Evaluation of Gradient Boosted Classifier in Atopic Dermatitis Severity Score Classification

Rivansyah Suhendra 1,*, Suryadi Suryadi 1, Noviana Husdayanti 2, Aga Maulana 3, Teuku Rizky Noviandy 3, Novi Reandy Sasmita 4, Muhammad Subianto 4, Nanda Earlia 5,6, Nurdjannah Jane Niode 7 and Rinaldi Idroes 8

1 Department of Information Technology, Faculty of Engineering, Universitas Teuku Umar, Aceh Barat 23681, Indonesia: rivansyahsuhendra@utu.ac.id (R.S.); suryadi@utu.ac.id (S.S.)
2 Teuku Umar Hospital, Aceh Jaya 23654, Indonesia: novianahdy@acehjayakab.go.id (N.H.)
3 Department of Informatics, Faculty of Mathematics and Natural Sciences, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia; agamaulana@usk.ac.id (A.M.); trizkynovianty@gmail.com (T.R.N.)
4 Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia; novireandys@usk.ac.id (N.R.S.); subianto@usk.ac.id (M.S.)
5 Department of Dermatology and Venereology, Faculty of Medicine Universitas Syiah Kuala, Banda Aceh, Aceh, Indonesia; nanda.earlia@usk.ac.id (N.E.)
6 Department of Dermatology and Venereology, Dr. Zainoel Abidin Hospital, Banda Aceh, Aceh, Indonesia;
7 Department of Dermatology and Venereology, Faculty of Medicine, University of Sam Ratulangi, Manado, 95163, Indonesia; niodejane@unsrat.ac.id (N.J.N.)
8 Department of Chemistry, Faculty of Mathematics and Natural Sciences Universitas Syiah Kuala, Banda Aceh 23111, Indonesia; rinaldi.idroes@usk.ac.id (R.I.)

* Correspondence: rivansyahsuhendra@utu.ac.id

Abstract

This study investigates the application of the Gradient Boosting machine learning technique to enhance the classification of Atopic Dermatitis (AD) skin disease images, reducing the potential for manual classification errors. AD, also known as eczema, is a common and chronic inflammatory skin condition characterized by pruritus (itching), erythema (redness), and often lichenification (thickening of the skin). AD affects individuals of all ages and significantly impacts their quality of life. Accurate and efficient diagnostic tools are crucial for the timely management of AD. To address this need, our research encompasses a multi-step approach involving data preprocessing, feature extraction using various color spaces and evaluating classification outcomes through Gradient Boosting. The results demonstrate an accuracy of 93.14%. This study contributes to the field of dermatology by providing a robust and reliable tool to support dermatologists in identifying AD skin disease, facilitating timely intervention and improved patient care.

1. Introduction

Atopic Dermatitis (AD), a chronic and frequently distressing skin condition, is a prominent dermatological concern affecting individuals of all ages, transcending geographical boundaries and cultural backgrounds [1, 2]. Characterized by symptoms such as pruritus (itching), erythema (redness), and lichenification (thickening of the skin), AD, colloquially known as eczema, exerts a
profund impact on the quality of life of those afflicted [3].

Statistical data shows the widespread prevalence of AD, emphasizing its status as one of the most common chronic inflammatory skin disorders globally. Statistics show that AD affects up to 2-10% of adults and 10-30% of children, making it a condition that spans generations [4, 5]. Its prevalence continues to rise, especially in low-income countries, highlighting the urgent need for practical diagnostic tools and interventions. In Indonesia, the prevalence of AD increases every year. Research [6] reported that the morbidity of allergic diseases in school children in metropolitan cities in Indonesia has the same pattern as in other developing countries. The research involved 499 children and teenagers from schools and universities in 5 cities.

