Hybrid Model Based on ReliefF Algorithm and K-Nearest Neighbor for Erythemato-Squamous Diseases Forecasting

Abdullah S. Alotaibi¹

Received: 6 March 2021 / Accepted: 23 June 2021 / Published online: 2 July 2021
© The Author(s) 2021

Abstract

Machine learning (ML) techniques have been used to solve real-world problems for decades. In the field of medical sciences, these approaches have been found to be useful in the diagnosis and prognosis of a variety of medical disorders. However, when dealing with voluminous, inconsistent, and higher-dimensional data, conventional ML approaches have failed to deliver the expected results. Researchers have suggested hybrid solutions to resolve these problems, which have been found to be more effective than conventional methods because these systems integrate their merits while reducing their drawbacks. In the current research article, hybrid model has been presented by coupling feature optimization with prediction approach. The proposed hybrid model has two stages: the first involves implementing the ReliefF Algorithm for optimum feature selection in erythemato-squamous diseases, and the second involves implementing k-nearest neighbor (KNN) for prediction of those selected optimum features. The experimentation was carried out on bench mark dataset for erythemato-squamous diseases. The presented hybrid model was also assessed with conventional KNN approach based on various metrics such as classification accuracy, kappa coefficient, recall, precision, and f-score.

Keywords KNN · ReliefF algorithm · Dermatology · Medical diagnosis · Machine learning

1 Introduction

Erythemato-squamous disease refers to skin disorders which cause the skin irritated, blocked, or inflamed. It results in various symptoms on skin such as rashes, lesions, macule, swelling, burning, and itching. It may be caused by inflammation, acne medications, asthma, solar radiation, photo-sensitization, acute radiation syndrome, bacteria, viruses or fungal infection, any of which may induce dilation of the capillaries, resulting in redness [1]. It can be of several types and its cure depends on the type of erythema. It may occur in numerous patterns and color variations in different areas of the body, and several types result in blisters erupting and causing sores. The severity of this disorder varies from moderate to life threatening. Mild conditions get recovered without cure in some days. Medication or emergency treatment is needed in case of more extreme cases [2].

As per the Global Burden of Disease report, skin disorders have been ranked as the 4th most common cause of non-fatal diseases across the globe [3]. According to [4], around 7.5 million people have been diagnosed with psoriasis, 1 million with melanoma, and 16 million with rosacea. In fact, it is difficult to diagnose erythemato-squamous disease without the assistance of a specialist since there are a number of possible skin conditions and these diseases have histopathological conditions also; so it is best to consult the doctor for any changes. Biopsy is recommended in order to diagnose these illnesses effectively [5].

Hence, it needs research in this domain and many researchers are working in this direction. In order to diagnose these disorders effectively, machine learning approaches have been practiced by researchers since recent years. These techniques have been found to be efficient as they aid dermatologists to make their diagnosis accurate [6].

However, when dealing with non-linear, inconsistent, and higher-dimensional data, these traditional ML techniques have performance issues. To address these issues, researchers have begun to introduce hybrid and ensemble models by combining conventional machine learning methods with other soft computing approaches or by using bagging/

✉ Abdullah S. Alotaibi
a.shawan@su.edu.sa

¹ Shaqra University, Shaqra, Saudi Arabia

boosting methodology. Hybridizing two or more soft computing techniques improves efficiency by incorporating the benefits of the each technique. These hybrid methods have proven to be very successful for medical diagnosis in cases where data is large, multi-dimensional, and non-linear [7]. This has motivated the authors of the paper to implement hybrid system based on feature optimization and prediction approach for erythemato-squamous disease diagnosis so that the undertaken disease can be predicted with better accuracy.

This paper presents a hybrid model by coupling ReliefF Algorithm with K-Nearest Neighbor (KNN) approach for the prediction of erythemato-squamous disorders. The proposed hybrid model works in two stages: the first stage deals with the implementation of ReliefF Algorithm for optimum feature selection involved in erythemato-squamous diseases; and the second stage deals with implementation of KNN for classifying those selected optimum features. The experiments have been conducted on benchmark dataset of erythemato-squamous disorders. This dataset comprises six erythemato-squamous diseases namely *psoriasis*, *seboric dermatitis*, *lichen planus*, *pityriasis rosea*, *chronic dermatitis*, and *pityriasis rubra pilaris*.

