Classification of Erythemato-Squamous Diseases of Skin Using Various Conventional Machine Learning Methods

Bhavana Kaushik[1], Ankur Vijayvargiya[2], Jayant Uppal[3], Ankit Gupta[4]
[1]School of Computer Science, University of Petroleum and Energy Studies, Dehradun, Uttarakhand, India
[2]Accenture, Noida, Uttar Pradesh, India
[3]PeopleStrong Noida, Uttar Pradesh, India
[4]Samsung R&D Noida, Uttar Pradesh, India

ABSTRACT
In the recent era, Machine Learning and Artificial Intelligence have come to a very great development point as we can use ML algorithms to predict the type of Erythemato-Squamous (Skin) diseases of the skin. In Dermatology, differential diagnosis of skin diseases is quite challenging in real life because most skin diseases share many histopathological features. And in this work, Psoriasis, Lichen Planus, Seborrheic Dermatitis, Chronic Dermatitis, Pityriasis Rosea, and Pityriasis Rubra Pilaris are among the skin illnesses for which eight different algorithm analytical comparison is done. Moreover, each classifier algorithm is discussed in detail with its pros and cons. The machine learning algorithms like Support Vector Machine, Decision tree, RandomForest, KNN, Naïve Bayes, Gradient Boosting, XGBoost, and Multilayer Perception have been proven to be successful in preserving state information through exact segmentation/classification. Random forest, Gradient Boosting, and XGBoost outperform all other methods and give an accuracy of 100% on the given ESD dataset. While Support Vector Machine gives the least accuracy of 72.97%. The paper alsodiscusses the difficulties connected with skin disease segmentation or categorization. Furthermore, the study proposes future potential directions that include real-time analysis.
Keywords: Erythemato-Squamous Diseases, Machine Learning, Classification, Skin Diseases, Comparative Analysis

I. INTRODUCTION

In recent times we observed that many people are suffering from skin diseases or skin cancer which can be curable at an early stage but now they are not curable thus after watching the continues advancement and development in technology and especially in AI & ML we decided to combine them and take the help from various machine learning algorithms to predict the skin diseases at the earliest stage so that patient can be cured within time and it can also help all medical field and especially in dermatology to predict the diseases atthe earliest stage. Many times, doctor is not able to get the type of skin disease in the earlier stage because at the beginning stage all types of skin diseases show the same symptoms and also share the same histopathological features so identifying in the earliest stage is also a very challenging task for the doctors thus this time is wasted and can cause the disease to grow very quickly and can cause to death and cancer. This wasted time is very crucial for the patient and in this time if our algorithm can make the accurate and right decision then we can save the life of the patient. The provision of high-quality services at reasonable prices is one of the most difficult difficulties that health-care institutions (hospitals, medical facilities) confront today. Quality service include appropriately diagnosing patients and giving effective therapies [1]. Most hospitals currently employ some kind of hospital information system to manage their healthcare or patient data. These systems generally produce massive volumes of data in the form of statistics, text, charts, and pictures [2]. Regrettably, these data are rarely used to guide clinical decisions. The outcomes can be obtained by utilising relevant computer-based information and/or decision support tools. This presents a crucial question: "How can we turn data into valuable information that allows clinicians to make informed therapeutic decisions?" This is the primary incentive for this study. Erythemato- squamous diseases (ESDs) arevery prevalent skin conditions. “Psoriasis, seboreic dermatitis, lichen planus, pityriasis rosea, chronic dermatitis, and pityriasis rubra pilaris are the six distinct types.” They all have the medical indications of erythema and scaling, by actual slight exclusion. [3]. Psoriasis - Psoriasis is assumed to be a problem with the immune system. Scales and itchy, dry spots emerge as skin cells stack up and create scales. Infections, stress, and the common cold are all triggers. A rash on the skin is the most common symptom, however, the rashcan also affect the nails or joints. The treatment's purpose is to remove scales and reduce the pace at which skin cells multiply. Topical ointments, light therapy, and medications can provide relief. [4]. Seborrheic dermatitis - This skin ailment is characterised by flaking patches and red skin, mainly on the scalp. It can also appear on oily body parts like the face, upper chest, and back. Seborrheic dermatitis can create obstinate dandruff in addition to scaly patches and red skin. Self-care and medicinal shampoos, creams, and lotions are used in the treatment. Treatments may need to be repeated [5]. Lichen planus -An inflammatory skin and mucous membrane disease. When the immune system mistakenly targets skin or mucous membrane cells, lichen planus develops. Lichen planus shows on the skin as reddish, itchy pimples with a flat top. It creates lacy, white patches on mucous membranes, such as the mouth, and sometimes severe ulcers. Lichen planus is frequently self-resolving. Topical treatments and antihistamines may help if symptoms are bothersome [6]. Pityriasis Rosea - A hasty that twitches as a huge spot on the abdomen, chest, or back and then spreads out into a pattern of smaller lesions. The cause of pityriasis rosea is unknown, but it is thought to be caused by a viral infection. The illness creates a rash on the torso, upper legs, and upper arms that is mildly irritating. Pityriasis rosea is frequently self-resolving. Antihistamines, steroid cream, and, in rare situations, antiviral medicines can all assist [7]. Chronic dermatitis - Dermatitis is a word used to define a group of itchy, inflammatory skin disorders marked by epidermal abnormalities. Dermatitis affects one out of every five people at some time in their lives. It can be caused by a range of factors and has a variety of patterns. The phrases dermatitis and eczema are frequently confused.

