathological variables into eight clinically meaningful fuzzy predictors while preserving diagnostically relevant information. Furthermore, the framework integrates a reproducible four-valued fuzzy transformation, an interpretable fuzzy rule base, supervised machine-learning classification, and external validation with an independent clinical dataset. These methodological characteristics distinguish the proposed framework from previous fuzzy dermatology classification studies.

Unlike conventional fuzzy classifiers that directly assign disease classes through a single fuzzy inference stage, the proposed method uses fuzzy reasoning to construct a reduced, clinically meaningful predictor space, which is then evaluated by multiple supervised machine-learning algorithms. This separation between fuzzy representation construction and machine-learning classification enables assessment of the fuzzy transformation's effect independently across different classifiers and datasets.

Fuzzy logic offers an effective mechanism for representing uncertainty by transforming crisp clinical observations into linguistic categories such as Absent, Mild, Moderate, and Severe.

When integrated with machine learning algorithms, fuzzy systems improve interpretability while preserving clinically meaningful information during prediction. Despite these advantages, relatively few studies have investigated fuzzy machine learning frameworks for multi-class dermatological disease classification, particularly those incorporating external clinical validation and constructing interpretable fuzzy predictors. We hypothesize that an interpretable fuzzy predictor framework can reduce data dimensionality while preserving clinically relevant information and maintaining high classification performance.

To address this gap, this study proposes to develop and externally validate a fuzzy machine learning framework for multi-class dermatological disease classification using symptom-based clinical and histopathological data. The framework was initially developed using a publicly available dermatology dataset containing six disease classes. A preliminary fuzzy representation based on four predictors ($P_1 - P_4$) was first investigated; however, experimental evaluation indicated information loss and reduced classification performance. Therefore, the framework was extended to eight fuzzy predictors ($P_1 - P_8$), providing a more comprehensive representation of clinically and histopathologically related features while substantially reducing dimensionality. Finally, the proposed methodology was externally validated using an independent clinical dataset collected from Fatima Medical Centre, Karachi, to assess its robustness and generalizability.

The principal novelty of this study lies in developing and externally validating an interpretable fuzzy machine learning framework that combines (i) a reproducible four-valued fuzzy transformation for ordinal dermatological symptom severity, (ii) the construction of eight clinically interpretable fuzzy predictors ($P_1 - P_8$) that reduce dimensionality from 34 original variables to eight predictors, (iii) an interpretable fuzzy rule-based inference mechanism integrated with supervised machine learning classifiers, and (iv) validation using an independent real-world clinical dataset. Together, these contributions provide a reproducible and clinically interpretable framework for symptom-based classification of dermatological diseases.

Beyond dermatological disease classification, the proposed framework introduces a generalized multi-valued fuzzy reasoning paradigm for clinical decision support. Rather than evaluating individual symptoms independently, the framework simultaneously models physiologically related groups of clinical variables to capture their combined contribution to disease severity and progression. The proposed predictor construction allows multiple interacting clinical features (*e.g.*, x_1, x_2, x_3 , and x_4) to be aggregated into clinically meaningful latent predictors, enabling the identification of disease patterns that cannot be represented by isolated variables alone. Likewise, the absence or reduced severity of specific combinations of clinical factors may indicate disease remission, lower risk, or improved treatment response. This generalized multi-valued fuzzy representation is independent of dermatology and can therefore be extended to other symptom-based medical domains where multiple physiological indicators jointly influence diagnosis, prognosis, and therapeutic decision-making.

Consequently, the proposed methodology is not limited to classifying dermatological diseases.

It provides a clinically interpretable framework that can support physician decision-making, personalized risk assessment, disease monitoring, and treatment planning across a wide range of healthcare applications. Furthermore, by identifying clinically relevant combinations of physiological features rather than isolated biomarkers, the framework may assist pharmaceutical research by enabling understanding of disease mechanisms, patient stratification, and the design of more targeted therapeutic strategies.

Therefore, the principal novelty of this work lies not only in introducing a four-valued fuzzy machine learning framework, but also in establishing a generalized multi-valued clinical reasoning methodology that is transferable to other healthcare decision-support systems^{12–16}.

PROTOCOL:

The external validation dataset consisted of retrospectively collected, fully anonymized clinical records from Fatima Medical Centre in Karachi, Pakistan. All patient identifiers were removed before data analysis to ensure patient confidentiality, and no personally identifiable information was accessed or retained. The use of the anonymized clinical records for research purposes was authorized by the clinic administration. The recruitment period covered two months, during which consecutive patient records were collected.

Method 1: Data acquisition and preprocessing for dermatology patient data

Dermatology considered the dataset description

The primary dataset collected from Fatima Medical Centre and the publicly available dermatology dataset obtained from the UCI Machine Learning Repository¹⁴ (accessed via online dataset-hosting and data-science platform) were loaded into a computational environment such as Python (using pandas) or MATLAB. Functions such as `read_csv()` in Python or `readtable()` in MATLAB were used to import both datasets into memory.

The dermatology dataset comprised clinical and histopathological features of multiple dermatological diseases, including psoriasis, seborrheic dermatitis, lichen planus, pityriasis rosea, chronic dermatitis, and pityriasis rubra pilaris. The dataset contained symptom-based clinical variables measured on an ordinal severity scale ranging from 0 to 3, where 0 indicated the absence of symptoms, and 3 indicated severe symptoms. Key features included erythema, scaling, itching, acanthosis, hyperkeratosis, spongiosis, and elongation of rete ridges, together with patient age and disease class labels.

The structured ordinal representation of symptom severity made the dataset suitable for fuzzification and fuzzy rule-based modeling. Clinical and histopathological variables were used as input features for multiclass dermatological disease classification within the proposed fuzzy machine-learning framework. The dermatology dataset consisted of 366 patient records, 34 clinical and histopathological attributes, and a disease class label representing six dermatological conditions. The proposed fuzzy framework reduced the dataset's dimensionality from 34 original features to 8 fuzzy predictors ($P_1 - P_8$) while preserving clinically relevant information for disease classification.

