VFAST Transactions on Software Engineering
http://vfast.org/journals/index.php/VTSE@ 2023, ISSN(e): 2309-3978, ISSN(p): 2411-6246
Volume 11, Number 1, January-March 2023 pp: 109-122

Classification of Skin Disease using Machine Learning

Azka Ahmed[1], Hafsa Ahmad[1], Mohsin Khurshid[2], Kamran Abid[1]

[1]Department of Computer Science, NFC Insitute of Engineering and Technology, Multan, Pakistan
[2]Department of Computer Science, The Islamia University of Bahawalpur, Pakistan
*Corresponding author email: azkaahamad289@gmail.com

ABSTRACT

Erythemato-squamous disease (ESD) is one of the dermatology field's complex diseases. Due to its common morphological features, it is challenging to diagnose and generally produces inconsistent results. In addition, the physician's expertise was used to make the diagnosis based on the observed symptoms. The accurate classification of erythemato-squamous disorders is one of the dermatology field's problems that need attention, and to help with this issue, by using clinical and histopathological data, this tool will differentiate the six classes of ESD. In this research, we have applied 3 different machine learning algorithms as base models i.e. Random Forest, Decision Tree, and Naïve Bayes to classify the ESD and 5 Ensemble Meta techniques such as Voting classifier, average classifier, Stacking, boosting, and bagging classifiers to measure the accuracy. In comparison to other classifier methods, the ensemble technique employed on dermatology dataset, original dataset and clinical feature extraction to identify which model performs better on both cases. The ensemble method provides a more precise and accurate prediction of skin diseases.

KEYWORDS:

Skin Disease, Ensemble Learning, Erythemato-Squamous, Machine learning.

JOURNAL INFO

HISTORY: Received: December 15, 2022 Accepted: March 10, 2023 Published: March 31, 2023

INTRODUCTION

The skin is the most essential part of the human body [3]. The skin helps in the production of vitamin D and protects the body from different harms such as UV radiation, various infections, wounds, heat, etc. [4]. Skin disease is one of the serious health issues these days, that frequently affects a person's quality of life in terms of their health [1]. In contrast to other infections like malaria, pneumonia [2], typhoid, etc., skin disorders receive less consideration even though they can have lethal implications. Some skin disorders can be mild, temporary, and effectively handled, while others can be quite dangerous and fatal, even for skilled medical professionals, detecting dermatological illnesses is a challenging task [5]. In dermatology outpatient clinics often see patients with Erythemato-Squamous disorders. The erythemato-squamous is a class of dermatological illnesses that causes redness in the skin due to inflammation and high blood flow in the affected area. ESD is presented as scaly inflammation like itchy dry skin, and dandruff it mostly affects the face and scalp.

ESD is a group of 6 skin disorders. “Seboric dermatitis, psoriasis, pityriasis rosea, lichen planus, chronic dermatitis, and pityriasis rubra” are included in this skin disease group [7]. These diseases are scaled with an extremely small difference. This disease can be affected by medicine for acne, UV radiation, asthma, bacterial and fungal infection, etc. This skin disease can affect any part of the body, especially elbows, feet, hands, ankles, knees, etc. The symptoms of erythemato-squamous can be itching, scaling, rashes, swelling, and burning [8].

A dermatologist with an extensive and relevant understanding of these disorders is required for the patient. The condition usually resembles erythema and scaling. During thorough observation, the ESD may present differently in various patients and may have distinct clinical symptoms in various body parts. The assessment of the erythemato-squamous disorder in dermatology is difficult [6]. Patients are initially clinically assessed according to 12 features which are crucial indicators in the classification of ESDs. The features are the degree of scaling and erythema, to define the current condition of "lesion borders, itching, and the Koebner phenomenon, papule formation, family history, and involvement of the oral mucosa, knees, elbows, and scalp”.

The ESD is classified into six classes. Psoriasis is a class of ESD that causes flaky patches on the skin. Psoriasis affects the elbow, scalp, and knee while lichen is predisposed to the oral mucosa [9]. Pityriasis rubra pilaris typically affects children. There is typically a familial record of psoriasis. Some patients are only diagnosed based on these clinical characteristics, although a biopsy is typically used to make a conclusive and correct diagnosis [10]. Pityriasis rosea and lichen planus are two different classes that might resemble psoriasis. Another class of ESD that can lead to skin issues is chronic dermatitis, which often results in less redness and scaling than psoriasis. It's essential to appropriately diagnose the skin issue to provide the correct treatment. Lichen planus exhibits irritation and polygonal papules. However, in pityriasis rubra pilaris, follicular papules are seen. Another challenge in the diagnostic process is when a disease exhibits the histological characteristics of another disease and develops distinctive traits later on. While some samples might not, some may demonstrate the typical histopathological symptoms of the disease. Chronic dermatitis is indicated by papillary dermis fibrosis, and lichen planus can be identified by melanin incontinence. Similarly, lichen planus is identified by the presence of a band-like infiltration, loss of the granular layer, saw-toothed rete ridges, vacuolization, and basal layer damage [11].

For the precise and effective classification of ESD, machine learning-based algorithms might be applied [12]. A machine learning model involves several stages, including gathering a dataset and downloading pertinent and verified data from an online source. Choosing the best feature out of those that are accessible can help to shorten processing times and simplify computations [13]. A machine learning model could experience the "curse of dimensionality" because it has more unnecessary features. Depending on the learning method, a machine learning model may classify into one of three groups.

A supervised machine learning method is a model that learns from prior knowledge or specific target samples. Both predicting a continuous value, known as the regression problem, and predicting a discrete value, known as the classification problem, fall under the category of supervised learning approaches. The use of machine learning algorithms is widespread in the medical industry for both disease prediction and clinical diagnosis. Initially, many tools are developed by using machine learning algorithms to detect various diseases e.g. breast cancer [14], thyroid diseases [15], kidney disease, Skin cancer[16], diabetes [17], liver disease [18], Covid [2], and brain tumor[19]. Various disease prediction models use classification to achieve a high accuracy rate. One of the effective machine learning techniques is ensemble learning, which combines the performances of numerous weak / base learners to improve the performance of models. It has a huge amount of potential to address a variety of issues in the healthcare field by providing medical systems to use the information statistically to uncover the best strategies that enhance healthcare and at the same time lower costs. The primary goal of the current work is the more accurate automatic detection of ESD.

To determine the class, machine learning approaches including a Decision tree classifier, Random Forest Classifier, Naïve Bayes Classifier, and various Ensemble Meta classifiers will be applied. There are two categories of machine learning: supervised and unsupervised. However, we do not classify the dataset using unsupervised learning. Voting Based Ensemble Model, Average Based Ensemble Model, Stacking Based Ensemble Model, Bagging Model, and Boosting Model were the five ensemble models that we used. The model's accuracy has been carefully examined. The dataset includes several features that can be used to classify skin disease. These models are applied to both feature extraction techniques and without feature extraction techniques to compare their accuracies and to find which technique provides the best accuracy.

The main contribution of this work is the creation of an Ensemble Meta strategy for the detection of ESD, which hasn't yet been discussed in the literature. Additionally, we discover that there haven't been many studies about the use of EMT for ESD classification described in the literature. Although there are numerous papers in the literature that used typical machine learning techniques to diagnose ESD, including Random forests, SVM, decision trees, and more. In this study, a real-world dermatology dataset is used to apply the Ensemble Meta approach and other traditional machine-learning techniques. When prediction accuracy is taken into account. In this research, we will compare different ML algorithms to attain a high accuracy rate. Also examining the significance of the feature space is the main research question this study addresses. With a cardinality of 12 and 22 features, respectively, there are two fundamental sets of features: clinical and histological.