Eczematous dermatitis is a term that is sometimes used. Dermatitis can be acute, chronic, or a combination of the two. Acute eczema (also known as dermatitis) is a fast-progressing red rash that can blister and swell. Eczema (or dermatitis) is a long- term irritating skin condition. It's usually darker than the rest of the skin, thicker, and scraped a lot [8]. Pityriasis Rubra Pilaris (PRP) – It is a term used to describe a collection of rare skin illnesses characterized by scaling patches that are reddish- orange in colour and have well-defined edges. They can cover the entire body or convinced parts like the prods and laps, palms, and soles. Islands of sparing are patches of uncomplicated skin, especially on the stem and limbs, which are frequently seen. The palms and soles are commonly affected, becoming swollen and yellowish in appearance (palmoplantar keratoderma).PRP is frequently misdiagnosed as psoriasis or another skin disorder [9].

Infection with one of these skin diseases results into the worsening of skin cells (squamous) and origins redness of the skin (erythema). Dermatologists often inspect patients clinically as well as on the basis of histopathological components [10]. Clinical inspections include examination of color, existence of pimples, size, position of, and other symptoms. All the above inspections lead to 12 clinical and 22 histologicalvariables for each person/patient. The investigation of these parameters may lead to indefinite and indistinct consequences, as they may intersect, especially in the early phases of ESD. So, there is a need to identify the appropriate classification technique to solve this problem and give a better result/prediction. This paper reviews all conventional techniques present to perform the differential diagnosis of ESD and also to identify the challenges existing with approaches. The contributions of this study are as follows:

(i) Analytical comparison of all relevant conventional machine learning techniques for differential diagnosis of ESD is done.
(ii) Also, the challenges and issues associated with each technique is identified.
(iii) The study discusses the future potential directions that include real- time analysis.
(iv) Discuss the difficulties connected with skin disease segmentation or categorization.

The structure of this paper is as follows: The description of material and methods used in this comparative analysis is given in Section 2. Section 3 contains the result and discussions of the analysis. Prominent challenges and future scope are described in the Section 4 and lastly Section 5 concludes the study.

II. MATERIALS AND METHODS

A. Materials

We used a standardized dermatology data set from the “University of California, School of Information and Computer Science's machine learning repository, or UCI. It has 34 properties, 12 of which are clinical and 22 of which are histological”. Age and family history are continuous characteristics in the data set, with values ranging from 0-1. “Every additional clinical and histological feature was assigned a degree from 0 to 3, with 0 indicating that the feature was not present, 3 indicating the maximum amount possible, and 1, 2 indicating relative intermediate values”. Naive Bayes, Random Forest, Support Vector Machines, XGBoost, Multi- layered perceptron, K-nearest neighbors, Decision tree, Gradient boosting DT are among the ML Classification Algorithms investigated in this paper. Figure 1 shows the extract of the dermatology data set from the University of California, School of Information and Computer Science's machine learning repository,or UCI [10].

	erythema	scaling	definite borders	itching	...	perifollicular parakeratosis	inflammatory monoluclear infiltrate	band-like infiltrate	age	class label
0	2	2	0	3	...	0	1	0	55.0	2
1	3	3	3	2	...	0	1	0	8.0	1
2	2	1	2	3	...	0	2	3	26.0	3
3	2	2	2	0	...	0	3	0	40.0	1
4	2	3	2	2	...	0	2	3	45.0	3
...	...	...	...	...	...	...	...	...	...	...
361	2	1	1	0	...	0	2	0	25.0	4
362	3	2	1	0	...	0	2	0	36.0	4
363	3	2	2	2	...	0	2	3	28.0	3
364	2	1	3	1	...	0	2	3	50.0	3
365	3	2	2	0	...	0	3	0	35.0	1

366 rows × 35 columns

Figure 1: Visualisation of the dataset.

Table 1: Six classes of ESD

Keys	Values (Class Labels)
1	Psoriasis
2	Seborrheic Dermatitis
3	Lichen Planus
4	Pityriasis Rosea
5	Chronic Dermatitis
6	Pityriasis Rubra Pilaris

B. Methods

a) Support Vector Machine (SVM)

The SVM's goal is to discover the hyper plane in n- dimensional space, where n is the number of variables. The hyper-plane is chosen such that the distance between the nearest data points and support vectors of the two separate modules is as little as possible. We may think of the hyper plane as a line in two dimensions and a plane in three dimensions. Hyper plane separates the data points of two different classes [11]. Numerous of the prevailing (non)convex soft-margin losses can be observed as one of the substitutes of the L0/1 soft- margin loss. SVM have gained huge consideration for the last two decades due to its wide-ranging usage, so many researchers have established optimization procedures to solve SVM with various soft-margin losses [12]. For the prediction corresponding to a new input can be obtained using Equation 1:

$$f(x) = B0 + sum(ai * (x, xi)) \quad (1)$$

Where f(x) is used to calculate the inner dot product which is the sum of the multiplication of each pair of the input values i.e., x as the new input and xi as each of the support vectors present in the training set. B0 and ai are the coefficients evaluated from the training data [12]. Figure 2 shows the visuals of the Support Vector Machines (SVMs) Model with all necessary features like absolute hyperplane with maximum margin, hyperplane with positive and negative trends and the nearest data points as the support vectors.