For the current study, ReliefF algorithm has been selected as feature optimization technique because this approach has been found to be better in case of multi-class output and categorical input attributes [8]. KNN approach has been selected as prediction technique because KNN performs better in case of heterogeneous data, i.e., data set has numeric and categorical input attributes. Also, KNN approach predicts highly convoluted decision boundary in case of multi-dimensional data [9]. The collected dataset of erythemato-squamous disease consists of 32 categorical and 02 numeric input attributes; and six output classes.

This research article is organized as follows: After introduction, Sect. 2 sheds light on the literature relevant to the present work. Section 3 elucidates methodology adopted to implement the proposed hybrid system. Section 4 analyzes and discusses about the experimental outcomes, and Sect. 5 outlines the conclusion of the paper.

2 Literature Review

In order to diagnose skin diseases, several ML based approaches have been implemented by researchers in recent years. These techniques include regression approach, classifiers, association rule mining, decision trees, and neural networks. These approaches assist dermatologists from diagnosis to customized cure [6].

Guenir et al. [10] proposed a *voting feature interval VFI5*-based method for the identification of Erythemato-squamous disorders. The proposed approach was

implemented with tenfold cross validation, and the proposed approach achieved classification accuracy of 99.2%.

A model for scaling segmentation in 2-D (two dimensional) digital images has been presented by Lu et al. [11] for evaluation of psoriasis severity. The proposed approach integrates Markov random field (MRF) approach with support vector machine (SVM). The experimental results predicted sensitivity of 81.79% with specificity of 87%.

Pomponiu et al. [12] applied pre-trained deep neural network for skin lesion identification. The experimentation was carried out on DermIS dataset consisting of 399 images of skin mole lesion. The proposed system achieved the accuracy of 93.64% along with recall of 92.10% and specificity of 95.18%.

Next year, Gustafson et al. [13] presented a machine learning based system for the diagnosis of atopic dermatitis in adults. The proposed system coupled natural language processing (NLP) with lasso logistic regression technique for the prediction of atopic dermatitis from electronic health records (EHR). The experimental results achieved sensitivity of 75% with positive predicted value (PPV) of 84%.

An automated system has been developed by Wei et al. [14] for detection of herpes, dermatitis, and psoriasis skin diseases from image color and texture features. The proposed system worked by pre-processing the input image for removal of noise and irrelevant background through the implementation of median filter and transformation approaches; followed by gray-level co-occurrence matrix (GLCM) implementation for feature extraction. Finally the extracted features were fed to SVM technique for disease identification undertaken in the study. The presented system predicted maximum accuracy of 95%.

An image processing based approach was presented by Alenezi [15] for the recognition of three types of skin ailments: Melanoma, Eczema, and Psoriasis. The presented approach worked by implementing AlexNet deep neural network for extracting features from images, and support vector machine (SVM) approach for categorization. The experimental results revealed the classification accuracy of 100% for the proposed system.

In the same year, Jamian et al. [16] presented an EHR-based approach for the prediction of systemic sclerosis by incorporating billing codes and random forest technique with clinical data. The presented system achieved positive predicted value of 84% with recall of 92%, and f-score of 88%.

Padmavathi et al. [17] applied deep neural networks such as convolutional neural network (CNN) and residual network (ResNet) for the prediction of dermatological diseases. The experiments have been performed on the dataset of 10,015 dermatoscopic images categorized into seven different classes. The experimental results revealed the classification accuracy of 77% for CNN and 68% for ResNet.

A hybrid model was presented by Rajasekaran et al. [18] for the identification of skin diseases. The proposed hybrid model comprises of four steps: pre-processing, segmentation, feature extraction, and classification. The step of pre-processing phase involves gray scale conversion, application of median filter, and binary mask. Segmentation is performed by otsu's method and histogram chart approach. Feature extraction involves application of wavelet transform for extracting the features from the image. And, the last stage employs CNN to classify the extracted features in order to identify the skin disease.

George et al. [19] introduced a framework based on bag-of-visual words approach, k-means clustering, and SVM for evaluating psoriasis severity in skin images. The presented framework achieved accuracy of 80.81%.