The rest of the article is organized as follows: Section 2 provides a detailed literature review about the prediction of skin diseases using various machine learning algorithms. While Section 3 presents the dataset description and proposed methodology. The results of the experiments are discussed in Section 4. Finally, Section 5 will conclude the article.

LITERATURE REVIEW

Much research has been proposed in the literature for automated diagnosis of the type of ESD. The literature has a variety of approaches. Such research can be divided into two major groups. The first group makes suggestions for various classifiers, their improvements, and combinations, while the second group focuses on choosing the optimum combination of attributes to differentiate the diseases. These works include “decision trees, support vector machines, artificial neural networks, and more”. In this section, we'll explore more detail about these works. Machine learning has seen great success in several industries recently thanks to the increase in processing power, the accessibility of inexpensive memory devices, and cloud computing.

The literature describes various proposed automated systems for differential diagnosis of ESD by using machine learning.

In [20] elephant herding algorithm was used with the parameters of support applied to identify the ESD. The authors of [9] combined neural networks and fuzzy logic's learning capabilities. To properly categorize the dataset into the correct classes of ESD, the integrated classifier was applied. Furthermore, to predict the kind of ESD.

The importance of definite and accurate classification of erythemato-squamous is mandatory because various cases of coronavirus show a relation with erythematous rashes on the skin. Coronavirus disease 2019 (COVID-19)-related skin symptoms have increased in frequency over time, concurrent with the coronavirus 2 associated with severe chronic pulmonary syndrome's widespread occurrence. Many cases were reported [21], [22], Several erythematous-squamous papules and plaques were seen during a dermatological examination, mostly on the upper limbs and trunk. Researchers[23] discuss the cases of two patients of COVID-19 diagnosed with Pityriasis rosea-like rashes. They also highlight those young adults are the most affected.

Anifah and Haryanto employed the Linear Vector Quantization approach to conducting the classification study in 2021[24]. The research conducted by Alotaibi, Abdullah S [25] in 2022 uses a hybrid algorithm KNN and ReliefF approaches, they have established that the suggested model is useful for diagnosing the condition under investigation.

In a study conducted in 2017 [26], a data mining algorithm was applied to the UCI machine repository by using boosting, begging, Bayes net, and naïve Bayes on only clinical features of eryhemato-squamous. They claimed that these algorithms are better algorithms for implementation when considering simply the clinical features of the erythematous disease and the authors make the conclusion that, in comparison to previously reported findings, their strategy is quite promising for clinical attributes.

The impact of altering the ratio of the training set to the testing set has been studied in [27] while comparing the two models based on the Extreme learning model (ELM) and SVM. According to experimental findings, ELM performed better than SVM for the testing set case on every front. High accuracy and consistency of results, especially in the face of variable training-testing dataset percentages, show that the proposed classifiers are practical tools in this important subject. SVM and ELM have thus proven to have a special capacity for achieving outstanding outcomes in the realm of biomedical diagnosis.

For the automatic detection of ESD, a novel strategy based on the application of Self Organizing Map (SOM) was proposed [28]. To determine the severity of erythemato-squamous dermatological diseases, researchers used various clustering approaches. In this study, the authors claimed that the characteristics accurately describe erythemato-squamous diseases, and the SOM algorithm's task obtained good classification accuracy. For psoriasis, seboreic dermatitis, lichen planus, and pityriasis rosea, the best accuracy was 85, 94%, 40,48%, 56,25%, and 82,61%, respectively.

[11] The derm2vec approach is a proposed hybrid deep learning for the detection of Erythmato-Squamous disease. The authors combined DNN and auto-encoder approaches to create the Derm2Vec. This Derm2Vec approach provides the best accuracy mean of cross-validation score is 96.92%

Using the Classification and Regression Tree to forecast the differential diagnosis of ESDs was the goal of researchers [29]. For the training and testing data sets, the classifier's accuracy was 95.31% and 93.69%. The findings of this study show that employing the CART classifier does not present adequate accuracy when compared to other classifiers. However, the authors strongly advised that ensemble approaches be utilized to categorize the differential diagnosis of ESD in their future work.

Table 1 Literature Review

References	***Model/ Algorithm***	***Dataset***	***Findings***
[30]	Linear Vector Quantization	UCI machine repository	Classification of ESD classes was performed by using data mining techniques.
[25]	ReliefF & k-nearest neighbors	UCI machine repository dermatology dataset	A comparison between traditional KNN and hybrid KNN is performed.

[26]	SVM, logic boost, MLP Begging, Bayes net and naïve bayes	UCI machine repository	Compare the algorithms to predict the best accuracy of ESD based on clinical features only.
[31]	Hybrid feature selection algorithm, beetle Swarm optimization	UCI-machine repository	This algorithm provides an accuracy rate of 95%.
[32]	Multi-layer feed-forward and ANN	UCI-machine repository	ANN algorithm produces optimum results
[33]	ANN model	UCI-machine learning repository	The accuracy rate of 98.3% on 6 groups of diseases.
[34]	Genetic algorithm with generalized wrapper-based feature selection	Dermatology dataset	97.6% of the accuracy rate was reported for the dermatology dataset.
[35]	SVM, FNN, Multilayer perceptron, and random forest	Dermatology dataset	This algorithm results in a maximum accuracy of 97%.
[36]	Fuzzy Logic	UCI-machine Learning repository	This fuzzy expert system reported an optimum accuracy rate.

[37] In this article, authors develop an "automated ontology framework" using data mining methods and then diagnose ESD using "knowledge-based and rule-based systems". To investigate the prediction of the ESD, the rule language (SWRL) and rule engine (Jess and Drools) were merged. 95% of performance accuracy was reported. The authors also reported in this study the comparison of two rule engines using some pre-selected criteria. The advantages and disadvantages of the rule engines are also examined in this research article in light of this comparison.

Using an SVM with a linear kernel to classify the ESD, we have worked to develop a diagnostic technique in this research [38]. SVM using linear, RBF, and polynomial kernels was investigated for this experiment. On the test set, SVM with a linear kernel has a 95.3125% accuracy rating.

[39] A mobile medical tool to assist doctors with psoriasis diagnosis using solely clinical criteria. The study compared and discovered six erythemato-squamous illness classification methods (SVM, NB, NN, OMK, and AB). In these algorithms, histopathological characteristics are utilized. This study created a mobile/desktop application with these 6 algorithms that can accurately identify these disorders into 2 categories—psoriasis or non-psoriasis—using only clinical criteria. 8 Doctors could use this mobile app to distinguish psoriasis from clinically identical illnesses.

A literature review was conducted in [40]. The purpose of a literature review was to find and critically evaluate studies that compared in-person Clinical classification of the PASI (psoriasis area and severity index) versus AI algorithms created to calculate the PASI in adult chronic patients with psoriasis. The techniques for erythema individually, erythema plus scale, and area PASI rating were discussed in 3 articles, 32 of which were identified. The investigations were tiny, and only one of them received outside validation. None of the scoring algorithms for the four PASI components, nor those that evaluated induration, were validated (a major issue for AI-based evaluation). Measures of psoriasis severity without the use of induration might be better suited for AI evaluation.