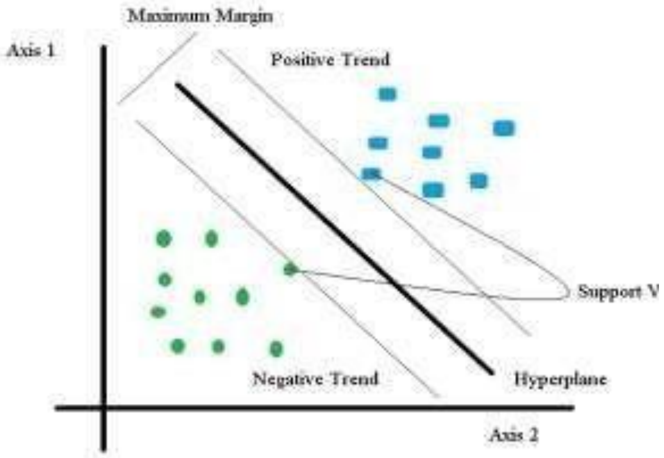


Figure 2: Visualization of Support Vector Machines (SVMs) Model [13]

b) Random Forest

It is, as the name implies, a group of various decision trees, each of which predicts some class, and the class having the most votes is accepted as the predicted class. These decision trees produce distinct outcomes relatively. This concept is highly effective in the reduction of prediction errors if predicted through a single decision tree. In this approach, an individual tree may be in the wrong direction but the common direction could be in the right direction [14]. The Figure 3 demonstrates the formation of the multiple decision trees with differently specified branching conditions to form a random forest and each decision tree having its own prediction.

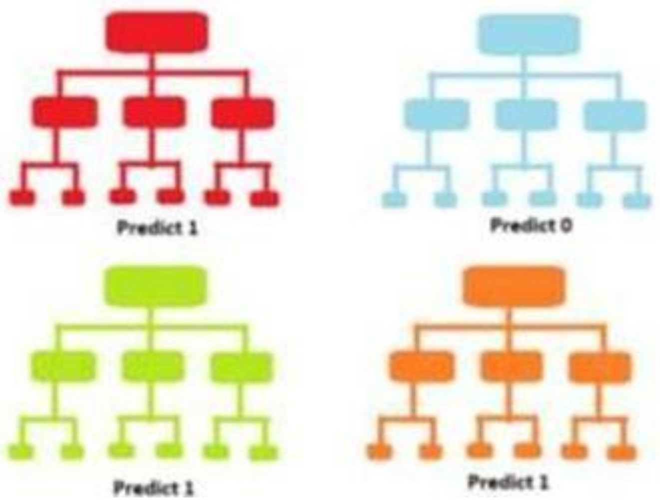


Figure 3: Visualization of a Random Forest Model [15]

c) Naive Bayes

This classifier is based on the Bayes theorem and approaches the probabilistic strategy in classification through prediction. Equation 2 states the approach of Bayes theorem:

$P(A|B) = P(B|A)\, P(A) / P(B)$ (2)

Where we discover the chance of happening of A assuming that B had already occurred. In this concept A is considered as the hypothesis and B is considered as the evidence. This approach is best when the features are not affected by each other [16]. The Figure 4 shows the visuals of Naive Bayes Classifier Algorithm.

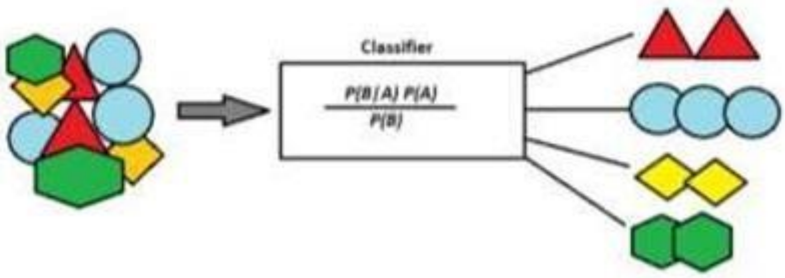


Figure 4: Visualization of Naive Bayes Classifier Model [17]

d) Decision tree

As the name suggests, we can find an analogy between a tree and a result tree. A result tree is similar to a tree by having split conditions as a node, directing edges as branches, and decisions as leaves. The formation of a tree involves feature decision and branching conditions and holds the decision by preventing further branching. This approach follows the greedy concept by splitting the branches with lower prediction cost i.e., the class with the maximum data points should be classified initially at 0 level/root node [18]. The Figure 5 demonstrates the formation of a decision tree with root node, sub-tree having branching from decision nodes and predicted outcomes as Terminal nodes.

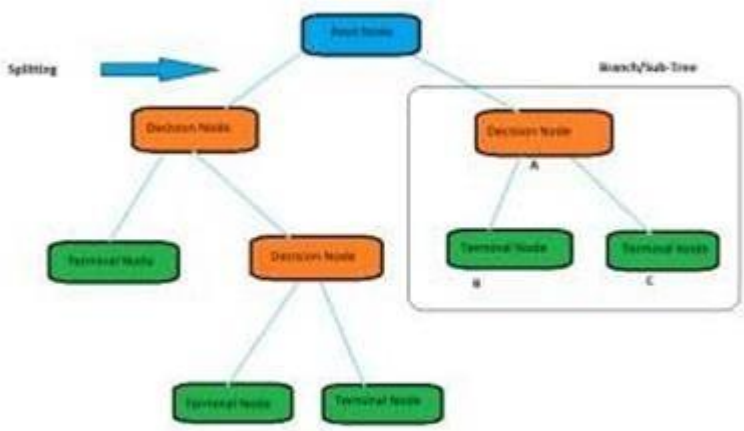


Figure 5: Visualization of Decision Tree Model [19]

e) K-nearest neighbors

As the name suggests, this algorithm finds the separate clusters of data points present in proximity i.e., near to each other based on the distance between the two data points. In this classification approach, the K refers to the number of neighbors as the class labels and the mode of k labels is considered as the predicted outcome. The efficiency of this algorithm decreases with an increase in the number of predictor variables [20]. For a new input having real values, the distance is most likely to be measured through Euclidean distance given by Equation 3:

$EuclideanDistance(x, xi) =$
$sqrt(\ sum(\ (xj - xij)\hat{}2\ )\ )$ (3)

Where x is the new input and xi is the existing point covering all the j input attributes [21]. The Figure 6 visualizes the formation of clusters of data points in proximity representing the nearest neighbors in K- Nearest Neighbors Model.