The preprocessing workflow consisted of several sequential steps. First, the dermatology dataset was inspected, and a data dictionary describing all clinical and histopathological variables was prepared. Records containing missing values were then identified. The dermatology dataset contained eight records with missing age values. These records were excluded prior to analysis, resulting in a final dataset of 358 complete observations. Categorical disease labels were encoded as numerical class labels for machine-learning implementation, whereas ordinal clinical severity variables retained their original four-level coding (0–3). The dataset was divided into training (80%) and testing (20%) subsets through stratified random sampling with a random seed of 42. Normalization parameters were estimated from the training dataset and subsequently applied to the testing dataset to prevent data leakage. Finally, the normalized ordinal variables were transformed into four-valued fuzzy membership grades of 0.00, 0.33, 0.66, and 1.00. This transformation was followed by fuzzy predictor construction ($P_1 - P_8$), machine-learning classification, and performance evaluation. Numerical variables were normalized where necessary using min–max scaling. Ordinal symptom-severity values ranging from 0 to 3 were retained to preserve their clinical meaning during fuzzification into multilevel logic.

Exploratory data analysis and visualization

Exploratory data analysis was performed to examine class distribution, feature correlations, and feature severity. Descriptive statistics, correlation heatmaps, and feature-severity plots were generated to verify data quality and guide fuzzy predictor construction.

Method 2: Fuzzification of dermatology dataset clinical variables

Considered dermatology datasets

This study utilized two independent datasets. The primary dataset was the Dermatology Dataset obtained from the UCI Machine Learning Repository (<https://archive.ics.uci.edu/ml/datasets/Dermatology>)¹⁴. The dataset contained 366 patient records, 34 clinical and histopathological attributes, and six dermatological disease classes. For convenient access and data management, the same dataset was downloaded through its mirrored online dataset-hosting and data-science platform repository. The online dataset-hosting and data-science platform version contained identical records and variables and did not represent a different dataset. Consequently, all experiments reported for the primary dataset were conducted using the original UCI Dermatology Dataset.

To evaluate the generalizability of the proposed framework, an independent external clinical validation dataset comprising 200 retrospectively collected, anonymized patient records was obtained from Fatima Medical Centre, Karachi, Pakistan. This dataset included 140 eczema cases and 60 psoriasis cases and was used exclusively for external validation. The same preprocessing, fuzzification, predictor construction, fuzzy inference, and machine-learning procedures applied to the primary dataset were subsequently applied to the external clinical dataset.

Selection of key features

Clinically relevant variables were selected from the dermatology dataset, including erythema,

scaling, itching, acanthosis, hyperkeratosis, spongiosis, elongation of rete ridges, and additional histopathological indicators such as parakeratosis and clubbing of rete ridges. These features were selected because of their diagnostic significance in distinguishing among multiple dermatological disease classes.

Definition of membership functions

The proposed framework followed the general concept of fuzzy membership representation used in conventional triangular and trapezoidal fuzzy systems. However, because the dermatology dataset recorded symptom severity using four ordinal categories—Absent, Mild, Moderate, and Severe—continuous overlapping membership functions were deemed inappropriate. Instead, a heuristic discrete membership assignment was employed to preserve the ordinal clinical interpretation while providing a simple, reproducible, and interpretable fuzzification procedure. The predefined membership grades were assigned as Absent = 0.00, Mild = 0.33, Moderate = 0.66, and Severe = 1.00. These values were selected based on established fuzzy-logic principles reported in the literature and were further reviewed by Dr. Khalida Perveen at Fatima Medical Centre for clinical consistency with the ordinal symptom-severity definitions. Consequently, breakpoint estimation, overlap widths, and parameter optimization required for conventional triangular or trapezoidal membership functions were not applicable to the proposed framework.

$$\mu(x) = \begin{cases} 0.00, & x = 0 \text{ (Absent)} \\ 0.33, & x = 1 \text{ (Mild)} \\ 0.66, & x = 2 \text{ (Moderate)} \\ 1.00, & x = 3 \text{ (Severe)} \end{cases} \quad (1)$$

In Equation (1), $x = \{0,1,2,3\}$, the first variable denotes the original ordinal clinical severity score, and $\mu(x)$, the second variable, represents the corresponding fuzzy membership grade.

Four-valued fuzzy transformation

The original dermatological symptom-severity values, which ranged from 0 to 3, were transformed into fuzzy membership values ranging from 0 to 1. Four linguistic categories—Absent, Mild, Moderate, and Severe—were used to represent different degrees of symptom manifestation. The transformed fuzzy values were subsequently used to construct predictors and generate fuzzy rules. This design preserved the ordinal interpretation of the clinical observations while providing a simple, reproducible, and computationally efficient framework for constructing the fuzzy predictors ($P_1 - P_8$) (Table 1 and Table 2)

Implementation of fuzzification

Each ordinal clinical and histopathological variable was transformed into its corresponding fuzzy membership grade according to Equation (1). The resulting fuzzy values of 0.00, 0.33, 0.66, and 1.00 were subsequently used to construct the fuzzy predictors ($P_1 - P_8$), which served as inputs to the fuzzy rule-based inference system and the machine-learning classifiers.

Method 3: Development of a fuzzy rule-based inference system

Feature-based rule design

Fuzzy IF–THEN rules were developed using clinically relevant dermatological features such as erythema, scaling, itching, acanthosis, hyperkeratosis, and elongation of rete ridges. Rule construction was based on established dermatological diagnostic patterns associated with multiple disease classes in the dermatology dataset, including psoriasis, seborrheic dermatitis, lichen planus, pityriasis rosea, chronic dermatitis, and pityriasis rubra pilaris.

Rule formulation

The fuzzy IF–THEN rules were manually formulated based on established dermatological knowledge, standard medical references (Bologna et al.; Guyton and Hall; Tortora)^{17–19}, and the clinical interpretation of the grouped fuzzy predictors ($P_1 - P_8$). The proposed rule base was further reviewed using clinical input from Dr. Khalida Perveen at Fatima Medical Centre to ensure consistency with common dermatological diagnostic patterns. The rules provided a transparent and interpretable reasoning process and were not automatically learned or optimized from the dataset²⁰. Representative fuzzy rules were presented in **Table 3**.