The authors of [41] offered four data mining methods for identifying skin conditions. The author took into account Decision Trees and Random Forests, two tree-based algorithms. Additionally, the author used K-Nearest Neighbor, a distance-based classifier, and Multi-Layer Perceptron, a neural classifier (KNN). The Random Forest algorithm, which has an accuracy of 98.64%, is a successful mechanism among the aforementioned techniques. This approach may suffer from computational expense because feature selection was not taken into account.

In [24] a comparative analysis was provided by applying various algorithms like logistic regression, KN classifiers, SV classification, GNB, DTC, and RFC. Furthermore, for results, a 5-fold method of cross-validation was applied. The results obtained from these machine learning methods were satisfactory.

A hybrid model was produced by combining the various approaches with feature optimization [25]. Initially, the Relief Algorithm is used as a hybrid model to pick the best features from the dataset, and KNN is used to predict those best features. This experimental result produced 9r5% accuracy.

Data mining approaches have been used to classify the classes of ESD in another article [26]. The algorithm was applied only to the clinical features of the dataset. The best accuracy provided only on clinical features by applied algorithms was 84.5%.

In [30] a decision support system has been proposed by the authors, this system is based on linear vector quantization. In this article, only clinical attributes of the ESDs are discussed. Results from this system are not satisfactory only one cluster out of six provides better accuracy in this regard. So, it can be concluded that selecting only clinical features from the dataset would not be sufficient for the detection of this disease.

Computer-aided methods are made to diagnose skin cancer at its earliest stage. Authors [31] use the preprocessing stage, which has a bigger impact on classification performance, as the most important phase in it. The relevant features from a set of qualities have previously been found using a variety of feature selection algorithms. To create a hybrid feature selection method for choosing pertinent feature subspace from datasets related to dermatology. The "Latent Semantic Index (LSI) and correlation-based Feature Selection" are combined to create the hybrid feature selection technique. "Beetle swarm optimization" is used to find the best feature subset. 95% and 92% are the obtained values for accuracy and sensitivity.

To categorize the disorders associated with dermatology, researchers [32] applied a neural network-based approach to the dermatological dataset that was downloaded from the UCI repository site. The best outcomes for classification and prediction issues are produced by artificial neural networks with backpropagation algorithms. It also has the capacity for generalization and is adaptable to problems encountered in the actual world. The study was expanded by using additional disease datasets, and a neural network-integrated automated diagnostic and advising system is unquestionably helpful in the problem of disease prediction.

To accurately categorize the six classes of ESD, the researchers of this article [33] employ artificial neural networks. They classify ESD for modeling using a backpropagation feed-forward algorithm. The UCI machine learning repository's dermatological dataset was gathered for this purpose. Using the pre-processed dataset, they trained and validated the suggested model before evaluating it. A 98.36% accuracy rate was attained using the suggested model. The proposed classifier is effective, according to the results.

Five ensemble classification algorithms for erythemato-squamous illnesses are compared in this study. We decided to categorize this collection of illnesses into the original dataset for dermatology's six classifications. This research is based on two steps.

When completing the classifications and computing the findings, both histopathological and clinical features from the Dermatology dataset were taken into consideration. Based on classification accuracy, recall, F-1 score, and precision, the classification algorithms were compared. Only clinical features are also tested for classification to determine the accuracy, recall, F-1 score, and precision, while ensemble classification algorithms were compared.

I. METHODOLOGY

The system uses machine learning models with various algorithms to predict erythemato-squamous dermatological diseases.. Each algorithm approaches the final prediction differently. This might be able to explain how each algorithm works. A flow chart is a type of diagram that shows a process, or a method. It connects different kinds of boxes to documents to show the order of the steps.

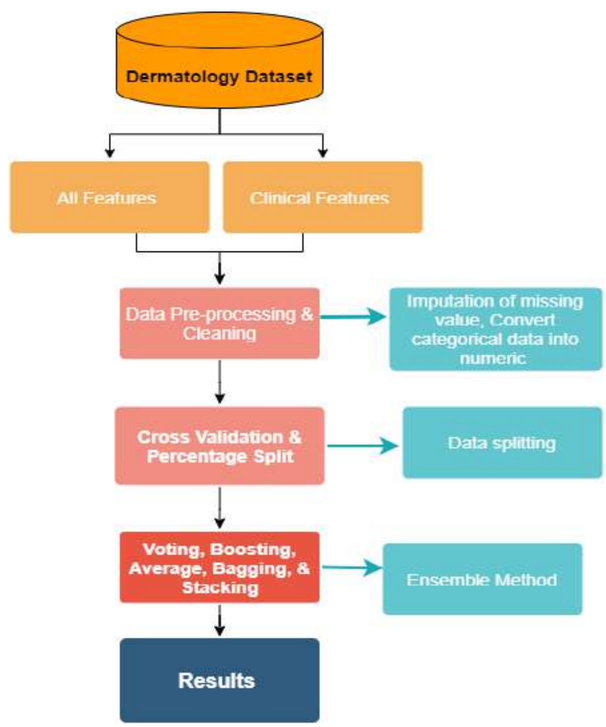


Figure 1. Proposed Framework

Performance Evaluation Framework

To assess the effectiveness of the suggested framework the figure is given below that describes each step. The ensemble method is used on the dataset to investigate its implication for the performance of the classifier. All of these steps increased the accuracy ratio of certain ML models. The methodology is based on two categories with feature extraction and without feature extraction. To evaluate the performance of the framework the ensemble learners are applied to the raw dataset that contains all the values and features of the dataset. And in feature extraction only clinical features from the dataset are selected. The accuracy of a raw dataset and the extracted feature will be compared to identify the best accuracy. The preprocessing steps and cross-validation method and percentage splitting will be the same. Six classes of erythemato Squamous are present in the dataset during feature extraction the classes remain the same no alteration is done on the dataset to make sure credibility. We analyzed the characteristics to create a flow chart for the machine learning techniques used on the dataset.

Dataset Description

The dataset used in this study is from the UCI Machine Learning repository and is owned by H. Altay and Nilsel Ilter. The dataset is used to identify the different classes of ESD. The dataset has 34 attributes, of which 33 attributes are linear and 1 attribute (Age) is noticed as nominal. Features of the dermatology dataset are divided into parts, (1) clinical features and (2) histopathological features. All the histopathological and clinical features are assigned the values 0, 1, 2, and 3. Where 0 means the absence of features, 1 and 2 are intermediate values and the highest amounts are assigned to 3. The dataset consists of 366 records. The investigated database includes a family history feature with a value of one; if none of these diseases run in the family, it has a value of zero. Age, or the patient's age, is another characteristic linked to the domain.