Machine learning algorithms can be used to predict skin diseases, according to a study of the related literature. Dermatology disorders have been diagnosed using deep learning techniques from imaging data. The traditional ML approaches have been found to be successful when dealing with clinical data. However, when dealing with voluminous, inconsistent, and higher-dimensional data; conventional ML approaches have failed to achieve the expected accuracy. Hybrid systems have been suggested by researchers as a means of resolving these problems. These hybrid models have been found to be more effective than conventional methods because they integrate their advantages while minimizing their disadvantages. This has prompted the authors to work on a hybrid method for diagnosing skin disease from clinical data, with the aim of improving the efficiency of traditional machine learning technique.

3 Materials and Methodology

This research work aims to present a hybrid system for *Erythemat-Squamous* ailments prediction. The proposed hybrid approach integrates ReliefF Algorithm with K-nearest neighbor (KNN) approach, and it works in two stages. The first stage deals with the implementation of *ReliefF Algorithm* for optimum feature selection involved in erythemat-squamous diseases; and the second stage deals with implementation of *KNN* for classifying those selected optimum features in order to predict form of Erythemat-Squamous disorder. The experimentation has been performed out on MATLAB 2018 tool.

3.1 Dataset for Erythemat-Squamous Diseases

In this research study, the experimentation has been conducted on benchmark dataset of Erythemat-Squamous disorder. This dataset has been collected from University of California, Irvine (UCI) Machine Learning Repository

(<https://archive.ics.uci.edu/ml/datasets/dermatology>). It contains 366 records, each with 34 input attributes and 1 output attribute. Among 34 input attributes, 12 present clinical attributes and 22 present histopathological attributes. 'Age' attribute represents patient's age, and it has integer value. 'Family History' attribute has nominal value, i.e., 1 (means this disease has occurred in the family) or 0 (this disease has not occurred). All the remaining 32 input attributes are having integer values from 0 to 3 [0-feature is absent, 1-mild value, 2-moderate value, 3-severe value]. The output attribute represents six classes namely *lichen planus*, *psoriasis*, *pityriasis rosea*, *seboric dermatitis*, *chronic dermatitis*, and *pityriasis rubra pilaris*. The particulars of the dataset attributes are presented in Tables 1 and 2.

3.2 Techniques used for Implementation of Hybrid Model

In the research work, a hybrid model has been developed by coupling ReliefF Algorithm with K-nearest neighbor (KNN) for erythemat-squamous disorders Prediction. The background of these techniques is presented as under:

3.2.1 ReliefF Algorithm

This algorithm computes the weights of features in scenarios where the Class is a multi-class and categorical. This algorithm penalizes the predictors which produces different values for the neighbors of a same Class and rewards those predictors which results different values for the different classes. At the onset this algorithm sets the weight of all the predictors to zero. Then, the algorithm iteratively selects any observation randomly and finds the k-nearest observations to the selected observation for each Class.

If x_r and x_q do belong to the same Class, then weights updating has been done for each nearest neighbor for a given predictor as:

$$W_j^i = W_j^{i-1} - \frac{\Delta_j(x_r, x_q)}{m} \cdot d_{rq}.$$

If x_r and x_q do belong to the different classes, then weights updating for each nearest neighbor for a given predictor as:

$$W_j^i = W_j^{i-1} + \frac{p_{y_q}}{1 - p_{y_r}} - \frac{\Delta_j(x_r, x_q)}{m} \cdot d_{rq}.$$

where W_j^i is the weight of the attribute F_j at the i th iteration, P_{y_r} is prior probability of the class to which X_r belongs, P_{y_q} is the probability of class to which X_q belongs and m is number of iterations.

ReliefF has actually a family of algorithms and is proven to be very successful feature estimator. The family of relief

Table 1 Description of dataset input attributes

S. no	Clinical attributes	S. no	Histopathological attributes
1	Erythema	12	Melanin incontinence
2	Scaling	13	Eosinophils in the infiltrate
3	Definite borders	14	PNL infiltrate
4	Itching	15	Fibrosis of the papillary dermis
5	Koebner phenomenon	16	Exocytosis
6	Polygonal papules	17	Acanthosis
7	Follicular papules	18	Hyperkeratosis
8	Oral mucosal involvement	19	Parakeratosis
9	Knee and elbow involvement	20	Clubbing of the rete ridges
10	Scalp involvement	21	Elongation of the rete ridges
11	Family history (0 or 1)	22	Thinning of the suprapapillary epidermis
34	Age (linear)	23	Spongiform pustule
		24	Munro microabscess
		25	Focal hypergranulosis
		26	Disappearance of the granular layer
		27	Vacuolization and damage of basal layer
		28	Spongiosis
		29	Saw-tooth appearance of retes
		30	Follicular horn plug
		31	Perifollicular parakeratosis
		32	Inflammatory mononuclear infiltrate
		33	Band-like infiltrate