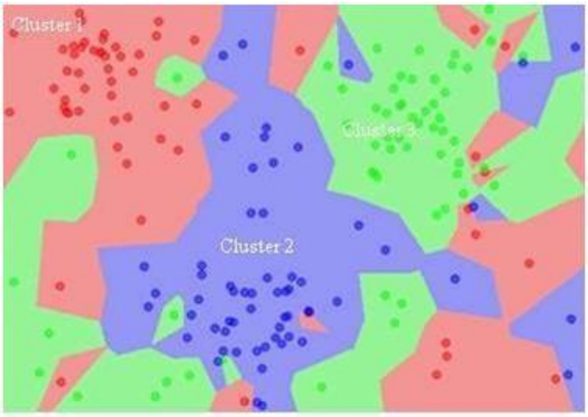


Figure 6: Visualization of K-Nearest Neighbors Model [22]

f) Gradient boosting DT

As the name suggests, in this approach small steps are initiated from a point in a direction by enhancing the weak learners to make them strong. It consists of a cost function, feeble leaner, and preservative sequential approach to improving the presentation of the predictive model [23]. This classifier algorithm is highly used to optimize the user-defined cost functions by using the gradients in the loss function to make them controlled and realistic [24]. The Figure 7 demonstrates the process of gradient boosting by the formation of efficient decision tree with repetitive error elimination from the previous decision trees.

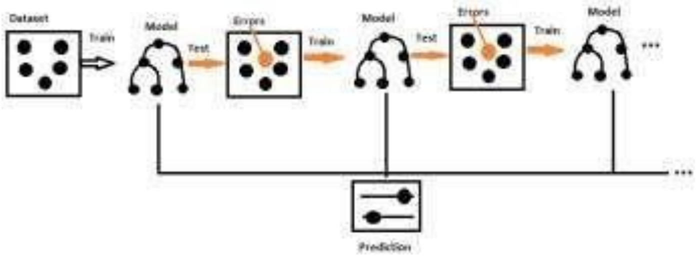


Figure 7: Visualization of Gradient Boosting DT Model [25]

g) Multi-layered perceptron

As the name suggests, it refers to the neural networks or system of input, output layers, and various hidden layers between them with multiple neurons connected. A perceptron is referred to as a neuron with a random activation function. This algorithm uses the technique of backpropagation, a repetitive approach of combining the weights and the inputs which are achieved through the threshold function to minimize the cost function [26]. The Figure 8 Shows the working of the Multilayer Perceptron Model by supplying weightsand inputs to the activation functions iteratively in the hidden layers to produce an output.

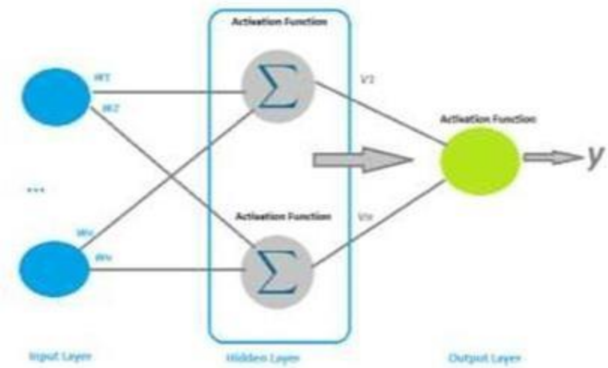


Figure 8: Visualization of Multilayer Perceptron Model [27]

h) XGBoost

This is known as an extreme gradient boosting algorithm. In this approach the framework of gradient boosting is conserved. It is a highly optimized algorithm in terms of software as well as hardware resources usage for supercilious prediction outcomes in a quick time with minimal computing cost. This approach involves Gradient descent methodology as gradient boosting for strengthening the weak learners like CARTs [28]. The Figure 9 describes the internal processes of XGBoost Model like interface compatibility, feature importance analysis, extendibility, flexibility, cross validation & model tuning, system processes, parameter tuning, optimization algorithms.

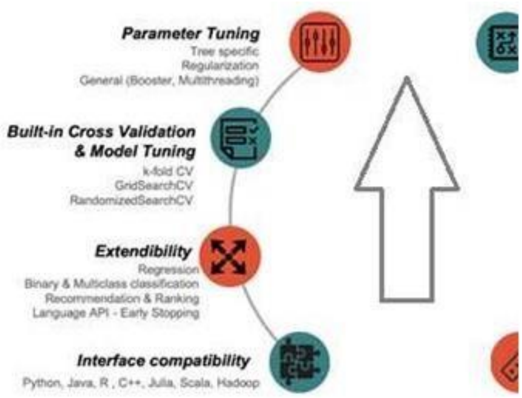


Figure 9: Visualization of XGBoost Model [29]

III. RESULTS AND DISCUSSION

This section will analyse all the classifier accuracy and performance using the confusion matrix and will give the proper comprehension regarding the best classifier algorithm. To do this analysis few pre-processing step is to be applied on used dataset. First the count of null values of attributes will be checked, which is shown in Table 2. So, we can clearly see that there are 8 rows of age attribute with null values. Now, describe the dataset to check the composition of the dataset. After this, the descriptive statistics of the dataset is collected, which is shown in Table 3.