Table 2 Dataset Description

Class	No. of Instances	Clinical features	Histopathological Features
4:Psoriasis	112	F1. Erythema F2. Scaling F3.Definite borders	F12. Melanin incontinence F13. Eosinophils in the infiltrate F14. PNL infiltrate F15. Fibrosis papillary dermis F16. Exocytosis F17. Acanthuses F19. Parakeratosis F20. Clubbing the rete ridge
5:Sebrrhenic dermatitis	61	F4. Itching F5.Koebner phenomenon F6. Polygonal papules	F21 elongation of the rere ridge F22. Thinning of the suprapapillary epidermis F24. Munro microabscess
2: Pityriasis rosea	49	F7.Follicular papules F8.Oral mucosal involvement F9.Knee and Elbow involvement	F25. Focal hypergranulosis F26. The disappearance of the granular layer F27. Vacuolization and damage of basal layer F28: follicular horn plug F29. Spongiosis F30. Per follicular par keratosis F31. Inflammatory mononuclear infiltrate F32: band-like infiltrate
0:Chronic dermatitis	52	F10.Scalp involvement F11.Family history	F33. Spongiform pustule
1:Lichenplanus 3:Pityriasis rubra	72 20	F34. Age	

RESULTS AND DISCUSSION

Comparative Analysis

In this part, a comparative analysis is conducted after each of the five models has been assessed using the four matrices on the original dataset and feature extracted dataset.

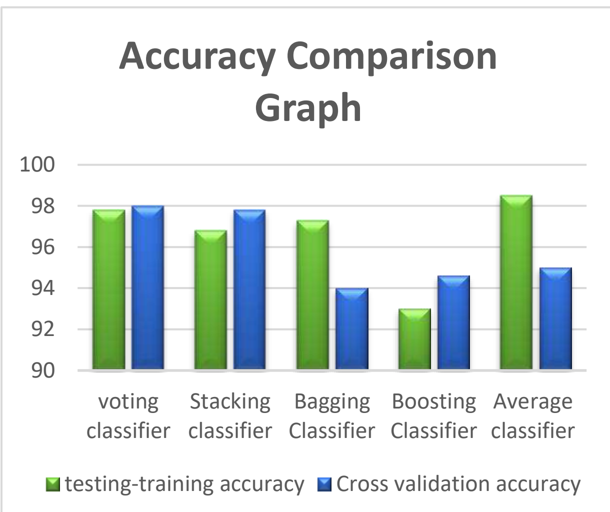
Accuracy Comparison

Accuracy data for five Voting Classifiers, Stacking Classifier, Average, Bagging classifier, and Boosting ensemble models are presented in the graph. Each model includes two distinct levels of accuracy: one for cross-validation and the other for training-testing splits. After the comparison of all classifiers, Voting Classifier provides the best Accuracy rate in both cross-validation and test and training accuracy.

Table 3 Accuracy of Ensemble method

Model	Cross-Validation Accuracy (%)	Training –testing accuracy (%)
Voting Classifier	98	97.8
Stacking Classifier	97.8	96.8
Bagging Classifier	94	97.3
Average Classifier	95	93
Boosting Classifier	94.	94.6

The table shows how accurate various models are. The model Voting classifier and the Bagging classifier have the highest accuracy on training and testing data (97.8%) and 97%, while the model Average Classifier has the lowest accuracy (93%). Similarly, the accuracy of cross-validation of the model voting classifier shows the highest percentage of 98 and Bagging is the least accurate model with 94%.



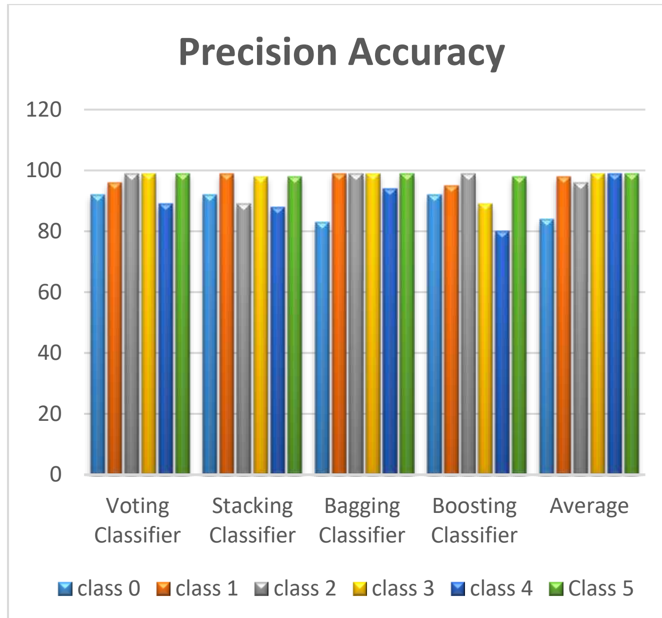
Graph 1 Accuracy Comparison

In the graph above, two metrics cross-validation and training testing are used to compare the accuracy of five models. The Boosting Classifier, Stacking Classifier, Voting classifier, Bagging, and average classifier were used to achieve the results.

A. Precision Comparison of Models

The precision score of all classifier models is presented in Table 4.16. The voting classifier performs well in classes 2, 3, and 5 with an accuracy of 99% and shows a low performance in class 0 with 92% accuracy. The accuracy of the predictions made by stacking classifiers is >97% for class 1,3 and 5 and 89% is the lowest prediction rate for class 2; Class 3, 4, and 5 has the highest accuracy according to the Average Classifier, with a precision score of 99%, whereas class 0 has a low precision score of 81%.

The Accuracy of the Boosting Classifier's predictions is highest in classes 0, 1,2, 3, and 5 (>90%), and lowest in class 4 (80%). Bagging Classifier performs admirably in classes 1, 2, 3, 4, and 5 accuracy % is more than 93% and poorly at class 0 with 83% accuracy.



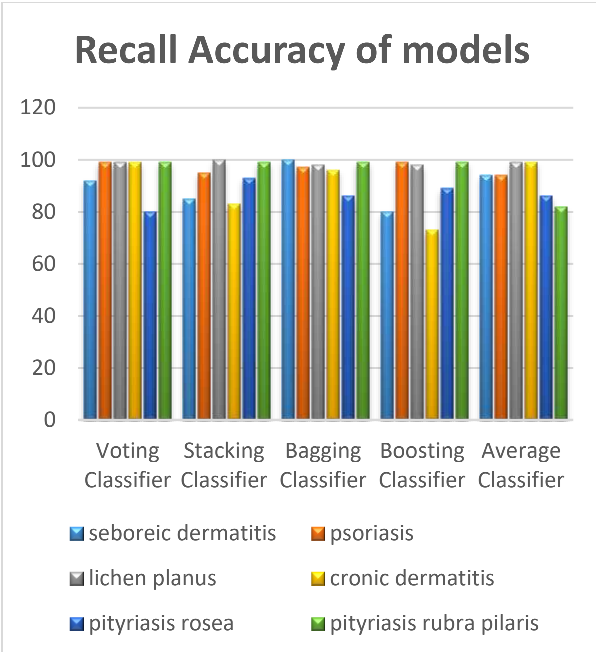
Graph 2 Comparison of precision Accuracy

The above graph displays the values of each class's precision and demonstrates that the Bagging Classifier calculates the highest precision while the other Classifiers perform relatively less than the bagging classifier. Class 3 and 5 have good performance across all classifiers, while Class 0 has poor performance across all classifiers.

B. Recall comparison

Recall for every classifier is provided with each class. Various classes have the highest recall with 99% accuracy, whereas class 0 and 4 has the lowest recall with 80% accuracy. The stacking Classifier provides class 2 with the highest recall and accuracy of 100% while providing class 3 with the lowest recall accuracy of 83%. Classes 2 and 3 have the highest accuracy provided by the average classifier, with a recall score of 99%, whereas class 5 has a poor recall accuracy score of 33%.