Table 2 Details of output attribute

S. no	Output class	No. of instances
1	Psoriasis	112
2	Seboric dermatitis	61
3	Lichen planus	72
4	Pityriasis rosea	49
5	Chronic dermatitis,	50
6	Pityriasis rubra pilaris	20

algorithms uses conditional dependencies among the features and produces a unified idea of feature selection in classification and regression problems. These algorithms are widely known as feature selection techniques where as they are also used in pre-processing stages before the model is trained [20].

In this study, relief algorithm has been implemented with input attributes as *predictor* variables and output attribute as *response* vector with seven nearest neighbors per Class. The rank values and weights for each input feature determined by relief algorithm are presented in Table 3. It has been analyzed that feature no. 15 (*Fibrosis of the Papillary Dermis*) is found to be the most important attribute whereas feature no. 3 (*Definite Borders*) is computed to be the least significant attribute in the Erythemato-Squamous disorder dataset. Out

of 34 input attributes, first 20 most significant attributes has been selected for classification to be done in the next phase of the proposed hybrid model. The details of these selected input features are presented in Table 4.

3.2.2 K-Nearest Neighbor

KNN is one of machine learning approaches, that is non-parametrically supervised and it is applied for both classification and regression. Classification of the given input record into the output Class label depends on the features of neighboring records and the value of K . It operates through the measurement of distance metrics by tracing the pattern of similarity between data points. Such metrics take into account different measurements of distance, such as *Manhattan*, *Euclidean*, *Hamming*, or *Cityblock*. In order to assess the nearest neighbors, the distance metric measures the minimum distance from the test instance to the training results. The output Class of the test example belongs to most of the groups of k-nearest neighbors. In many research publications, *Euclidean measure* has been implemented to compute minimum distance between two data points in the dataset, which is represented as under:

$$\text{Euclid}(V, W) = \sqrt{(v_1 - w_1)^2 + (v_2 - w_2)^2 + \dots + (v_n - w_n)^2}$$

Table 3 Weight and rank of each attribute determined by ReliefF algorithm

S. no	Attribute	Rank	Weight
1	Erythema	29	−0.035479
2	Scaling	11	−0.000208
3	Definite borders	34	−0.052189
4	Itching	19	−0.01129
5	Koebner phenomenon	5	0.04283
6	Polygonal papules	21	−0.02564
7	Follicular papules	9	0.002027
8	Oral mucosal involvement	26	−0.03231
9	Knee and elbow involvement	17	−0.00681
10	Scalp involvement	32	−0.0407
11	Family history	33	−0.05083
12	Melanin incontinence	23	−0.02854
13	Eosinophils in the infiltrate	6	0.03677
14	PNL infiltrate	2	0.092184
15	Fibrosis of the papillary dermis	1	0.10199
16	Exocytosis	18	−0.0085
17	Acanthosis	24	−0.0285
18	Hyperkeratosis	30	−0.0377
19	Parakeratosis	10	0.0008494
20	Clubbing of the rete ridges	16	−0.006112
21	Elongation of the rete ridges	12	−0.0032859
22	Thinning of the suprapapillary epidermis	3	0.07161
23	Spongiform pustule	28	−0.03440
24	Munro microabscess	31	−0.03859
25	Focal hypergranulosis	27	−0.03335
26	Disappearance of the granular layer	20	−0.025147
27	Vacuolization and damage of basal layer	25	−0.028681
28	Spongiosis	4	0.052684
29	Saw-tooth appearance of retes	22	−0.02676
30	Follicular horn plug	15	−0.00538
31	Perifollicular parakeratosis	8	0.00867
32	Inflammatory mononuclear infiltrate	14	−0.0048
33	Band-like infiltrate	7	0.01065
34	Age	13	−0.00340

where V and W correspond to the records in *training set* and the *test set*, respectively, with n input attributes in the data [21].