The fuzzy predictor values were calculated as the average fuzzy membership values of their corresponding feature groups. These predictors served as condensed representations of clinical symptoms, epidermal changes, inflammatory characteristics, and histopathological findings. Predictor construction was shown in Equations (2)–(5).

$$P_1 = \frac{\text{Erythema} + \text{Scaling} + \text{Itching} + \text{Definite Borders}}{4} \quad (2)$$

$$P_2 = \frac{\text{Acanthosis} + \text{Hyperkeratosis} + \text{Parakeratosis} + \text{Clubbing Rete Ridges}}{4} \quad (3)$$

$$P_3 = \frac{\text{PNL Infiltrate} + \text{Eosinophils Infiltrate} + \text{Exocytosis} + \text{Spongiosis}}{4} \quad (4)$$

$$P_4 = \frac{\text{Elongation Rete Ridges} + \text{Thinning Suprapapillary Epidermis} + \text{Munro Microabscess} + \text{Spongiform Pustule}}{4} \quad (5)$$

The fuzzy predictor values were calculated as the arithmetic mean of the corresponding fuzzy membership grades, because all grouped features represented clinically related manifestations of the same dermatological characteristic and were measured on the same four-level ordinal scale. The arithmetic mean provided a simple and interpretable summary while assigning equal weight to each feature. In the absence of established clinical evidence or validated feature-importance measures indicating that some variables contributed more than others, equal weighting was adopted to avoid subjective bias. Consequently, no weight optimization was performed.

The eight fuzzy predictors ($P_1 - P_8$) were calculated as the arithmetic means of their corresponding feature groups presented in **Table 4**. The same procedure described in Equations (1)–(4) for the ($P_1 - P_4$) four-predictor framework was used.

For each predictor, the fuzzy value was calculated as the arithmetic mean of the four associated fuzzy-transformed features. The predictor definitions were provided in **Table 2** and **Table 4**, in

which each predictor represented the aggregated contribution of clinically related dermatological characteristics. Each feature was first transformed into a four-valued fuzzy membership value of 0.00, 0.33, 0.66, or 1.00, after which the corresponding predictor value was calculated using the mean of the grouped features.

Rule integration

The fuzzy IF–THEN rules presented in **Table 5** were integrated into a rule-based inference system using logical AND operators. The fuzzy IF–THEN rules were integrated into a rule-based inference system using logical AND operators. For each patient, the fuzzy predictor values ($P_1 - P_8$) were evaluated against the antecedent conditions of all predefined rules. Rules that satisfied the specified conditions were activated, and their corresponding disease-class membership values were aggregated using the maximum operator. The aggregated membership values were then passed to the decision stage, in which the disease class with the highest membership value was selected as the final prediction. The complete rule base and its implementation were provided in raw data.

Defuzzification

The maximum-membership principle was employed to convert the fuzzy inference outputs into crisp disease-class predictions. The disease class associated with the highest aggregated fuzzy membership value was selected as the final classification outcome.

Integration of the fuzzy inference system with machine learning

The proposed framework integrated fuzzy reasoning with supervised machine learning through a sequential computational pipeline. Initially, the raw clinical and histopathological variables were preprocessed and transformed into four-valued fuzzy membership grades. The fuzzified variables were subsequently aggregated into eight clinically interpretable fuzzy predictors ($P_1 - P_8$) using the corresponding predictor equations. These predictors were evaluated using the predefined fuzzy IF–THEN rule base, in which logical AND operators determined rule activation and the activated rule outputs were combined using the maximum aggregation operator. The maximum-membership principle was then employed to assign the final fuzzy state of each predictor. The resulting fuzzy predictor values constituted a reduced feature matrix that served as the input to the supervised machine-learning classifiers, including Random Forest (RF), Support Vector Machine (SVM), and Extreme Gradient Boosting (XGBoost). Finally, the trained classifiers generated disease predictions, and their performance was evaluated using accuracy, precision, recall, F1-score, Cohen’s kappa coefficient, cross-validation accuracy, and training time.

Complete fuzzy inference procedure

The proposed fuzzy inference system operated through a sequence of steps. First, the clinical and histopathological variables were transformed into four-valued fuzzy membership grades of 0.00, 0.33, 0.66, and 1.00. The fuzzy-transformed variables were then aggregated to construct the ($P_1 - P_8$) eight fuzzy predictors using the arithmetic mean of clinically related feature groups. The calculated predictor values were evaluated against the predefined fuzzy IF–THEN rules using the logical **AND** operator. Activated rules contributed membership values to their

corresponding dermatological disease classes, and multiple activated rules were combined using the maximum aggregation operator. Finally, the maximum-membership defuzzification principle assigned each patient to the disease class with the highest aggregated membership value, which was subsequently used as input for supervised machine-learning classification. This sequential fuzzy inference procedure provided a complete and reproducible implementation of the proposed fuzzy machine-learning framework^{9,10,11} for dermatological disease classification.

Method 4: Applied a fuzzy machine learning framework

Dataset splitting

To prevent data leakage, the dataset was first divided into training and testing subsets using stratified sampling. All preprocessing operations that required parameter estimation, including missing-value imputation and normalization, were fitted exclusively using the training dataset. The learned preprocessing parameters were subsequently applied unchanged to the testing dataset. The predefined four-valued fuzzy transformation was then applied consistently to both the training and testing samples using the same membership-assignment scheme. This procedure ensured that no information from the testing dataset influenced model training or preprocessing.

Machine learning (ML) models

Random Forest (RF)

Random Forest²¹ was an ensemble learning algorithm that constructed multiple decision trees and combined their predictions via majority voting. The method was robust to noise, reduced overfitting, and effectively processed high-dimensional clinical and histopathological data. In this study, Random Forest was employed to classify dermatological diseases using both the original and fuzzified feature representations.

Support vector machine (SVM)

Support Vector Machine²² was a supervised machine-learning algorithm that identified an optimal decision boundary for separating multiple disease classes. SVM was particularly effective for high-dimensional classification problems and provided strong generalization performance. The classifier was trained using both the original and fuzzy predictor datasets to evaluate the effect of fuzzy transformation on classification accuracy and computational efficiency.

Extreme gradient boosting (XGBoost)

XGBoost²³ was a gradient-boosting algorithm that sequentially constructed decision trees to minimize classification errors. The method incorporated regularization and optimization strategies to improve predictive performance and reduce overfitting. XGBoost was used to investigate the effectiveness of the proposed fuzzy predictor²⁴ framework in improving classification accuracy and reducing computational complexity.