With an accuracy of 99%, Boosting Classifier has the highest recall for classes 1, and 5 and the lowest accuracy for class 3 with 73%. Bagging Classifier fared poorly in class 4 with an accuracy of 86% while doing well in the rest of the other classes with a recall value of >90% accuracy.



Graph 3 Comparison of Recall Accuracy

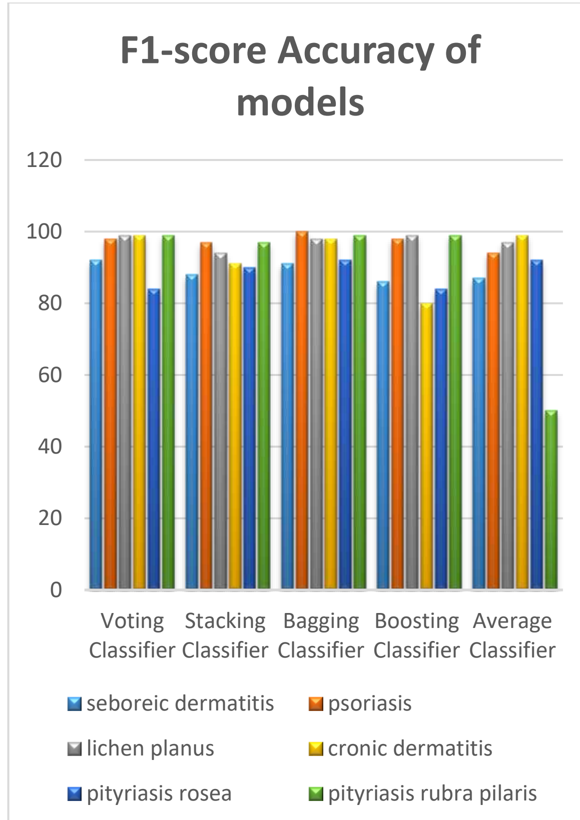
The graph displays the recall values for each class and demonstrates that the Voting Classifier, with an accuracy of 93%, 96%, and 99%, calculates the highest recall values. The Stacking Classifier calculates the lowest recall values for each class. In every classifier, class C2 has a high recall, while class C0 has a low recall.

C. F-1 Scores Comparison of model

Every class is given the F-1 Score for each classifier. Classes 2, 3, and 5 have the greatest F-1 Score of the Voting Classifier with an accuracy of 99%, and class 0 has the lowest F-1 Score with an accuracy of 92%. Stacking Classifier performs poorly on class 0 with an accuracy of 88% and gives classes 5 & 1 the highest F-1 Score with 97% accuracy. Class 3 receives the highest accuracy from the average classifier, with an F-1 score of 99%, while class 5 receives a low F-1 score, with an accuracy score of 50%.

Classes 2 and 5 get the best F-1 Score from Boosting Classifier with an accuracy of 99%, while class 4 receives the lowest score with an accuracy of 84%. Bagging Classifier fared poorly in class 0 with an accuracy of 91% while doing well in class 5 with a precision of 99%.

The graph displays the F-1 Score values for each class and demonstrates that the Bagging Classifier calculates the F-1 Score with an accuracy of>then 90% in each class, respectively. The Stacking Classifier and average classifier calculate relatively less than other classifiers' F-1 Scores for each class. In every classifier, classes 1 and 3 have high F-1 scores, while class 0 has a low F-1 score.



Graph 4 F-1 score accuracies

D. Analysis of results

The following conclusions are drawn from our results after evaluating the confusion matrix and all classifier results:

- When compared to simple models, ensemble machine learning models produce more accurate findings.
- The annotated dataset increases each class's precision, recall, and F-1 score.
- Voting and bagging classifiers function, but are more effective than the other ensemble classifiers.
- In terms of precision, classes 3 and 5 award the greatest score, while class 0 scores the lowest.
- In the Recall, class 2 scored highest while class 4 scored lowest.
- F-1 Score, with classes 3 and 5 having the highest and class 0 having the lowest.

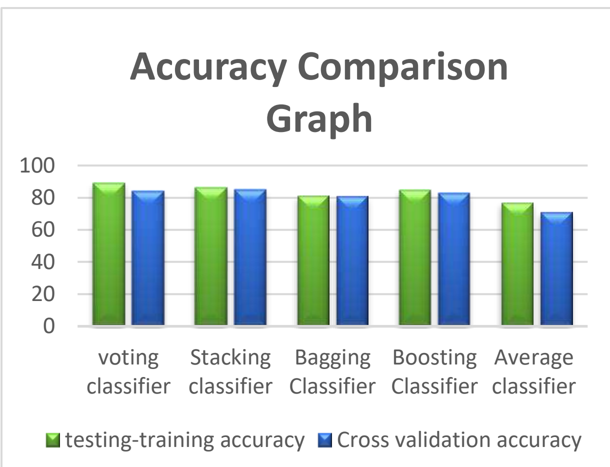
Comparative Analysis of Clinical Feature

An additional method for enhancing the accuracy of erythemato-squamous illnesses is using only clinical features. To help doctors distinguish the classes of ESD, show the value of ensemble Meta classification algorithms that only accept clinical features as input, and introduce these Machine Learning algorithms as a first step in creating an expert system for the classification of erythemato-squamous. Only the clinical features from the dataset for erythemato-squamous disorders acquired from the UCI (University of California, Irvine) machine repository are utilized as input for five ensemble classification algorithms.

E. Accuracy Comparison

Accuracy data for five Voting Classifiers, Stacking Classifier, Average, Bagging classifier, and Boosting ensemble models are presented in the form of a graph. Each model includes two distinct levels of accuracy: one for cross-validation and the other for training-testing splits. After the comparison of all classifiers, Voting Classifier provides the best Accuracy rate in both cross-validation and test and training accuracy.

The model Voting classifier has the highest accuracy on testing data (89.1%), while the model boosting Classifier has the lowest accuracy (77%). Similarly, the accuracy of cross-validation of the stacking voting classifier shows the highest percentage of 85 and Boosting is the least accurate model with 71%.



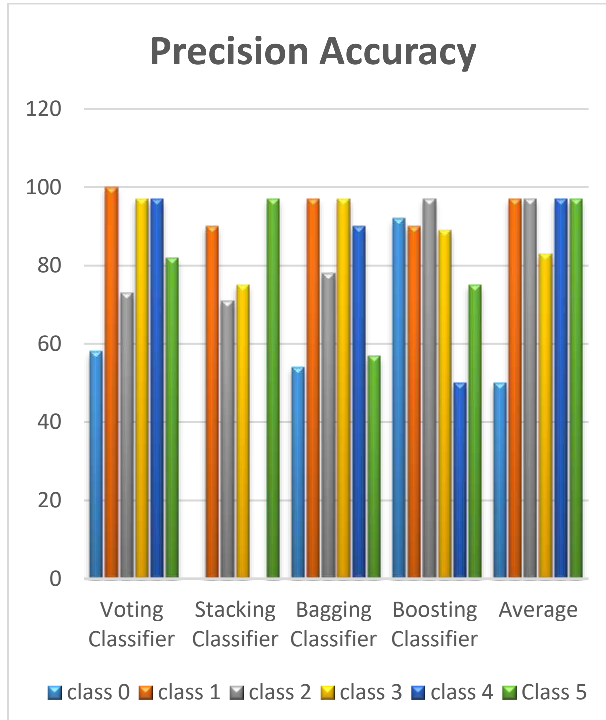
Graph 5 Accuracy Comparison

In the graph above, two metrics cross-validation and training testing are used to compare the accuracy of five models. The Boosting Classifier, Stacking Classifier, Voting classifier, Bagging, and average classifier were used to achieve the results.