Table 2: Number of null values in attributes

Attributes	Null Values
erythema	0
scaling	0
definite borders	0
itching	0
koebner phenomenon	0
polygonal papules	0
follicular papules	0
oral mucosal involvement	0
knee and elbow involvement	0
scalp involvement	0
family history	0
melanin incontinence	0
eosinophils in the infiltrate	0
PNL infiltrate	0
fibrosis of the papillary dermis	0
exocytosis	0
acanthosis	0
hyperkeratosis	0
parakeratosis	0
clubbing of the rete ridges	0
elongation of the rete ridges	0
thinning of the suprapapillary epidermis	0
spongiform pustule	0
munro microabscess	0
focal hypergranulosis	0
disappearance of the granular layer	0
vacuolisation and damage of basal layer	0
spongiosis	0
saw-tooth appearance of retes	0
follicular horn plug	0
perifollicular parakeratosis	0
inflammatory monoluclearinfiltrate	0
band-like infiltrate	0
Age	8

Table 3: Description of dataset without replacing null values

Parameters	Values
count	358.000000
mean	36.296089
std	15.324557
min	0.000000
25%	25.000000
50%	35.000000
75%	49.750000
max	75.000000

Now, in lieu of the null values with the median value of the attribute and describe the dataset again. Table 4 describes the altered dataset, after replacing the null values present in age attribute.

Table 4: Description of dataset after replacing null values

Parameters	Values
count	366.000000
mean	36.363388
std	15.037366
min	7.000000
25%	25.000000
50%	35.000000
75%	48.000000
max	75.000000

Now, it can be clearly seen that there is only a slight difference in the mean frequency of the dataset. Mean frequency difference percentage = 0.18%. Now, import all the 8 classifier algorithms i.e., Naive Bayes, Support Vector Machines, Random Forest, XGBoost, Multi-layered perceptron, K-nearest neighbors, Decisiontree, Gradient boosting DT. By meeting the internal classes ration in both sets, divide the dataset into a train set and a test set with a test size of 0.1. Now, the next step is to check the distribution of classes in trainingand test set. Figure 10 show that the distribution of both the training and test set are in proportion.

Figure 10: Class distribution in training and test set

The internal ratio of classes is same in both the sets as shown in Figure 10. After training these models with various classifier algorithms mentioned above and test for the accuracy scores. Figure 11 describes the accuracy scores of the classifier algorithms. It is the number of correct predictions made divided by the total number of predictions made, multiplied by 100 to turn it into a percentage [30]. Now, plot the confusion matrices for the various classifier algorithms mentioned above. Confusion Matrix is a performance measurement for machine learning classification. Well, it is a performance measurement for machine learning classification problem where output can be two or more classes [30]. The Figures 12 shown beloware the confusion matrices of the classifiers algorithms which compares the predicted class of ESD byclassifiers algorithm to the actual class of ESD for a particular set of attribute values from test set.

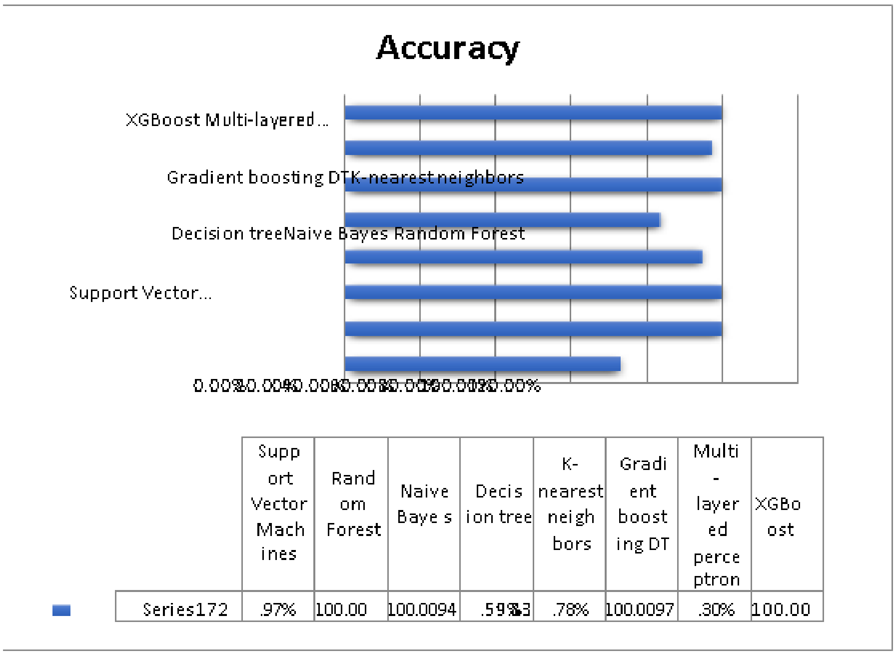
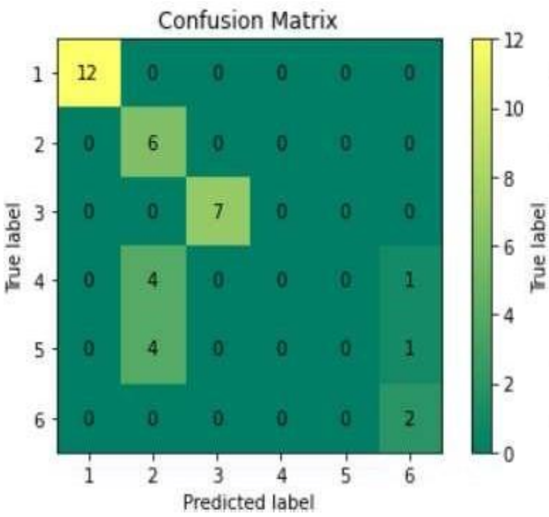
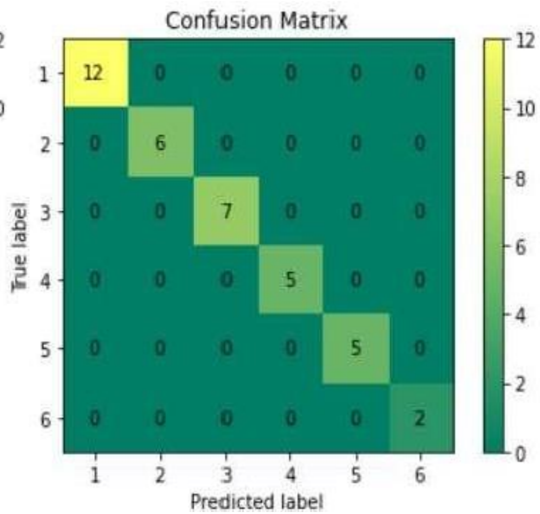