The machine-learning implementation and parameter settings are available with the raw data.

Model hyperparameter configuration

The machine-learning models were implemented using fixed hyperparameter settings. Random Forest was configured with 100 decision trees ($n_estimators = 100$), the Gini impurity criterion, unlimited tree depth ($max_depth = None$), and $random_state = 42$. The Support Vector Machine used a radial basis function (RBF) kernel with $C = 10$ and $gamma = "scale."$ XGBoost was configured with 100 estimators ($n_estimators = 100$), $learning_rate = 0.1$, $max_depth = 4$, $objective = "multi:softmax,"$ $num_class = 6$, and $random_state = 42$. Default values were retained for all remaining hyperparameters.

ML performance evaluation metrics

Accuracy is measured by the proportion of correctly classified instances among all observations⁹, as shown in Equations (6)–(9), in which TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

1. **Precision:** It evaluates the proportion of positive predictions that are correct⁹. (See Equation (7))

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

2. **Recall:** It measures the ability of the classifier to correctly identify relevant disease instances⁹. (See Equation (8))

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

3. **F1-Score:** The F1-score⁹ provides a balanced measure of precision and recall. (See Equation (9))

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (9)$$

4. **Cohen's Kappa (k):** Cohen's kappa²⁵ evaluates agreement between predicted and actual classes while accounting for agreement that occurs by chance.

$$k = \frac{P_o - P_e}{1 - P_e} \quad (10)$$

Cross-validation accuracy

Tenfold cross-validation was employed to assess model robustness and generalization capability²⁶. The average classification accuracy across the 10 folds was reported.

Training time

Computational training time was recorded in seconds to compare the efficiency of the different

machine-learning models and feature representations. In Equation (10), P_0 and P_e represent the first and second terms, respectively, which represent the observed agreement between the predictions and true labels and the expected agreement resulting from random chance. Higher kappa values indicated stronger classification reliability and predictive consistency.

Model training

The selected model was trained using the fuzzified input features. Training was performed using commands such as `model.fit(X_train, y_train)`.

Hybrid integration

The fuzzy machine-learning predictions were integrated by combining the outputs of the fuzzy rule-based system with those of the machine-learning model. This integration improved classification accuracy and enhanced the interpretability of the dermatological diagnoses.

Method 5: Model evaluation and validation

Prediction generation

The trained model was applied to the testing dataset using functions such as `model.predict(X_test)`. The predicted disease classes were recorded.

Performance evaluation

Model performance was evaluated using metrics such as accuracy, precision, recall, and the confusion matrix. Functions such as `classification_report()` and `confusion_matrix()` in Python were used. Comparative performance results for the original and fuzzy feature representations are reported in **Table 6**.

Interpretation of results

The fuzzy outputs were analyzed alongside the classification results to assess uncertainty and confidence levels. The results were interpreted in the context of dermatological diagnosis.

Method 6: External Clinical Validation

Clinical Dataset Preparation

An independent external clinical validation dataset was retrospectively collected at Fatima Medical Centre in Karachi, Pakistan, to evaluate the generalizability of the proposed framework. The dataset consisted of 200 fully anonymized patient records collected over a 2-month period. Patient selection included individuals with dermatologist-confirmed diagnoses of eczema or psoriasis who had complete clinical records that included all required clinical variables. Records with incomplete clinical information were excluded from the analysis.

The final dataset comprised 140 eczema cases and 60 psoriasis cases. Among patients with eczema, 90 had contact dermatitis, and 50 had chronic dermatitis. The dataset included 103 male and 97 female patients. The age distribution included 37 patients younger than 20 years, 103 patients aged between 20 and 40 years, and 60 patients older than 40 years.

The recorded clinical variables included redness, itching, watering, secondary infection, boil

formation, pus discharge, dryness, dandruff, and clinician-assessed weak immunity. These variables represented clinically relevant manifestations commonly used during dermatological assessment to differentiate eczema from psoriasis.

To ensure methodological consistency, the same preprocessing, four-valued fuzzification, fuzzy predictor construction, rule-based inference, and machine-learning procedures applied to the primary dermatology dataset were subsequently applied to the external clinical dataset. Classification performance was evaluated using accuracy, precision, recall, F1-score, Cohen's kappa coefficient, tenfold cross-validation accuracy, and computational training time.

The collected clinical records were reviewed for completeness and consistency before analysis. The same preprocessing procedures applied to the primary dermatology dataset, including data cleaning, normalization, feature selection, and four-valued fuzzy transformation, were subsequently applied to the clinical validation dataset. This approach ensured methodological consistency and enabled an objective evaluation of the proposed fuzzy machine-learning framework using real-world clinical data.

The anonymized external validation dataset was made available as raw data.

Validation of fuzzy rules

The developed fuzzy rule base was evaluated using the clinical dataset. The predicted disease classifications were compared with the clinician-confirmed diagnoses. For the external clinical validation dataset, three fuzzy predictors ($P_1 - P_3$) were constructed from nine clinically relevant dermatological features. The predictors represented symptom severity, infection-related manifestations, and psoriasis-associated manifestations. These predictors were subsequently used for machine-learning classification and performance evaluation. The clinical variables assigned to each fuzzy predictor and the corresponding predictor-construction formulas are summarized in **Table 7**.

A simplified fuzzy rule base was developed for the external validation dataset using the three fuzzy predictors ($P_1 - P_3$). The rules were formulated according to the predominant clinical manifestations of eczema and psoriasis observed in the collected patient records. The eczema-related rules primarily incorporated symptom severity and infection-associated manifestations represented by P_1 and P_2 , whereas the psoriasis-related rules were predominantly associated with dryness, dandruff, and immune status indicators represented by P_3 . The resulting rule base was subsequently integrated with the machine-learning classifiers to evaluate the robustness of the proposed fuzzy machine-learning framework using independent clinical data. The representative fuzzy rules used to differentiate eczema and psoriasis are presented in **Table 8**.

Performance assessment

Accuracy, precision, recall, F1-score, Cohen's kappa coefficient, cross-validation accuracy, and training time were calculated.