F. Precision Comparison of Models

A precision score of all classifier models is presented. The voting classifier performs well in classes 1,3 and 4 with an accuracy of 97% and shows lower performance in class 0 with 58% accuracy. The accuracy of the predictions made by stacking classifiers is 97% for class 5 and 90%, 71%, and 75%, in other classes and performs the lowest prediction rate for class 0 with the percentage of 0; Class 1, 2, 4, and 5 has the highest accuracy according to the Average Classifier, with a precision score of 97%, whereas class 0 has a low precision score of 50%.

The Accuracy of the Boosting Classifier's predictions is highest in classes 2 (99%), and lowest in classes 0, and 4 (50%). Bagging Classifier performs admirably in classes 1 and, 3 with 97% accuracy and poorly at class 0 with 54% accuracy.



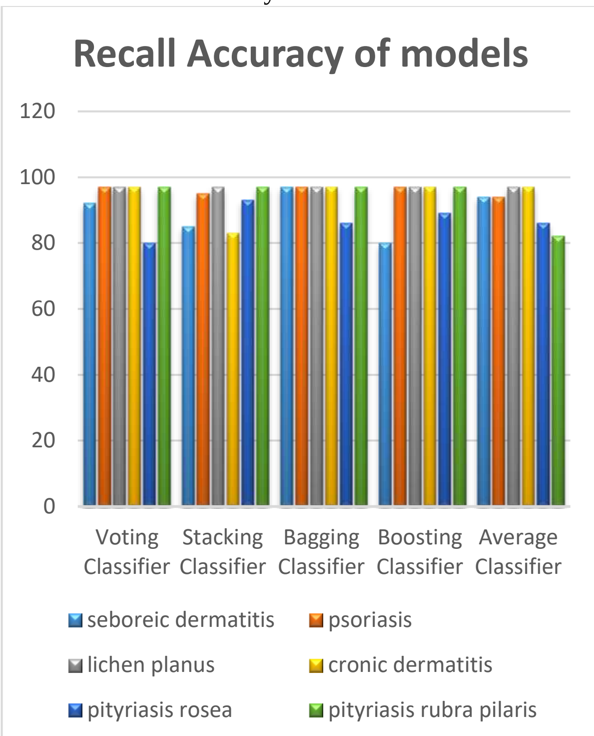
Graph 6 Comparison of Precision Accuracy

The graph above displays the values of each class's precision and demonstrates that the Bagging Classifier calculates the highest precision while the other Classifiers perform relatively less than the bagging classifier. Classes 2 and 5 have good performance across all classifiers, while Classes 0, and 4 have poor performance across all classifiers.

G. Recall comparison

Recall for every classifier is provided with each class. Various classes have the highest recall with 97% accuracy, whereas class 0 and 4 has the lowest recall with 80% accuracy. The stacking Classifier provides class 2 with the highest recall and accuracy of 97% while providing class 3 with the lowest recall and accuracy of 83%. Classes 2 and 3 have the highest accuracy provided by the average classifier, with a recall score of 97%, whereas class 5 has a poor recall and an accuracy score of 82%.

With an accuracy of 97%, Boosting Classifier has the highest recall for classes 2, and 5 and the lowest accuracy for class 3 with 53%. Bagging Classifier fared poorly in class 5 with an accuracy of 44% while doing well in class 1 with a recall value of 97% accuracy.



Graph 7 Comparison of Recall Accuracy

The graph displays the recall values for each class and demonstrates that the Voting Classifier, with an accuracy of 78, 90, 92, 97, and 64 calculates the highest recall values. The Stacking Classifier calculates the lowest recall values for each class. In every classifier, class C2 has a high recall, while class C0 has a low recall.

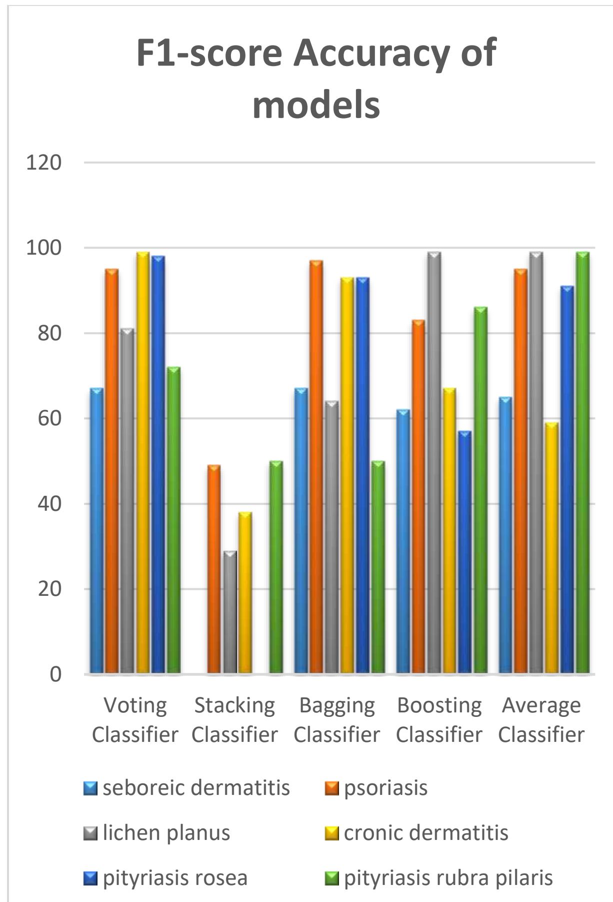
H. F-1 Scores Comparison of model

Every class is given the F-1 Score for each classifier. Classes 3 have the greatest F-1 Score of the Voting Classifier with an accuracy of 99%, and class 0 has the lowest F-1 Score with an accuracy of 67%. Stacking Classifier performs poorly on classes 0 and 4 with 0% and gives class 5 the highest F-1 Score with 50% accuracy. Class 2 receives the highest accuracy from the average classifier, with an F-1 score of 99%, while class 2 and 5 receives a low F-1 score, with an accuracy score of 65% in class 0.

Classes 2 get the best F-1 Score from Boosting Classifier with an accuracy of 99%, while class 0 class 4 receives the lowest score with an accuracy of 57%. Stacking Classifier fared poorly in class 0 and class 4 with an accuracy of 0% while class 1 has the highest accuracy with a precision of 49%. Whereas the average classifier performs best from all other classifiers with 99% in class 2 and class 5 and the lowest accuracy of the average classifier is in class 3 with a percentage of 59%.

The graph displays the F-1 Score values for each class and demonstrates that the Bagging Classifier calculates the F-1 Score with an accuracy of 67%, 97%, 64%, 93%, 93%, and 50%, respectively. The Stacking Classifier and average classifier calculate relatively poorly than other classifiers' F-1 Score for each class. In every classifier, classes 3 and 5 have high F-1 scores, while class 0 has a low F-1 score.

The voting Classifier calculates the F-1 Score with an accuracy of 67%, 95%, 82%, 99%, 98%, and 72%, respectively. Overall Class 2 shows the highest accuracy in boosting and average classifier.