Here, KNN model is implemented as a classifier in the second phase of the hybrid model. The presented hybrid model has been trained via KNN approach with 20 *significant input parameters* (selected by ReliefF Algorithm) and 1 *output parameter* with 6 classes from the dataset under study. The KNN ML approach was applied with *tenfold cross validation* and *Euclidean* distance metric. It was found that the most favorable number of nearest neighbors

Table 4 Optimum features selected by ReliefF algorithm

S. no	Attribute no	Feature
1	A_2	Scaling
2	A_4	Itching
3	A_5	Koebner phenomenon
4	A_7	Follicular papules
5	A_9	Knee and elbow involvement
6	A_{13}	Eosinophils in the infiltrate
7	A_{14}	PNL infiltrate
8	A_{15}	Fibrosis of the papillary dermis
9	A_{16}	Exocytosis
10	A_{19}	Parakeratosis
11	A_{20}	Clubbing of the rete ridges
12	A_{21}	Elongation of the rete ridges
13	A_{22}	Thinning of the suprapapillary epidermis
14	A_{26}	Disappearance of the granular layer
15	A_{28}	Spongiosis
16	A_{30}	Follicular horn plug
17	A_{31}	Perifollicular parakeratosis
18	A_{32}	Inflammatory mononuclear infiltrate
19	A_{33}	Band-like infiltrate
20	A_{34}	Age

(k) was 5. KNN approach classifies the input data into one of six classes of erythematous-squamous diseases.

3.3 Implementation of Proposed Hybrid Architecture

In order to implement and evaluate the proposed hybrid architecture, benchmark dataset has been collected. After this, *data pre-processing* has been conducted to deal with missing values using imputation. The outliers present in the data have been identified through application of statistical technique. In order to convert all the input attributes in the range $[0, 1]$, *Min–Max* normalization method has been applied; as presented in following equation.

$$M' = \frac{M - \text{minimum}(A)}{\text{maximum}(A) - \text{minimum}(A)}$$

where M corresponds to the true value of the attribute A , M' corresponds to normalized value of the respective attribute, $\text{minimum}(A)$ refers to smallest value and $\text{maximum}(A)$ is the largest value in the attribute A .

After normalization, the proposed hybrid model has been implemented. This hybrid model works in two phases. In the first phase, *ReliefF Algorithm* has been implemented on the pre-processed data for selecting the most significant attributes in the dataset. In a higher-dimensional dataset with fewer instances, majority of the features are not important

for data classification most times. The large number of features with limited records results in high computational cost that affects the model's efficiency and can result in overfitting of the learning model. Under these circumstances, features reduction plays an important role since it reduces space and time complexity of learning algorithm. In the second phase, after feature selection, *k-nearest neighbor* (KNN) method has been applied for the model training to identify Erythematous-Squamous disorders. In this phase, the data (with selected features) has been categorized as per ratio of 80:20 to two sub-sets: *training set* and *test set*; thus *training set* had 292 samples and *test set* had 74 samples. The validation of the proposed hybrid model was done with *test set* in order to appraise its efficiency. The presented hybrid model has also been compared with conventional KNN approach by five main measures namely *accuracy*, *F-score*, *recall*, *kappa coefficient* and *precision*. The overall outline of the hybrid architecture is presented in Fig. 1.

4 Results and Discussion

In the current research work, hybrid model based on ReliefF Algorithm and KNN approach has been presented for Erythematous-Squamous disorders prediction. From the experimental findings, it has been evaluated that the proposed hybrid model can be well utilized to predict undertaken medical diseases. The hybrid model has been validated based on performance metrics such as *accuracy*, *recall*, *precision*, *f-score* and *kappa coefficient*. In case of

medical diagnosis, the doctors want that diagnostic system should predict accurate results. This diagnostic system is also expected to prevent *false positives* as well as *false negatives* in order to avoid misclassification of the disease. In this research work; accuracy, recall, precision, kappa coefficient, and f-score metrics have been adopted for validation of the hybrid model. The reason is that *Accuracy* depicts performance of the model; *Precision* tells about False Positives, *Recall* points out false negatives. *F-score* helps in resolving the issue of *low variance and high bias* vs *high variance and low bias*.