Method 7: Visualization and reproducibility

Visualization of membership functions

Plots of the fuzzy membership functions were generated to illustrate the severity levels of the clinical features.

Workflow diagram generation

A schematic diagram representing the complete pipeline was created. The diagram included data preprocessing, fuzzification, rule-based inference, machine learning, and prediction.

Documentation of parameters

All preprocessing steps, membership-function definitions, and model parameters were documented. Scripts and configurations were saved to ensure reproducibility.

Preparation for video demonstration

The step-by-step execution of the workflow, including data preprocessing, fuzzification, rule implementation, and model training, was recorded. Clarity and logical sequencing were maintained throughout the visual demonstration.

Reproducibility resources

The computational workflow was implemented using a Python-based computational environment with standard data-processing, machine-learning, statistical-analysis, and visualization libraries. Complete software names, versions, sources, and URLs were provided in the **Table of Materials**.

The preprocessing procedures, fuzzy membership assignments, predictor-construction equations, fuzzy rule base, machine-learning hyperparameters, evaluation metrics, and statistical analyses were fully documented in the manuscript. The Python implementation, workflow scripts, and raw data were provided to facilitate reproduction of the proposed methodology.

Method 8: Statistical analysis

Classification performance was evaluated using accuracy, precision, recall, F1-score, Cohen's kappa coefficient, and tenfold cross-validation accuracy. Statistical significance between the paired classification outcomes from the original and fuzzy feature representations was assessed using McNemar's test, which compares paired predictions on the same testing samples. The continuity-corrected McNemar test statistic was calculated using Equation (11).

$$\chi^2 = \frac{(|b - c| - 1)^2}{b + c} \quad (11)$$

Where b denotes the number of samples correctly classified by the original model but misclassified by the fuzzy model, and c denotes the number of samples misclassified by the original model but correctly classified by the fuzzy model. Statistical significance was defined as $p < 0.05$. All statistical analyses were performed using a standard Python-based computational environment. Complete software information is provided in the **Table of Materials**.

REPRESENTATIVE RESULTS:

The dermatology dataset consisted of 366 patient records and 34 clinical and histopathological features representing six dermatological disease classes. Exploratory data analysis (**Figure 1**, **Figure 2**, and **Figure 3**) demonstrated the class distribution, correlations among clinical variables, and average severity scores for dermatological features, providing the basis for subsequent fuzzy transformation and machine learning analysis. The complete computational workflow was summarized in **Figure 4**, which illustrated the sequential stages of data acquisition, preprocessing, four-valued fuzzification, fuzzy predictor construction, fuzzy rule-based inference, machine-learning classification, and external validation.

Detailed class-wise precision, recall, F1-score, and support for the Random Forest, Support Vector Machine, and XGBoost classifiers are provided in raw data, while the corresponding confusion matrices are presented in raw data.

On the fixed stratified test set of 74 patients, RF achieved an accuracy of 95.95% (95% CI: 88.75%–98.61%) using the original 34 features and 97.30% (95% CI: 90.67%–99.26%) using the $P_1 - P_8$ predictors. The original and $P_1 - P_8$ models produced discordant outcomes for three patients: one patient was correctly classified only by the original model, and two were correctly classified only by the $P_1 - P_8$ model. The 1.35-percentage-point difference was not statistically significant (exact McNemar $p = 1.000$; paired-bootstrap 95% CI for the difference: -2.70 to 6.76 percentage points). Therefore, the holdout analysis indicates that the proposed $P_1 - P_8$ framework maintained classification performance while reducing the number of predictors from 34 to 8, with no statistically significant difference in accuracy compared with the original feature representation.

In repeated stratified 10-fold cross-validation, RF achieved $97.70\% \pm 0.29\%$ with the original features and $97.24\% \pm 0.20\%$ with $P_1 - P_8$. The small difference favored the original representation (Wilcoxon $W = 2.0$, Holm-adjusted $p = 0.0117$). These findings indicate that $P_1 - P_8$ provides substantial dimensionality reduction with a small average performance tradeoff.

The proposed framework was further evaluated using an independent external clinical validation dataset collected from Fatima Medical Centre. The external validation performance of Random Forest (RF), Support Vector Machine (SVM), and XGBoost classifiers using the original clinical features and the proposed fuzzy predictor representation is summarized in **Table 9**. The results demonstrate that the proposed fuzzy framework maintains high classification performance while reducing feature dimensionality on independent clinical data.

The robustness and statistical validity of the proposed fuzzy machine learning framework were further evaluated using complementary analyses. McNemar's statistical comparison (**Table 10**) showed no statistically significant differences between the original feature representation and the proposed fuzzy predictor framework, indicating that dimensionality reduction was achieved

without a significant loss in classification performance. Stratified 10-fold cross-validation (**Table 11**) further confirmed the stability and generalizability of the proposed framework across repeated data partitions. Finally, a qualitative comparison with recent dermatological disease classification studies (**Table 12**) highlights the distinctive characteristics of the proposed framework, including interpretable four-valued fuzzy reasoning, clinically meaningful feature reduction, supervised machine learning integration, and independent external clinical validation.

As summarized in **Table 12**, recent dermatological disease classification studies have primarily focused on image-based deep learning models, conventional fuzzy expert systems, or disease-specific prediction frameworks. Although these approaches have reported promising predictive performance, differences in datasets, disease categories, input modalities, and evaluation protocols preclude direct quantitative comparison with the present study. Therefore, a structured qualitative comparison was performed. The proposed framework distinguishes itself by combining interpretable four-valued fuzzy reasoning, clinically meaningful feature reduction from 34 variables to eight fuzzy predictors, supervised machine learning classification, and independent external clinical validation within a unified and reproducible computational workflow. These characteristics enhance both interpretability and practical applicability for symptom-based clinical decision support.

To further evaluate the proposed fuzzy feature representation, **four conventional baseline classifiers** were investigated using identical preprocessing and train-test settings. As shown in **Table 13**, the fuzzy predictors improved the performance of **Decision Tree** and **k-Nearest Neighbors**, with the largest gain observed for **k-Nearest Neighbors** (87.84% to 95.95%). Logistic Regression and **Naïve Bayes** exhibited slightly lower performance with the fuzzy representation, indicating that the proposed feature transformation is particularly beneficial for non-linear and instance-based learning algorithms rather than linear or probabilistic classifiers.