Graph 8 F-1 score accuracies

I. Analysis of results

The following conclusions are drawn from our results after evaluating the confusion matrix and all classifier results:

- When compared to simple models, ensemble machine learning models produce more accurate findings.
- The annotated dataset increases each class's precision, recall, and F-1 score.

- Here we compared the accuracy of extracted features of the dataset.
- Voting, and average classifiers function, but are more effective than the other ensemble classifiers.
- In terms of precision, classes 1 and 3 award the greatest score, while class 0 scores the lowest.
- In the Recall, classes 2, and 5 scored highest while class 0 scored lowest.
- F-1 Score, with classes 2 having the highest and class 0 having the lowest.

The results of the proposed framework are compared with various existing studies using the same dermatology dataset for the classification of erythemato-squamous disease. The results are evaluated in terms of accuracy and are compared with existing research.

Table 4 Comparison of state-of-the-art techniques

References	Model	Dataset	Accuracy
[15]	Linear Vector Quantization	UCI machine repository dermatology dataset	The accuracy of 40% for psoriasis, 45% for seborrheic dermatitis, 30% for lichen planus, 45% for pityriasis rosea, 50% for chronic dermatitis, and 90% for pityriasis rubra pilaris, classification investigations were conducted.
[9]	ReliefF & k-nearest neighbors	UCI machine repository dermatology dataset	Traditional KNN results in an accuracy of 85.13% and the accuracy of hybrid KNN is 94.59%.
[10]	SVM, logic boost, MLP Begging, Bayes net and naïve bayes	UCI machine repository dermatology dataset	The Bayesian network provides 83.6% accuracy and the logit boost gives 84.5% accuracy is the best accuracy to predict this disease based on clinical features.
[17]	Multi-layer feed-forward and ANN	UCI-machine repository dermatology dataset	ANN algorithm produces optimum results
[21]	Fuzzy Logic	UCI-machine repository dermatology dataset	This fuzzy expert system reported a 90.2% of accuracy rate.
Our Proposed Methodology	Ensemble Meta Technique	UCI Machine Repository dermatology dataset	97.8 % is the best accuracy by the proposed methodology

Conclusion and Future Work

The health of the skin, hair, nails, and mucous membranes are concern in the field of dermatology. According to dermatology data from the UCI Repository, many diseases share many clinical and histological characteristics. The identification of various ESDs is a difficult task. The proposed study focuses on the medical diagnosis of erythemato-squamous disorders such as “Lichen planus, Pityriasis rosea, Pityriasis rubra pilaris, psoriasis, seboreic dermatitis, and chronic dermatitis” using solely clinical characteristics.

The performance of 5 ensemble-based predicting techniques based on the Voting classifier, Average classifier, gradient boost, and bagging classifier algorithms is measured in this research. Using different parameters from the dermatological dataset, these models are utilized to forecast erythemato-squamous illness. There are 366 data samples gathered. The study's initial phase involved preprocessing the data before using it as an input for base models. Next, classification was done, and the effectiveness of the technique was evaluated in terms of precision, recall, accuracy, and f-1 score.

The model Voting classifier and the Bagging classifier have the highest accuracy on training and testing data (97.8%) and 97%, while the model Average Classifier has the lowest accuracy (93%). Similarly, the accuracy of cross-validation of the model voting classifier shows the highest percentage of 98 and Bagging is the least accurate model with 94%. In terms of precision, classes 3 and 5 award the greatest score, while class 0 scores the lowest. In the Recall, class 2 scored highest while class 4 scored lowest. F-1 Score, with classes 3 and 5 having the highest and class 0 having the lowest.

Furthermore, we also concluded that the case of clinical feature yielded classification accuracies that were relatively less than the original dataset. The highest classification accuracy was supplied by the whole dataset, which was comparable to earlier reported results.

A. Future Work

It is noted that the bulk of the features was unnecessary. The future researcher should examine the dataset more thoroughly about the assignment of feature values to further investigate this result. Future work will involve testing and refining the algorithm to improve accuracy for a range of heterogeneous datasets where disease differential diagnosis is difficult because of overlapping variables. The implementation of the suggested classifier in hospitals, starting with the classification for the differential diagnosis of erythemato-squamous disorders, where unseen datasets would be used, would be our next step. The proposed methodology can also be used to detect other diseases.

In the future, an Android application can be developed to assist in the healthcare sector. This application will allow the user to enter the required details that are required to detect the skin disease. This application will also prescribe medicine according to the detected skin disease.

CREDIT AUTHOR STATEMENT

Azka Ahmed: Conceptualization, Methodology, Software, Data curation, Writing- Original draft preparation. **Hafsa Ahmad**: Visualization, Validation, Investigation. ***Kamran Abid****:* Supervision. **Mohsin Khurshid**: Software, Writing- Reviewing and Editing.

COMPLIANCE WITH ETHICAL STANDARDS

It is declare that all authors don’t have any conflict of interest. Furthermore, informed consent was obtained from all individual participants included in the study.

REFERENCES

[1] Z. Wu *et al.*, "Studies on different CNN algorithms for face skin disease classification based on clinical images," vol. 7, pp. 66505-66511, 2019.

[2] H. Malik, T. Anees, M. Din, and A. Naeem, "CDC_Net: multi-classification convolutional neural network model for detection of COVID-19, pneumothorax, pneumonia, lung Cancer, and tuberculosis using chest X-rays," *Multimedia Tools and Applications*, pp. 1-26, 2022.

[3] I. Abunadi and E. M. J. E. Senan, "Deep learning and machine learning techniques of diagnosis dermoscopy images for early detection of skin diseases," vol. 10, no. 24, p. 3158, 2021.

[4] E. Proksch, J. M. Brandner, and J. M. J. E. d. Jensen, "The skin: an indispensable barrier," vol. 17, no. 12, pp. 1063-1072, 2008.

[5] K. Jyothilakshmi and J. Jeeva, "Detection of malignant skin diseases based on the lesion segmentation," in *2014 International Conference on Communication and Signal Processing*, 2014, pp. 382-386.

[6] S. K. Singh, A. Sinha, and S. Yadav, "Performance Analysis of Machine Learning Algorithms for Erythemato-Squamous Diseases Classification," in *2022 IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*, 2022: IEEE, pp. 1-6.

[7] S. O. Olatunji and H. J. I. J. o. S. C. Arif, "Identification of erythemato-squamous skin diseases using extreme learning machine and artificial neural network," vol. 4, no. 1, pp. 627-32, 2013.

[8] "Erythema," ed: drugwatch, 2022.

[9] J. B. Idoko, M. Arslan, and R. Abiyev, "Fuzzy neural system application to differential diagnosis of erythemato-squamous diseases," *Cyprus J Med Sci*, vol. 3, no. 2, pp. 90-97, 2018.

[10] B. Lahijanian, F. V. Farahani, and M. F. Zarandi, "A new multiple classifier system for diagnosis of erythemato-squamous diseases based on rough set feature selection," in *2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2016: IEEE, pp. 2309-2316.

[11] S. Putatunda, "A hybrid deep learning approach for diagnosis of the erythemato-squamous disease," in *2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, 2020: IEEE, pp. 1-6.

[12] K. V. Swamy and B. Divya, "Skin Disease Classification using Machine Learning Algorithms," in *2021 2nd International Conference on Communication, Computing and Industry 4.0 (C2I4)*, 2021: IEEE, pp. 1-5.