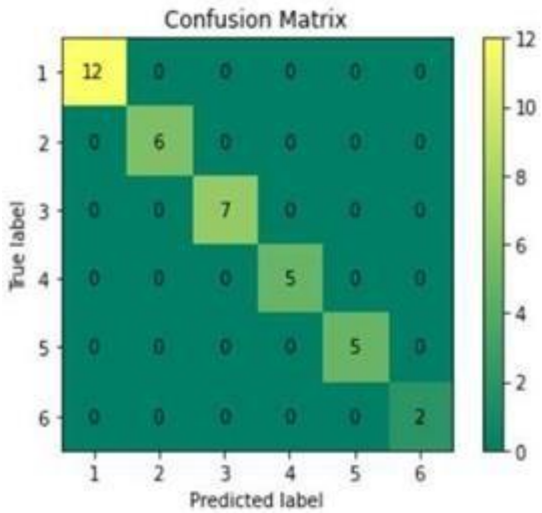
Figure 11: Accuracies of classifier algorithms

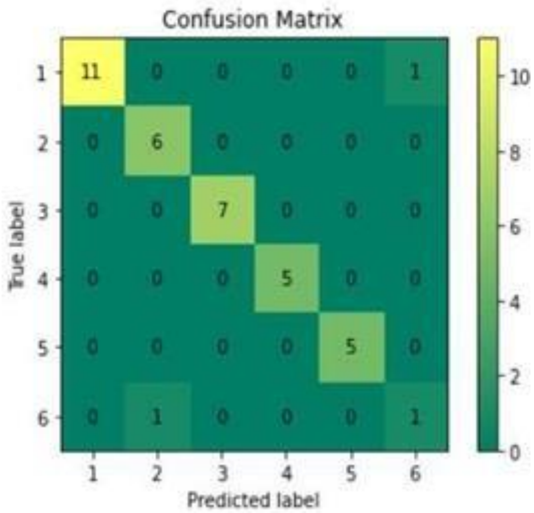


(a)

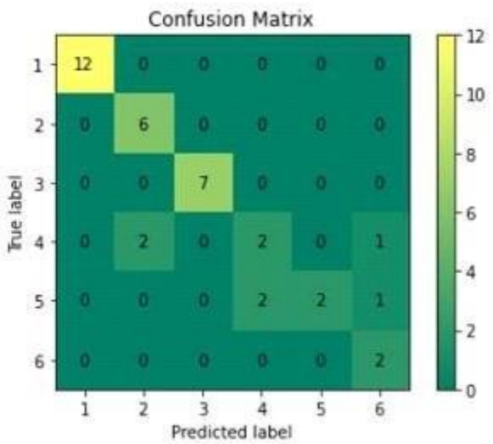


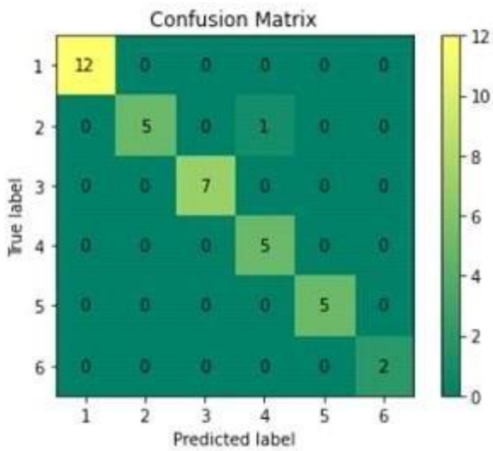
(c)