McNemar's statistical analysis (**Table 14**) further evaluated whether the observed performance differences between the original and fuzzy feature representations were statistically significant. Although the proposed fuzzy representation improved the predictive performance of Decision Tree and k-Nearest Neighbors, while slightly reducing the performance of Logistic Regression and Naïve Bayes, none of these differences reached statistical significance (all $p > 0.05$). These findings suggest that the fuzzy feature representation maintains classification performance while improving selected classifiers without introducing statistically significant changes in predictive behavior. The largest practical improvement was observed for k-Nearest Neighbors, indicating that distance-based learning algorithms may benefit more from the proposed fuzzy representation than linear or probabilistic classifiers.

DATA AVAILABILITY:

The primary dermatology dataset, the fully anonymized external clinical validation dataset from Fatima Medical Centre, Karachi, Pakistan, raw datasets (where permitted), processed datasets, prediction outputs, Python source code, computational workflow, and supplementary materials supporting this study have been deposited in the Zenodo repository and are publicly available

at <https://doi.org/10.5281/zenodo.21384359>. The corresponding GitHub repository is available at <https://github.com/Ubaidafatima/Fuzzy-Dermatology-Classification>.

FIGURE AND TABLE LEGENDS:

Figure 1: Distribution of dermatological disease classes in the UCI Dermatology dataset. The dermatology dataset contains six disease categories: Psoriasis (112 cases), Seborrheic Dermatitis (61 cases), Lichen Planus (72 cases), Pityriasis Rosea (49 cases), Chronic Dermatitis (52 cases), and Pityriasis rubra Pilaris (20 cases). The figure illustrates the class distribution and shows moderate class imbalance, with psoriasis as the largest class and pityriasis rubra pilaris as the smallest.

Figure 2: Correlation heatmap of dermatological clinical and histopathological features. Pearson correlation coefficients were computed among the dermatological variables and visualized using a heatmap. Warmer colors indicate positive correlations, whereas cooler colors indicate negative correlations. The heatmap helps identify relationships among clinical and histopathological features used in fuzzy predictor construction and machine-learning classification.

Figure 3: Average severity scores of dermatological features. The figure presents the mean severity score of each clinical and histopathological feature across all patient records. Erythema, acanthosis, inflammatory mononuclear infiltrate, and scaling exhibit the highest average severity values, indicating their prominence in the dermatology dataset.

Figure 4: Proposed fuzzy machine learning workflow for dermatological disease classification. The workflow begins with the acquisition and preprocessing of the dermatology dataset, followed by four-valued fuzzification of clinical and histopathological features. Fuzzy predictors ($P_1 - P_8$) are subsequently constructed and utilized to formulate an interpretable fuzzy rule base. The fuzzy predictors are then used as inputs to Random Forest, Support Vector Machine, and XGBoost classifiers. Model performance is evaluated using multiple classification metrics, and the framework is externally validated on anonymized psoriasis and eczema patient records collected from Fatima Medical Centre in Karachi.

Table 1: Discrete four-level fuzzy membership assignment used to transform ordinal clinical severity scores into fuzzy membership grades. Original symptom severity levels ranging from 0 to 3 were mapped to four fuzzy membership grades (Absent, Mild, Moderate, and Severe), with corresponding membership values between 0 and 1, to facilitate fuzzy reasoning and uncertainty modeling.

Table 2: Preliminary construction of fuzzy predictors $P_1 - P_4$. Clinically and histopathologically related dermatological features were initially grouped into four fuzzy predictors to reduce dimensionality. Experimental evaluation revealed information loss and reduced classification performance, motivating the development of the extended $P_1 - P_8$ framework.

Table 3: Preliminary fuzzy rule base for dermatological disease classification using the $P_1 - P_4$ framework. Representative IF-THEN rules were developed from clinical symptoms and histopathological characteristics to support interpretable disease classification and fuzzy inference.

Table 4: Construction of the proposed fuzzy predictors ($P_1 - P_8$) from dermatological clinical and histopathological features. The original dermatology dataset features were grouped into eight clinically interpretable fuzzy predictors. Each predictor was computed as the average fuzzy membership value of four related dermatological features and used as input for machine learning classification models.

Table 5: Fuzzy rule base for dermatological disease classification using the proposed $P_1 - P_8$ fuzzy predictor framework. Representative IF-THEN rules were formulated using combinations of fuzzy predictors to model disease-specific clinical patterns associated with psoriasis, seborrheic dermatitis, lichen planus, pityriasis rosea, chronic dermatitis, and pityriasis rubra pilaris.

Table 6: Construction of fuzzy predictors ($P_1 - P_3$) for the external clinical validation dataset. The clinical variables collected from Fatima Medical Centre, Karachi, were grouped into three fuzzy predictors representing common dermatological symptoms, infection and inflammatory manifestations, and psoriasis-associated manifestations. Predictor P_1 was constructed using redness, itching, and watering; predictor P_2 was constructed using secondary infection, boils, and pus discharge, and the predictor P_3 was constructed using dryness, dandruff, and weak immunity. Each predictor was computed as the arithmetic mean of the corresponding fuzzy-transformed clinical features and subsequently used for fuzzy rule generation and machine-learning-based external validation.

Table 7: Fuzzy rule base for eczema and psoriasis classification using the $P_1 - P_3$ validation framework. Representative IF-THEN rules were developed based on the clinical manifestations observed in the external validation dataset collected from Fatima Medical Centre. The rules incorporated combinations of symptom severity, infection-related manifestations, and psoriasis-associated manifestations to facilitate interpretable classification of eczema and psoriasis cases within the fuzzy machine learning framework.

Table 8: Comparative performance evaluation of Random Forest, Support Vector Machine, and XGBoost classifiers using the original dermatology dataset and the proposed fuzzy predictor representations. The $P_1 - P_8$ framework reduced dimensionality from 34 original features to eight predictors while retaining broadly comparable holdout performance. Statistical comparisons and 95% confidence intervals are reported to quantify uncertainty in the observed performance differences. Training times represent descriptive model-fitting measurements obtained under the specified experimental environment.