AD presents as an intricate dermatological challenge, often manifesting with itching as its hallmark symptom. This relentless itch is accompanied by erythematous (red) skin patches, which may progress to excoriated lesions due to persistent scratching [5]. The condition is frequently found on flexural surfaces of the body, such as the inner elbows and behind the knees, though it can affect virtually any body part [7]. The pathophysiology underlying AD involves complex interactions between genetic predispositions, immune system dysregulation, and environmental triggers, making it a condition of substantial clinical and scientific interest [8]. As the prevalence of AD continues to rise, the demand for accurate and efficient diagnostic tools becomes increasingly urgent, aiming to alleviate suffering and improve the lives of those affected by this chronic skin ailment [2].

While the clinical presentation of AD is often distinctive, making an accurate diagnosis can be complex, particularly in milder cases or individuals with overlapping skin conditions. Dermatologists rely on clinical examination, patient history, and sometimes invasive testing to confirm AD. However, this process can be time-consuming and error-prone, leading to diagnostic uncertainties and potential delays in treatment [3].

Artificial Intelligence (AI) has brought significant advancements across a range of fields, including education [9], finance [10], environmental science [11], chemistry [12-15], agriculture [16], healthcare [17], and etc. In recent years, AI and machine learning (ML) researchers, specifically in the dermatology field, have demonstrated human-expert-level performance in various domains, including skin condition classification, lesion detection, and segmentation. This progress has prompted significant interest in the dermatology field, particularly in the context of AD. Researchers have explored multiple imaging techniques for AD detection, including multiphoton tomography, clinical images, and electronic health records. For instance, electronic health records have been investigated as a resource for AD diagnosis [18-20], while the utility of clinical images in AD diagnosis has been examined [21]. Moreover, multiphoton tomography has shown promise in AD detection [22].

Concurrently, similar attention has been directed towards addressing diagnostic challenges in skin pathologies like psoriasis [23], which share commonalities with AD regarding subjective scoring systems and time-consuming diagnostic processes. Dash [24] demonstrated the effectiveness of convolutional neural networks in segmenting psoriasis lesions with high accuracy, sensitivity, and specificity, surpassing existing methods. Additionally, Pal [25] highlighted the efficacy of convolutional neural networks in visual sign classification, an essential element in automated severity grading. Building on these foundations, [26] introduced a comprehensive approach combining lesion segmentation and severity grading, developing a computer-aided diagnosis (CADx) system for psoriasis lesion grading. These advances collectively underscore the significant potential of AI and machine learning in transforming the classification and diagnosis of dermatological conditions, including Atopic Dermatitis.

This underscores the pressing need for innovative approaches, such as advanced machine learning techniques, to enhance the accuracy and efficiency of AD classification from visual data. In this context, our research explores the application of Gradient Boosting machine learning to provide a robust and reliable approach for classifying the severity score, ultimately contributing to timely intervention and improved patient care of Atopic Dermatitis.

2. Materials and Methods

In the design of our methodology, as shown in Fig. 1, we adopt the data mining paradigm to utilize the potential of color features in AD image classification. We seamlessly integrate the machine learning paradigm within this framework, specifically utilizing the Gradient Boosting Classifier. This machine learning approach enables us to establish robust generalization capabilities, utilizing a meticulously curated training dataset comprising color image data and physician-classified labels [18].

These labels serve as the foundation for training our model, enabling it to learn and adapt to the complexities of AD classification. Subsequently, this trained model
predicts the unknown risk class of test color image datasets. This methodology incorporates 90 color space features extracted from various color spaces. These features form the foundation of our feature extraction process, allowing our classifier to discern intricate AD patterns. This comprehensive approach, seamlessly integrating data mining and machine learning paradigms, ensures the accuracy and effectiveness of our AD image classification model.

2.1. Data Acquisition and Preparation

This study's AD image dataset was procured from the Dermatology Division of Zainoel Abidin Regional Public Hospital. The data collection process involved digitally photographing areas of patients' bodies displaying AD lesions. Subsequently, a dataset preparation phase was undertaken, wherein skin images were manually cropped from each patient's photographs. This process ensured that the dataset exclusively comprised relevant AD lesion images, eliminating extraneous information.