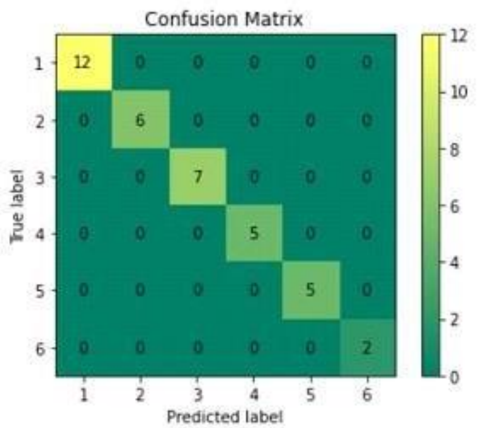


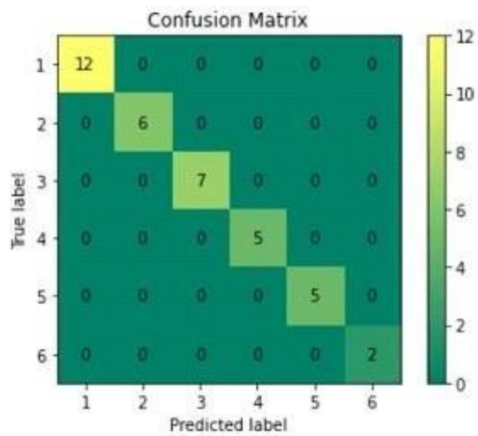


(b) (d)









(e)(f)(g)(h)

Figure 12 (a): SVM Confusion matrix, Figure (b): Random Forest Confusion matrix, Figure (c): Naive Bayes Confusion matrix, Figure (d): Decision tree Confusion matrix, Figure (e): KNN Confusion matrix, Figure (f): Gradient boosting DT Confusion matrix, Figure (g): MLP Confusion matrix, Figure (h): XGBoost Confusion matrix

From the above confusion matrices, we can clearly see predictions as per the accuracy scores of all the classes of ESD. The algorithms with 100 percent accuracy or the Ensemble of classifier algorithms with the best accuracies were employed for better classification about the prediction of the differential analysis of erythemato-squamous diseases constructed on their accuracies (ESD). By satisfying the internal classes ratio in both sets and checking the classification distribution in the trained and trial sets, the dataset was divided into a train set and a trial set with a test size of 0.1. After that, we can see that the internal class ratio is the same in both sets, and we also trained these models using the various classifier techniques discussed above, and we tested their accuracy: "Support Vector Machines (72.97%), Random Forest (100.0%), Naive Bayes (100.0%), Decision Tree (94.59%), K-nearest neighbors (83.78%), Gradient Boosting DT (100.0%), Multi- layered perceptron (97.3%), and XGBoost (100.0%) were the most popular".

IV. CHALLENGES AND FUTURE SCOPE

Challenges in the current study is as follows: (i) In the data set, there is 2.2 percent missing data for the age attribute. The mean frequency was used to replace missing data with true values. As a result, training M.L. model would have been more effective if we had used real data. Our dataset size is small so it can lead tomany problems like overfitting, Measurement errors, Missing values, Sampling Bias, etc. Due to this model accuracy will be low and can produce very bad results also at sometimes. Like in if have the biased data thenit can lead to the worst prediction [30]. (ii) The system will take time even if we use the best method with massive data. In some circumstances, this may result in the use of more CPU power. Furthermore, the data may take more storage space than is available. (iii) Vast quantity of data for training and testing is acquired. As a result of this technique, data inconsistencies may emerge. This is due to the fact that some data is updated on a frequent basis. As a result, we'll have to wait for more information. If this is not the case, the oldand new data may produce contradictory results.

Future scope of the current study is as follows: (i) Automatic diagnosis of these illness groupings could aid physicians in making decisions. (ii) Medical testing in hospitals must be kept to a minimum. They can achieve these results by utilising appropriate computer-based info and/or decision-making technology. (iii) We can improve our app by using a larger dataset and creating an app that can predict a huge number of diseases. (iv) We can establish the link between clinical and histological features using our feedback methodologies.

V. CONCLUSION

"Psoriasis, seboreic dermatitis, lichen planus, pityriasis rosea, chronic dermatitis, and pityriasis rubra pilaris are all classified as erythemato-squamous disorders (ESD), is possible with an artificial-intelligence-based approach through machine learning models using different classification algorithms, including Support Vector Machines, Random Forest, Naive Baye." We can see from the preceding calculations that utilizing the attribute's median to replace missing values results in a very small percentage change in the dataset's mean frequency. Furthermore, while comparing eight other classification methods, we can find that XGBoost fared exceptionally well with a 100% accuracy rate for three of them, namely Random Forest, Naive Bayes, and Gradient Boosting DT. XGBoost could be useful since it uses a mixture of software and hardware enhancement methods to provide supercilious prediction results with minimal computer assets in a small quantity of period.

VI. REFERENCES

[1]. K. E. Greiner AC, Health Professions Education: A Bridge to Quality., Washington: National Academies Press (US), 2003.

[2]. C. Brook, "What is a Health Information System?," 2020.

[3]. M. A.-D. A. M. N. Alaa M. Elsayad, "Analysis and Diagnosis of Erythemato-Squamous Diseases Using CHAID Decision Trees," Yasmine Hammamet, Tunisia, 2018.

[4]. M. C. Staff, "Psoriasis," 2020.

[5]. S. M. P. K. A. J. GARY W. CLARK, "Diagnosis and Treatment of Seborrheic Dermatitis," no. Feb 1, 2015 Issue, 2015.

[6]. R. P. USATINE, "Diagnosis and Treatment of Lichen Planus," no. Jul 1, 2011 Issue, 2011.