The experimental results are presented in Table 5, and these results revealed that the proposed hybrid model achieved *accuracy* of 94.59% with *recall* of 90.94%, *precision* of 94.72%, kappa coefficient of 0.932 and *f-score* of 0.9279. The traditional KNN approach predicted *accuracy* of 85.13% with *recall* of 81%, *precision* of 79.89%, kappa coefficient of 0.812 and *f-score* of 0.7993. The comparative analysis of the proposed hybrid model with traditional KNN approach is presented in Fig. 2.

Table 5 Experimental results for hybrid model and traditional KNN

Parameter	Hybrid model	Traditional KNN
Accuracy	94.59%	85.13%
Recall	90.94%	81%
Precision	94.72%	79.89%
F-score	0.9279	0.7993
Kappa's coefficient	0.932	0.812

Fig. 1 Framework for the hybrid architecture

The confusion matrix for the proposed hybrid model and traditional KNN is presented in Fig. 3a and b, respectively. From the confusion matrices, it has been found out that traditional KNN predicted somewhat good results for Class 1 (*psoriasis*), Class 3 (*lichen planus*), and Class 5 (*chronic dermatitis*). But it predicted low performance for Class 2 (*seboric dermatitis*) and Class 4 (*pityriasis rosea*). For Class 1, the performance of both hybrid model and traditional KNN was same. For Class 2, traditional approach achieved *precision* of 44.4% and *recall* of 50%. For Class 4, it has shown *recall* of 50% and *precision* of 55.6%. For Class 6, it had achieved *recall* of 83%. But the proposed hybrid model achieved better results as compared to traditional approach for all the Classes. Hybrid model predicted *Class-wise recall* as 100%, 75%, 100%, 100%, 93%, and 100%.

It had achieved *Class-wise precision* as 100%, 100%, 93%, 77.8%, 100%, and 75%.

Similarly, *Receiver Operating Characteristic* (ROC) curve for the proposed hybrid model and traditional KNN is presented in Fig. 4a and b, respectively. It also shows that traditional KNN approach predicted poor performance for Class 2 and Class 4 as true positive rate predicted by hybrid model is better than that of traditional KNN approach.

The hybrid model outperformed conventional KNN in terms of accuracy, recall, precision, kappa coefficient and *f*-score values, according to the results of the experiments. The hybrid model predicts a higher true positive rate than conventional KNN. The conventional approach had more false positives and false negatives than those of hybrid model. Thus, the proposed hybrid model outperformed

Fig. 2 Performance metrics for hybrid model versus traditional KNN

Fig. 3 Confusion matrix for test data **a** hybrid model, **b** traditional KNN

Fig. 4 ROC (receiver operating characteristic) curve for hybrid model versus traditional KNN

the traditional KNN approach. The explanation for this is that conventional KNN used all of the features in the disease dataset to classify the disease; as a result, the data has become more dimensional, and certain features may be less relevant. It has a higher computational cost, which reduces the efficiency of the trained model and causes overfitting.

Only significant attributes present in the data were taken into account in the proposed hybrid model, resulting in the data's dimensions being compacted and the learned model's performance being improved.