Our dataset comprises a total of 500 image samples, thoughtfully categorized into four distinct severity scores: "none," "mild," "moderate," and "severe" with total of 129, 184, 106, 81 samples for each class respectively. Moreover, dataset labeling was carried out using the SCORAD intensity standard score, ensuring accurate categorization and establishing a robust foundation for subsequent phases of our research.

2.2. Data Preprocessing

In the data preprocessing phase, our primary aim was to prepare the AD image dataset to ensure effective training and classification. To achieve this, we implemented several pivotal steps to enhance data consistency and elevate the dataset's overall quality.

The first step involved uniformly resizing all images in the dataset to a resolution of 250x250 pixels. This resizing process had a dual benefit: it standardized image dimensions across the entire dataset, enabling a seamless and consistent analysis and facilitating model training by ensuring a uniform input format.

Additionally, we employed augmentation techniques to enhance the robustness of our model and its ability to generalize effectively. Specifically, we utilized skew transformation, a controlled variation technique that introduced subtle alterations in image orientation while...
preserving the essential characteristics of AD lesions. We followed a systematic approach for model evaluation, splitting the dataset into two sets: a training set (comprising 80% of the data) and a validation set (comprising 20% of the data). This partitioning enabled us to train and fine-tune our Gradient Boosted classifier effectively.

2.3. Feature Extraction

Our primary action in the feature extraction phase was systematically extracting features from the AD images. The extraction has been done by computing two statistical components (mean and standard deviation) for each color space, including RGB, normalized-RGB, YCbCr, HSV, HLS, CIE XYZ, CIE Lab, CIE Lch, opponent, CMY, YUV, VIQ, YDbDr, and YPbPr, resulting in a feature set of 90 color features. All these color spaces are known in the literature [27–29]. These statistical components provided valuable insights into the data distribution and variation within each color space, effectively enhancing our feature set and facilitating our classification model's discernment of critical AD patterns.

2.4. Gradient Boosted Classifier Approach

Our methodology involved applying the Gradient Boosting machine learning approach for AD image classification. We leveraged this powerful technique to enhance the classification accuracy of AD skin disease images and reduce the potential for manual classification errors. The Gradient Boosting Classifier (GBC) represents an ensemble classifier strategy designed to reduce errors through resampling and adjusting the weights assigned to individual weak learners, ultimately enhancing classification accuracy [30].

The GBC demonstrated superior overall performance in a practical assessment comparing various supervised learning algorithms, including Random Forests and Boosted Decision Trees [31]. Encouraged by this notable success, the authors chose gradient boosting as their classifier. One significant advantage of the GBC is its capability to produce outputs in severity score probabilities, offering valuable insights for clinical interpretation. Furthermore, GBCs provide a means to gauge the relative influence of each parameter integrated into the classifier, adding a useful layer of transparency and interpretability to the classification process [32].

2.5. Model Evaluation

We employed a comprehensive set of performance metrics to evaluate the classifier's performance. These metrics include accuracy, recall, precision, and the F1 score. Accuracy measures the overall correctness of classifications, while recall quantifies the classifier's ability to identify true positive cases correctly. Precision assesses the classifier's capability to minimize false positives, and the F1 score balances precision and recall to provide a comprehensive measure of classifier performance [33].

We employed cross-validation techniques to ensure robustness and minimize the risk of overfitting. By splitting the dataset into multiple folds and iteratively training and testing the model on different subsets, we more accurately assessed the model's generalization capabilities. This approach provides a more reliable estimate of the model's performance on unseen data. Additionally, we generated Receiver Operating Characteristic (ROC) curves and calculated the Area Under the Curve (AUC). These metrics provide insights into the classifier's ability to discriminate between different AD severity levels. A higher AUC indicates improved discriminatory power. GBCs also provide a metric to assess the relative influence of each parameter included in the classifier. This analysis enables us to identify the most critical features and parameters contributing to the classifier's decisions, enhancing transparency and interpretability.

3. Results and Discussion

The primary objective of this paper is to assess the effectiveness of the Gradient Boost Classifier within the framework of classifying AD disease using color features.