Table 9: External validation performance of Random Forest, Support Vector Machine, and XGBoost classifiers using the Fatima Medical Centre dataset. Classification performance was

evaluated using accuracy, precision, recall, F1-score, Cohen's kappa coefficient, cross-validation accuracy, and computational training time. Original dataset refers to the nine clinical features collected from dermatology patients, whereas the fuzzy representation corresponds to the three fuzzy predictors ($P_1 - P_3$) constructed using the proposed fuzzy transformation framework. The results demonstrate the ability of the fuzzy predictor framework to reduce dimensionality while maintaining strong predictive performance on independent clinical data.

Table 10: McNemar's statistical comparison between the original 34-feature dermatology dataset and the proposed eight-predictor fuzzy ($P_1 - P_8$) representation using Random Forest, Support Vector Machine, and XGBoost classifiers. McNemar's test was performed on paired predictions from the same test set to determine whether differences in classification accuracy between the original and fuzzy feature representations were statistically significant ($p < 0.05$).

Table 11: Stratified 10-fold cross-validation performance of the proposed framework on the external clinical validation dataset. Mean cross-validation (CV) accuracy, standard deviation (SD), and 95% confidence interval (CI) are reported for Random Forest (RF), Support Vector Machine (SVM), and Extreme Gradient Boosting (XGBoost). Stratified 10-fold cross-validation preserved the original class distribution (140 eczema and 60 psoriasis cases) in each fold to assess model stability and robustness.

Table 12: Qualitative comparison of the proposed fuzzy machine learning framework with recent dermatological disease classification studies. Most published studies employ dermoscopic image datasets and deep learning methods, whereas the proposed framework utilizes symptom-based clinical and histopathological variables, interpretable four-valued fuzzy reasoning, dimensionality reduction, supervised machine learning, and independent external clinical validation. Because of differences in datasets, disease classes, input modalities, and evaluation protocols, direct quantitative comparison is not appropriate. Instead, the table summarizes the principal methodological characteristics of representative recent studies. Detailed confusion matrices together with class-wise precision, recall, and F1-score values for the Random Forest, Support Vector Machine, and XGBoost classifiers are provided in the raw data.

Table 13: Performance comparison of conventional baseline classifiers using original and fuzzy feature representations. Comparison of Logistic Regression, Decision Tree, Naïve Bayes, and k-Nearest Neighbors trained using the original 34 clinical features and the proposed eight fuzzy predictors. Classification performance is reported in terms of accuracy, highlighting the impact of the fuzzy feature representation on conventional machine learning algorithms.

Table 14: McNemar's statistical comparison between original and fuzzy feature representations using conventional baseline classifiers. McNemar's statistical test comparing the classification performance of conventional baseline classifiers trained on the original clinical features and the proposed fuzzy predictors. The table reports McNemar's χ^2 statistic, p-value, and statistical interpretation to assess whether the observed performance differences are

statistically significant.

DISCUSSION:

Successful implementation of the proposed framework depends on accurate preprocessing, consistent four-valued fuzzification, and correct construction of the fuzzy predictors ($P_1 - P_8$). Care should be taken to preserve the original ordinal clinical severity scores during preprocessing, as incorrect encoding or inconsistent fuzzy membership assignment may affect predictor construction and rule activation. Verification of the fuzzy predictor values and rule outputs before classifier training improves reproducibility and reduces implementation errors^{11,12,24}.

Several measures were adopted to minimize potential bias throughout the computational workflow. Stratified data partitioning preserved disease-class distributions, while normalization parameters were estimated exclusively on the training dataset and then applied to the testing dataset, thereby preventing data leakage. The external validation dataset was processed independently using the identical preprocessing and fuzzy transformation procedures to provide an unbiased assessment of model performance²¹⁻²⁸.

Initially, a preliminary fuzzy representation comprising of four predictors ($P_1 - P_4$) was developed by grouping clinically and histopathologically related features. Although this approach substantially reduced the dataset's dimensionality, the experimental results showed a noticeable decline in classification performance, suggesting that important disease-specific information was not fully preserved. To address this limitation and minimize information loss, the fuzzy representation was extended to include eight predictors (P_1-P_8), incorporating nearly all the clinical and histopathological attributes available in the dermatology dataset. The expanded eight-predictor framework ($P_1 - P_8$) provided a more comprehensive representation of disease characteristics while still achieving substantial dimensionality reduction, from 34 original features to 8 fuzzy predictors. Comparative experiments demonstrated that the eight-predictor ($P_1 - P_8$) framework substantially improved classification accuracy and preserved clinically relevant information more effectively than the preliminary four-predictor ($P_1 - P_4$) representation.

The reproducibility of the proposed methodology is supported by publicly available datasets, complete source code, workflow documentation, parameter settings, and supplementary materials deposited in the associated GitHub and Zenodo repositories. Nevertheless, classifier performance may vary slightly across different computational environments, software versions, or hardware configurations. Although the framework demonstrated robust performance on an independent clinical dataset, additional validation on larger multicenter datasets that represent broader demographic and disease diversity would further strengthen the model's generalizability.

Beyond dermatological disease classification, the proposed fuzzy machine learning protocol may be adapted to other clinical decision-support applications involving ordinal symptom

severity, uncertainty, and interpretable prediction. Future studies may investigate weighted fuzzy predictor construction, optimized fuzzy membership functions, automated rule learning, explainable artificial intelligence techniques, and integration with deep learning and multimodal clinical data to further improve predictive performance and clinical applicability^{11–13}.

A limitation of this study is that the external validation dataset consisted solely of clinically confirmed cases of eczema and psoriasis. Therefore, it does not represent a complete external validation of the proposed six-class dermatological classification framework. However, the external evaluation demonstrates that the proposed eight-predictor fuzzy feature-reduction methodology can be successfully applied to an independent clinical dataset while preserving diagnostic performance. Future studies will extend the external validation to larger multicenter datasets encompassing all six dermatological disease classes to further evaluate the generalizability of the proposed framework. The high classification performance observed on the external validation dataset should be interpreted in the context of its binary classification design (eczema versus psoriasis), which represents a less complex prediction task than the six-class primary dermatology dataset. Although external validation supports the robustness and generalizability of the proposed framework, future studies should evaluate the model on larger multicenter datasets, including all six dermatological disease classes, to further assess model stability and clinical applicability.