5 Conclusion

In this paper, hybrid system has been presented for the forecasting of Erythematous-Squamous disorders from clinical data. From the experimental findings, it has been found out that the proposed model can be effectively used for the diagnosis of the disease under study. This model has improved the performance of traditional KNN by eliminating less contributing features present in the dataset, thus increasing the efficiency of trained model. The main limitation of this model is that it can't be used on imaging data or biomedical signals as it has been trained on clinical data. In the future, the same hybrid model may be used to diagnose other medical conditions. This model can also be updated to diagnose medical conditions using imaging data or biomedical signals.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Erythema: <https://en.wikipedia.org/wiki/Erythema>. Accessed 3 Jan 2021
2. Erythema: <https://www.drugwatch.com/health/rash-and-skin-disorders/erythema/>. Accessed 3 Jan 2021
3. Seth, D.; Cheldize, K.; Brown, D.; Freeman, E.F.: Global burden of skin disease: inequities and innovations. *Current Dermatology Reports* **6**(3), 204–210 (2017). <https://doi.org/10.1007/s13671-017-0192-7>
4. Skin Conditions by the Numbers: <https://www.aad.org/media/stats-numbers>. Accessed 3 Jan 2021
5. Dermatoses: <https://www.healthline.com/health/dermatoses#diagnosis>. Accessed 3 Jan 2021
6. Chan, S.; Reddy, V.; Myers, B.; Thibodeaux, Q.; Brownstone, N.; Liao, W.: Machine learning in dermatology: current applications, opportunities, and limitations. *Dermatol. Ther. (Heidelb.)* (2020). <https://doi.org/10.1007/s13555-020-00372-0>
7. Ardabili, S.; Mosavi, A.; Annamaria, R.: Advances in machine learning modeling reviewing hybrid and ensemble methods. In: *Proceedings of the International Conference on Global Research and Education, Lecture Notes in Networks and Systems*, vol. 101, pp. 215–227 (2020). https://doi.org/10.1007/978-3-030-36841-8_21
8. Durgabai, R.P.L.: Feature selection using ReliefF algorithm. *Int. J. Adv. Res. Comput. Commun. Eng.* **3**(10), 8215–8218 (2014)
9. Ali, N.; Neagu, D.; Trundle, P.: Evaluation of k-nearest neighbour classifier performance for heterogeneous data sets. *SN Appl. Sci.* **1**, 1559 (2019). <https://doi.org/10.1007/s42452-019-1356-9>
10. Guvenir, H.A.; Demiroz, G.; Ilter, N.: Learning differential diagnosis of erythematous squamous diseases using voting feature intervals. *Artif. Intell. Med.* **13**, 147–165 (1998)
11. Lu, J.; Kazmierczak, E.; Manton, J.H.; Sinclair, R.: Automatic segmentation of scaling in 2-D psoriasis skin images. *IEEE Trans. Med. Imaging* **32**(4), 719–730 (2013). <https://doi.org/10.1109/TMI.2012.2236349>
12. Pomponiu, V.; Nejati, H.; Cheung, N.M.: Deepmole: Deep neural networks for skin mole lesion classification. In: *Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP)*, Phoenix, AZ, pp. 2623–2627 (2016). <https://doi.org/10.1109/ICIP.2016.7532834>
13. Gustafson, E.; Pacheco, J.; Wehbe, F.; Silverberg, J.; Thompson, W.: A machine learning algorithm for identifying atopic dermatitis in adults from electronic health records. In: *Proceedings of the IEEE International Conference on Healthcare Informatics*, pp. 83–90 (2017). <https://doi.org/10.1109/ICHI.2017.31>
14. Wei, L.; Gan, Q.; Ji, T.: Skin disease recognition method based on image color and texture features. *Comput. Math. Methods Med.* (2018). <https://doi.org/10.1155/2018/8145713>
15. Alenezi, N.: A method of skin disease detection using image processing and machine learning. In: *Proceedings of the 16th International Learning and Technology Conference 2019, Procedia Computer Science*, vol. 163, pp. 85–92 (2019)
16. Jamian, L.; Wheless, L.; Crofford, L.J.; Barnado, A.: Rule-based and machine learning algorithms identify patients with systemic sclerosis accurately in the electronic health record. *Arthritis Res. Ther.* (2019). <https://doi.org/10.1186/s13075-019-2092-7>
17. Padmavathi, S.; Mithaa, E.M.; Kiruthika, T.; Ruba, M.: Skin diseases prediction using deep learning framework. *Int. J. Recent Technol. Eng.* (2020). <https://doi.org/10.35940/ijrte.F9038.038620>
18. Rajasekaran, G.; Aiswarya, N.; Keerthana, R.: Skin disease identification using image processing and machine learning techniques. *Int. Res. J. Eng. Technol.* **7**(3), 1368–1371 (2020)
19. George, Y.; Aldeen, M.; Garnavi, R.: Automatic scale severity assessment method in psoriasis skin images using local descriptors. *IEEE J. Biomed. Health Inform.* **24**(2), 577–585 (2020). <https://doi.org/10.1109/JBHI.2019.2910883>
20. Robnik-Sikonja, M.; Kononenko, I.: Theoretical and empirical analysis of ReliefF and RReliefF. *Mach. Learn.* **53**, 23–69 (2003)
21. Brownlee, J.: *K Nearest Neighbors for Machine Learning* (2016). <https://machinelearningmastery.com>. Accessed 23 Dec 2020