[13] G. N. Raj, U. N. Raju, M. K. Venkat, and V. K. Shrivastava, "Comparative performance analysis of different classifiers on diagnosis of erythmato-

squamous diseases," in *2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, 2017: IEEE, pp. 1-6.

[14] J. Wu and C. Hicks, "Breast cancer type classification using machine learning," *Journal of personalized medicine*, vol. 11, no. 2, p. 61, 2021.

[15] H. Abbad Ur Rehman, C.-Y. Lin, Z. Mushtaq, and S.-F. Su, "Performance analysis of machine learning algorithms for thyroid disease," *Arabian Journal for Science and Engineering,* vol. 46, no. 10, pp. 9437-9449, 2021.

[16] A. Naeem, T. Anees, M. Fiza, R. A. Naqvi, and S.-W. Lee, "SCDNet: A Deep Learning-Based Framework for the Multiclassification of Skin Cancer Using Dermoscopy Images," *Sensors,* vol. 22, no. 15, p. 5652, 2022.

[17] S. Uddin, A. Khan, M. E. Hossain, M. A. J. B. m. i. Moni, and d. making, "Comparing different supervised machine learning algorithms for disease prediction," vol. 19, no. 1, pp. 1-16, 2019.

[18] R. A. Khan, Y. Luo, and F.-X. Wu, "Machine learning based liver disease diagnosis: A systematic review," *Neurocomputing,* vol. 468, pp. 492-509, 2022.

[19] A. Naeem, T. Anees, R. A. Naqvi, and W.-K. Loh, "A Comprehensive Analysis of Recent Deep and Federated-Learning-Based Methodologies for Brain Tumor Diagnosis," *Journal of Personalized Medicine,* vol. 12, no. 2, p. 275, 2022.

[20] E. Tuba, I. Ribic, R. Capor-Hrosik, and M. Tuba, "Support vector machine optimized by elephant herding algorithm for erythemato-squamous diseases detection," *Procedia computer science,* vol. 122, pp. 916-923, 2017.

[21] A. Sanchez *et al.*, "Digitate papulosquamous eruption associated with severe acute respiratory syndrome coronavirus 2 infection," *JAMA dermatology,* vol. 156, no. 7, pp. 819-820, 2020.

[22] E. Farinazzo, C. Dianzani, I. Zalaudek, C. Conforti, S. Grabbe, and M. Goldust, "Synthesis of the data on COVID-19 skin manifestations: underlying mechanisms and potential outcomes," *Clinical, Cosmetic and Investigational Dermatology,* vol. 14, p. 991, 2021.

[23] S. Veraldi, M. Romagnuolo, and V. Benzecry, "Pityriasis rosea-like eruption revealing COVID-19," *The Australasian Journal of Dermatology,* vol. 62, no. 2, p. e333, 2021.

[24] M. N. BOZOK and A. ÇALHAN, "Diagnosis of Erythemato-Squamous Skin Diseases with Machine Learning Algorithms," 2022.

[25] A. S. Alotaibi, "Hybrid model based on ReliefF algorithm and K-nearest neighbor for erythemato-squamous diseases forecasting," *Arabian Journal for Science and Engineering,* vol. 47, no. 2, pp. 1299-1307, 2022.

[26] B. Geetha, "Diagnosis of Erythemato Techniques," *Australian Journal of Basic and Applied Sciences,* vol. 11, no. 5, pp. 45-51, 2017.

[27] S. O. Olatunji and H. Arif, "Identification of erythemato-squamous skin diseases using support vector machines and extreme learning machines: a comparative study towards effective diagnosis," *Transactions on Machine Learning and Artificial Intelligence,* vol. 2, no. 6, p. 124, 2015.

[28] H. Haryanto *et al.*, "The Erythemato-Squamous Dermatology Diseases Severity Determination using Self-Organizing Map," *IPTEK Journal of Proceedings Series,* vol. 1, no. 1, 2015.

[29] K. Maghooli, M. Langarizadeh, L. Shahmoradi, M. Habibi-Koolaee, M. Jebraeily, and H. Bouraghi, "Differential diagnosis of erythmato-squamous diseases using classification and regression tree," *Acta Informatica Medica,* vol. 24, no. 5, p. 338, 2016.

[30] L. Anifah, "Decision Support System Erythemato-Squamous Diseases Classification Diagnosis Using Linear Vector Quantization Based Clinical Atributes," in *2021 3rd International Conference on Research and Academic Community Services (ICRACOS)*, 2021: IEEE, pp. 254-258.

[31] J. Rajeshwari and M. Sughasiny, "Modified Filter Based Feature Selection Technique for Dermatology Dataset Using Beetle Swarm Optimization," *EAI Endorsed Transactions on Scalable Information Systems,* pp. e78-e78, 2022.

[32] S. Patel and A. Patel, "An Empirical Study of Applying Artificial Neural Network for Classification of Dermatology Disease," *Indian Journal of Science and Technology,* vol. 10, no. 17, 2017.

[33] B. I. Al-Kahlout, M. M. Naeem, and M. J. Shepherd, "ANN for the Classification of Eryhemato-Squamous Disease," 2021.

[34] G. Sahebi, P. Movahedi, M. Ebrahimi, T. Pahikkala, J. Plosila, and H. Tenhunen, "GeFeS: A generalized wrapper feature selection approach for optimizing classification performance," *Computers in biology and medicine,* vol. 125, p. 103974, 2020.

[35] I. J. Bush, M. Arslan, and R. Abiyev, "Intensive investigation in differential diagnosis of erythemato-squamous diseases," in *International Conference on Theory and Applications of Fuzzy Systems and Soft Computing*, 2018: Springer, pp. 146-153.

[36] M. A. A. Raza, M. S. Liaqat, and M. Shoaib, "A Fuzzy Expert System Design for Diagnosis of Skin Diseases," in *2019 2nd International Conference on Advancements in Computational Sciences (ICACS)*, 2019: IEEE, pp. 1-7.

[37] S. Sivasankari, D. Punnoose, and D. Krishnamoorthy, "A comparative study on the performance of rule engines in automated ontology learning: a case study with erythemato-squamous disease (ESD),"

International Journal of Intelligent Unmanned Systems, vol. 8, no. 4, pp. 267-280, 2020.

[38] A. Basu, S. S. Roy, and A. Abraham, "A novel diagnostic approach based on support vector machine with linear kernel for classifying the erythemato-squamous disease," in *2015 International Conference on Computing Communication Control and Automation*, 2015: IEEE, pp. 343-347.

[39] S. M. Banu and G. Toacşe, "A mobile/desktop medical application for automatic differential diagnosis of psoriasis lesions," in *2013 Second International Conference on E-Learning and E-Technologies in Education (ICEEE)*, 2013: IEEE, pp. 186-191.

[40] M. Brito, B. Stevens, and Z. Yiu, "Can artificial intelligence be used for accurate remote scoring of the Psoriasis Area and Severity Index in adult patients with plaque psoriasis?," *British Journal of Dermatology*, vol. 185, no. 6, pp. 1262-1264, 2021.

[41] A. K. Verma, S. Pal, and S. Kumar, "Classification of skin disease using ensemble data mining techniques," *Asian Pacific journal of cancer prevention: APJCP*, vol. 20, no. 6, p. 1887, 2019.