ACKNOWLEDGMENTS:

This research contains no funding.

AUTHOR CONTRIBUTIONS:

Ubaida Fatima conceived and designed the study, performed data collection and analysis, developed the proposed methodology, conducted the experiments, interpreted the results, prepared the figures and tables, wrote the manuscript, and approved the final manuscript.

DISCLOSURES:

The author declares no conflicts of interest related to this work.

REFERENCES:

1. Bickers DR, et al. The burden of skin diseases: 2004: a joint project of the American Academy of Dermatology Association and the Society for Investigative Dermatology. *J Am Acad Dermatol*. 2006;55(3):490–500.
2. Wehner MR, Levandoski KA, Kulldorff M, Asgari MM. Research techniques made simple: an introduction to use and analysis of big data in dermatology. *J Invest Dermatol*. 2017;137(8):e153–e158.
3. Szeto MD, et al. Dermatologic data from the Global Burden of Disease Study 2019 and the PatientsLikeMe online support community: comparative analysis. *JMIR Dermatol*. 2024;7:e50449. doi:10.2196/50449.
4. Javed R, Rahim MSM, Saba T, Rehman A. A comparative study of features selection for skin lesion detection from dermoscopic images. *Netw Model Anal Health Inform Bioinform*. 2020;9:4. doi:10.1007/s13721-019-0209-1.

5. Khan MA, et al. Skin lesion segmentation and classification: a unified framework of deep neural network features fusion and selection. *Expert Syst.* 2022;39(7):e12497. doi:10.1111/exsy.12497.
6. Karimi M, et al. Feature selection methods in big medical databases: a comprehensive survey. *International Journal of Theoretical & Applied Computational Intelligence.* 2025;2025:181–209.
7. Shehu Aliyu B, et al. Enhanced feature selection for imbalanced microarray cancer gene classification using chaotic salp swarm algorithm. *International Journal of Theoretical & Applied Computational Intelligence.* 2025;2025:258–283.
8. Nasir S, Bilal M, Khalidi H. Detection and classification of skin cancer using a CNN-enabled cloud storage data access control algorithm based on blockchain technology. *International Journal of Theoretical & Applied Computational Intelligence.* 2025;2025:145–169.
9. Laskaris N, et al. Fuzzy description of skin lesions [conference paper]. Presented at: SPIE Medical Imaging 2010; San Diego, CA, USA; 2010. Available from: <https://doi.org/10.1117/12.845294>
10. Ayo FE, et al. Severe acne skin disease: a fuzzy-based method for diagnosis. In: Gervasi O, et al., editors. *Computational Science and Its Applications—ICCSA 2020*. Cham: Springer; 2020. p. 320–334. Available from: https://doi.org/10.1007/978-3-030-58817-5_25
11. Fatima U, Khushal R. Enhancing explainability in epidemiological predictions using fuzzy logic integrated with machine and deep learning algorithms. *Sci Rep.* 2025;15(1):36139. doi:10.1038/s41598-025-10689-5.
12. Khushal R, Fatima U. Fuzzy logic and machine learning for diabetes risk prediction using modifiable factors. *Int J Adv Appl Sci.* 2024;11(12):225–231.
13. Khushal R, Fatima U. Fuzzy computing in healthcare [conference paper]. Presented at: 2024 International Visualization, Informatics and Technology Conference; Kuala Lumpur, Malaysia; 2024:78–83. Available from: <https://doi.org/10.1109/IVIT62102.2024.10692652>
14. Ilter N, Guvenir H. Dermatology [dataset]. UCI Machine Learning Repository; 1998. Available from: <https://archive.ics.uci.edu/dataset/33/dermatology>. doi:10.24432/C5FK5P.
15. Sari MO, Keser K. Classification of skin diseases with deep learning-based approaches. *Sci Rep.* 2025;15:27506. doi:10.1038/s41598-025-13275-x.
16. Baral S, Satpathy S, Satpathy RN, Kumar Baral A. Comparative evaluation of ensemble machine learning approaches for heart disease prediction. *J Vis Exp.* 2026;(230):e70124. doi:10.3791/70124.
17. Bologna JL, Schaffer JV, Cerroni L, editors. *Dermatology*. 4th ed. Elsevier; Philadelphia (PA); 2018.
18. Hall JE. *Guyton and Hall textbook of medical physiology*. 14th ed. Elsevier; Philadelphia (PA); 2021.
19. Tortora GJ, Derrickson B. *Principles of anatomy and physiology*. 16th ed. John Wiley & Sons; Hoboken (NJ); 2021.
20. Shahab N, Perveen K, Hussain M. Sticking to traditions: anatomy education and research in Pakistan. *Anat Sci Educ.* 2016;9(1):101–102.
21. Breiman L. Random forests. *Mach Learn.* 2001;45(1):5–32.
22. Cortes C, Vapnik V. Support-vector networks. *Mach Learn.* 1995;20(3):273–297.

- 957 23. Chen T, Guestrin C. XGBoost: a scalable tree boosting system [conference paper].
958 Presented at: 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data
959 Mining; San Francisco, CA, USA; 2016:785–794. Available from:
960 <https://doi.org/10.1145/2939672.2939785>
- 961 24. Mamdani EH. Application of fuzzy algorithms for control of simple dynamic plant. Proc
962 Inst Electr Eng. 1974;121(12):1585–1588.
- 963 25. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas.
964 1960;20(1):37–46.
- 965 26. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model
966 selection [conference paper]. Presented at: 14th International Joint Conference on Artificial
967 Intelligence; Montreal, Quebec, Canada; 1995:1137–1143. Available from:
968 <https://www.ijcai.org/Proceedings/95-2/Papers/016.pdf>
- 969 27. McNemar Q. Note on the sampling error of the difference between correlated
970 proportions or percentages. Psychometrika. 1947;12(2):153–157.
- 971 28. Rainio O, Teuvo J, Klén R. Evaluation metrics and statistical tests for machine learning.
972 Sci Rep. 2024;14:6086. doi:10.1038/s41598-024-56706-x.