[7]. M. C. Staff, "Pityriasis rosea," 2020.

[8]. D. A. Oakley, "Dermatitis," 1997.

[9]. D. A. O. Vanessa Ngan, "Pityriasis rubra pilaris," 2015.

[10]. A. Verma, S. S. Pal and S. Kumar, "Comparison of skin disease prediction by feature selection using ensemble data mining techniques," Inf. Med.Unlocked, pp. 1-16, 2019.

[11]. H. A. G. a. H. A. G. Nilsel Ilter, "Dermatology Data Set," 01 January 1998. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/dermatology.

[12]. R. Gandhi, "Support Vector Machine — Introduction to Machine Learning Algorithms," 7 June 2018. [Online]. Available: https://towardsdatascience.com/support-vector- machine-introduction-to-machine- learning-algorithms-934a444fca47.

[13]. H. Wang, Y. Shao, S. Zhou, C. Zhang and N. Xiu, " Support Vector Machine Classifier via Soft-Margin Loss," IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1-11, 2021.

[14]. J. Brownlee, "Support Vector Machines for Machine Learning," 20 April 2016. [Online]. Available: https://machinelearningmastery.com/support- vector-machines-for-machine-learning/.

[15]. C. GOYAL, "Top 15 Questions to Test your Data Science Skills on SVM," 20 May 2021. [Online]. Available: https://www.analyticsvidhya.com/blog/2021/05/top- 15-questions-to-test-your-data-science-skills-on- svm/.

[16]. T. Yiu, "Understanding Random Forest," 12 June 2019. [Online]. Available: https://towardsdatascience.com/understanding- random-forest-58381e0602d2.

[17]. T. Yiu, "Understanding Random Forest," 12 June 2019. [Online]. Available: https://towardsdatascience.com/understanding- random-forest-58381e0602d2.

[18]. R. Gandhi, "Naive Bayes Classifier," 5 May 2018. [Online]. Available: https://towardsdatascience.com/naive-bayes- classifier-81d512f50a7c.

[19]. R. Gusain, "Naïve bayes classifier from scratch with hands on examples in r," 04 April 2020. [Online]. Available:https://insightimi.wordpress.com/2020/04/04/naive- bayes-classifier-from-scratch-with-hands-on-examples-in-r/.

[20]. P. Gupta, "Decision Trees in Machine Learning,"18 May 2017. [Online]. Available: https://towardsdatascience.com/decision-trees-in-machine-learning-641b9c4e8052.

[21]. N. S. Chauhan, "Decision Tree Algorithm, Explained," January 2020 . [Online]. Available: https://www.kdnuggets.com/2020/01/decision-tree- algorithm-explained.html.

[22]. O. Harrison, "Machine Learning Basics with the K- Nearest Neighbors Algorithm," 11 September 2018. [Online]. Available:https://towardsdatascience.com/machine-learning- basics-with-the-k-nearest-neighbors-algorithm- 6a6e71d01761.

[23]. J. Brownlee, "K-Nearest Neighbors for Machine Learning," 15 April 2016. [Online]. Available: https://machinelearningmastery.com/k-nearest- neighbors-for-machine-learning/.

[24]. J. Brownlee, "Develop k-Nearest Neighbors in Python From Scratch," 24 October 2019. [Online]. Available: https://machinelearningmastery.com/tutorial-to- implement-k-nearest-neighbors-in-python-from- scratch/.

[25]. J. Brownlee, "A Gentle Introduction to the Gradient Boosting Algorithm for Machine Learning," 9 September 2016. [Online]. Available: https://machinelearningmastery.com/gentle- introduction-gradient- boosting-algorithm-machine- learning/.

[26]. H. Singh, "Understanding Gradient Boosting Machines," November 4 2018. [Online]. Available: https://towardsdatascience.com/understanding- gradient-boosting-machines-9be756fe76ab.

[27]. "Gradient Boosting," [Online]. Available: https://docs.paperspace.com/machine- learning/wiki/gradient-boosting.

[28]. C. Bento, "Multilayer Perceptron Explained with a Real-Life Example and Python Code: Sentiment Analysis," 21 September 2021. [Online]. Available: https://towardsdatascience.com/multilayer- perceptron- explained-with-a-real-life-example-and- python-code-sentiment-analysis-cb408ee93141.

[29]. C. Bento, "Multilayer Perceptron Explained with a Real-Life Example and Python Code: Sentiment Analysis," September 2021. [Online]. Available: https://towardsdatascience.com/multilayer- perceptron- explained-with-a-real-life-example-and- python-code-sentiment-analysis-cb408ee93141.

[30]. V. Morde, "XGBoost Algorithm: Long May She Reign!," 8 April 2019. [Online]. Available: https://towardsdatascience.com/https-medium-com- vishalmorde-xgboost-algorithm-long-she-may-rein-edd9f99be63d.

[31]. S. Malik, "XGBoost: A Deep Dive Into Boosting,"23 March 2020. [Online]. Available: https://dzone.com/articles/xgboost-a-deep-dive- into-boosting.

[32]. R. Agrawal, "The 5 Classification Evaluation metrics every Data Scientist must know," 17 September 2019. [Online]. Available: https://towardsdatascience.com/the-5-classification- evaluation-metrics-you-must-know-aa97784ff226. [Accessed 9 Decemeber 2021].

[33]. EduPristine, "Problems of Small Data and How to Handle Them," 2016.

[34]. M. C. Staff, "Pityriasis rosea," 2020.