3.1. Gradient Boosting Model Performance

We employed the Gradient Boosting machine learning technique for AD classification, utilizing the feature-rich dataset prepared through rigorous preprocessing and feature extraction. The model was trained with an approach, incorporating k-fold cross-validation with ten folds to fine-tune the parameters for optimal AD classification. Through this process, we identified the most effective parameters, which include: Learning Rate: 0.3, Maximum Depth: 5, Maximum Features: 1.0, Minimum Samples per Leaf: 3, Minimum Samples per Split: 7, Minimum Weight Fraction per Leaf: 0.0, Number of Estimators: 160, and Criterion: 'friedman_mse'. These parameter settings were crucial in achieving the better classification performance observed in our results.

Figure 2 presents the Confusion Matrix for the validation data, showcasing the performance of the Gradient Boosting Classifier model in classifying AD images across different severity levels. The matrix highlights the model's capacity to categorize samples into their respective classes accurately.
Table 1. Performance of several models on the validation data.

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
<th>AUC</th>
<th>Recall</th>
<th>Precision</th>
<th>F1-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM - Linear Kernel</td>
<td>.551</td>
<td>.000</td>
<td>.551</td>
<td>.526</td>
<td>.492</td>
</tr>
<tr>
<td>Naive Bayes</td>
<td>.654</td>
<td>.876</td>
<td>.654</td>
<td>.678</td>
<td>.643</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>.774</td>
<td>.923</td>
<td>.774</td>
<td>.786</td>
<td>.771</td>
</tr>
<tr>
<td>Linear Discriminant Analysis</td>
<td>.794</td>
<td>.931</td>
<td>.794</td>
<td>.811</td>
<td>.791</td>
</tr>
<tr>
<td>K Neighbors Classifier</td>
<td>.820</td>
<td>.954</td>
<td>.820</td>
<td>.834</td>
<td>.815</td>
</tr>
<tr>
<td>Decision Tree Classifier</td>
<td>.851</td>
<td>.895</td>
<td>.851</td>
<td>.859</td>
<td>.850</td>
</tr>
<tr>
<td>Gradient Boosting Classifier</td>
<td><strong>.931</strong></td>
<td><strong>.989</strong></td>
<td><strong>.931</strong></td>
<td><strong>.936</strong></td>
<td><strong>.929</strong></td>
</tr>
</tbody>
</table>

Figure 2. Confusion matrix on validation data.

As shown from the Confusion Matrix, the GBC model has demonstrated remarkable classification results. Each class exhibits an average True Positive (TP) rate exceeding 94%, underscoring the model's proficiency in correctly identifying AD cases within these categories. Notably, the model exhibits minimal False Positives and False Negatives, indicating its high precision and reliability.

As seen in Figure 2, there is 1 image in the mild category which is classified as severe. This can happen because there are outlier data in the mild class data so that the model fails to classify according to its class. The misclassification of a single image of mild AD as severe can occur due to the inherent complexity and variations in AD images. Mild and severe cases can sometimes exhibit overlapping visual characteristics, making them challenging to distinguish.

In addition to assessing the performance of our GBC model, we conducted a comprehensive comparative analysis that involved other popular machine learning algorithms. These algorithms included SVM with a Linear Kernel, Naive Bayes, Logistic Regression, Linear Discriminant Analysis, K Neighbors Classifier, and Decision Tree Classifier. We aimed to evaluate how our GBC model fared against these alternative methods in classifying AD images.

The performance comparison results are summarized in Table 1 that presents a detailed overview of the performance metrics of all the models considered in our analysis. It is worth noting that the GBC model achieved the most optimal values for each parameter, including accuracy, AUC (Area Under the Curve), precision, recall, and F1-score. These results highlight the superior performance of the GBC model in comparison to other machine learning algorithms. The GBC model consistently outperformed its counterparts across all relevant metrics, indicating its effectiveness in AD image classification.

The superior performance of the Gradient Boosting Classifier (GBC) model can be attributed to its ensemble learning approach, which combines multiple weak learners for robustness and improved generalization compared to others approach. GBC's gradient boosting optimization minimizes errors by adjusting weights during training, making it effective for complex tasks like Atopic Dermatitis classification.

As the final analysis of the GBC model's performance, we conducted an ROC Curve analysis for each severity score class in AD classification. The results were indicative of the model's exceptional discriminatory power and precision.

The ROC curve is a crucial evaluation tool in binary and multi-class classification tasks, including the classification of AD severity levels. It offers valuable insights into a classification model's performance by illustrating the trade-off between the True Positive Rate (TPR or Sensitivity) and the False Positive Rate (FPR or 1 - Specificity) at various decision thresholds. A perfect classifier's ROC curve passes through the top-left corner (0, 1), indicating maximum TPR (Sensitivity) and no FPR. The diagonal line (from bottom-left to top-right) represents random guessing, and a model that performs no better than random closely follows this line. The farther the ROC curve is from the diagonal line and closer to the top-left corner, the better the model's ability to distinguish between positive and negative cases.

The Area Under the ROC Curve (AUC) is a quantitative metric derived from the ROC curve. It provides an overall
measure of the model’s performance. A higher AUC value indicates better discrimination ability, with an AUC of 1.00 signifying perfect classification and 0.50 indicating no discriminatory power.

Figure 3 visually represents the ROC Curve for each severity score class. The AUC results for the GBC model were exceptionally high, with values of 1.00 for “none,” “moderate,” and “severe,” and 0.99 for “mild.” This analysis offers valuable insights into the performance of the Gradient Boosting Classifier model across different severity levels in Atopic Dermatitis classification.

### 3.2. Analysis of Color Feature

The significance of color features in the GBC model for AD classification became evident in our analysis. Color information plays a pivotal role in dermatological image analysis, as it encapsulates valuable insights into the visual characteristics of skin lesions. Color features empower the GBC model to discriminate between different AD severity levels effectively, enhancing its accuracy and reliability.

Figure 4 offers a visual representation of the eight most crucial features. The result reveals that the feature “MEAN_YIQ_Q,” denoting the mean value of channel Q in the YIQ color space, stands out as the most critical feature from GBC’s model. This feature has been assigned a noteworthy importance score of 0.17, underscoring its pivotal role in the model’s decision-making process. These features have proven instrumental in the Atopic Dermatitis classification process within the GBC model. The importance of these features is evident in their capacity to significantly influence the model’s decision-making and classification outcomes. While the specific details of the features are outlined in the figure, it’s worth noting that these attributes contribute to the model’s ability to categorize AD images into different severity levels accurately.
4. Conclusions

In conclusion, our study delves into the domain of Atopic Dermatitis image classification, primarily focusing on applying the Gradient Boosting Classifier model. AD, a prevalent and chronic skin condition, presents multifaceted challenges in diagnosis and classification. Our research addresses these challenges by meticulously exploring color features within AD images. Our study's results demonstrate the GBC model's remarkable potential in accurately categorizing AD images across varying severity scores. With an impressive accuracy of 93.14% and exceptional AUC values, the model showcases its robust discriminative power. The significance of color features underscores the pivotal role of visual information in dermatological image analysis.

Our findings emphasize the relevance of advanced machine learning techniques, such as Gradient Boosting, in augmenting the diagnostic capabilities of dermatologists. The GBC model's reliability and precision offer valuable support to healthcare professionals, contributing to timely interventions and improved Atopic Dermatitis patient care. While this study has yielded promising results, it is essential to acknowledge its limitations. Firstly, the study relied on a single dataset for training and evaluation, which may limit the model's generalizability to different patient populations or image acquisition conditions. Additionally, variations in image quality, lighting, and patient demographics may have introduced some level of variability in the dataset, potentially affecting classification performance. To address these limitations, future research should involve larger and more diverse datasets, including data from multiple healthcare institutions.


Funding: This research was funded by Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi through the “Penelitian Dosen Pemula” scheme, grant number: 107/UN59.7/PG.02.00.PT/2023 and “Penelitian Fundamental Reguler” Scheme with grant number: 168/ES/PG.02.00.PL/2023.

Ethical Clearance: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: All the authors declare no conflicts of interest.

References


