Improving Differential Diagnosis of Pathologically Similar Dermatological Conditions to Minimize Invasive Procedures

Jason Bouhenguel **Ryan Gallagher**
Stanford University, Department of Bioengineering

Abstract

The erythemato-squamous dermatological diseases share many clinical and histopathological attributes making their differential diagnosis a challenging task for physicians. In this paper, we develop a classification algorithm based on previous clinical data to assist dermatologists in differentiating among these conditions. The final embodiment of the model was an optimized SVM with a Guassian kernel (γ = 2-8) combined with principal component analysis (PCA), achieving a maximum classification accuracy of 97.8%, an average sensitivity and specificity of 97.64% and 99.57%, respectively. Feature selection revealed a subset of features with relatively high positive predictive impact (e.g., itchiness of affected area) and relatively negative predictive impact (e.g., age). Optimized models trained on only non-invasive feature subsets (features not requiring invasive biopsy) yielded an average sensitivity of 81.56% and a specificity of 96.78%, suggesting the utility of our model as a first-line screening tool for dermatologists to rule-out the more severe diseases, and thus minimize subsequent invasive testing.

Introduction

Computer-aided diagnostic tools hold much promise for assisting in the differential diagnosis of diseases that bear minor differences in presentation. Such tools often combine elements of artificial intelligence, digital image processing, and computer vision to improve diagnostic accuracy of radiologists when interpreting X-ray, MRI, ultrasound, and other radiological results. Since the initial use of these technologies in radiology, there has been a demand for similar tools in numerous other medical specialties. However, because of difficulty of image standardization and a paucity of readily extractable features, such tools have developed more slowly for some specialties than others.

Traditionally, dermatologists have based their diagnoses of skin conditions on visual inspection of the pathological area and assessment of histological samples biopsied from the area of concern. Diagnostic accuracy in dermatology has been shown to be highly dependent on the experience of the practitioner and their visual acumen. To avoid potential misdiagnoses resulting from these human factors, the dermatological community has shown increasing interest in automated and aided diagnostic tools.

In particular, these tools hold much promise to assisting in narrowing the differential diagnosis of skin diseases that bear only small differences in symptoms and clinical features. The erythemato-squamous dermatological diseases, literally “red-skin diseases”, share many clinical attributes including erythema (redness), itching, and scaling, for example, making their differential diagnosis a challenging task for physicians. Beyond visual similarity, these conditions share several histopathological similarities, as well, and the differences are often nuanced. It is from these indistinguishable differences that dermatologists often elect to perform more, sometimes potentially extraneous biopsies in an effort to bolster the certainty of their diagnosis. In much of today’s medical practice, the consequences of biopsies are overlooked; however, similar to other invasive procedures, biopsies are not without consequence—increasing the risk of procedure-related complications (e.g.,

infection, scarring, chronic pain, etc.), patient anxiety, and overall medical costs.

In this project we first investigated the use of support vector machines (SVM) to predict the correct diagnosis of the erythemato-squamous diseases based on non-invasive clinical and histopathological features of 350+ patients. The model was optimized by exploring the effects of feature selection and PCA, in addition to parameter optimization specifically, l_1-regularization and kernel selection—Gaussian, sigmoid, linear, and polynomial. Utilizing this optimization scheme we subsequently assessed model performance when trained on a subset of features, specifically comparing the use of non-invasive features to those derived from invasive biopsy.

Data Set

The data set used for this project was obtained from the University of California Irvine Machine Learning Repository. The data include a total of 34 discrete-valued features and 366 examples collected in a clinical setting. A 35th nominal attribute is included which indicates the dermatological diagnosis of the patient. Of the features, 12 features were obtained from an initial non-invasive clinical evaluation and 22 features from an additional histopathological examination of biopsied skin samples under a microscope. Data was de-identified before it was published on the repository.

The age feature is linearly valued and indicates the age of the patient in years. If any of the six skin diseases have previously been found in the patient's family, the family history feature takes a value of 1, and 0 otherwise. All other clinical and histopathological features were subjectively assessed in the range of 0 to 3. A values of 0 indicated the feature was not present, 3 indicated the largest amount possible for the feature, and 1 and 2 indicated relative intermediate values. All patients received only one of the six possible diagnoses.

Class Code	Class Name	Instances in Data
1	Psoriasis	112
2	Seboreic dermatitis	61
3	Lichen planus	72
4	Pityriasis rosea	49
5	Chronic dermatitis	52
6	Pityriasis rubra pilaris	20

Table 1. Erythemato-squamous disease classes.

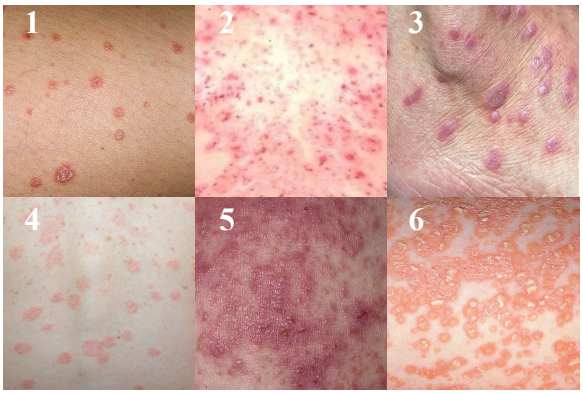

Figure 1. Erythemato-squamous disease skin visuals.

Data Processing

The data set contained eight missing values in the Age attribute, distinguished with a '?'. To reduce unwanted issues introduced by the absence of these data, the eight training examples for which this issue occurred were omitted, leaving 358 examples.

Methods

Below is a description of the workflow that was followed in building the differential diagnosis classification algorithm.

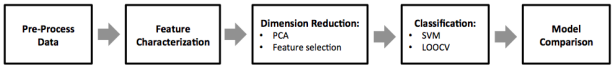

Figure 2. Flowchart showing high-level process of model development.

1. Preprocess data (clean up and normalize dataset)

2. Using LOOCV: (1) Optimize kernel selection comparing linear, polynomial, Gaussian, and logistic kernels with respective parameter optimization; (2) Optimize model performance, assessing impact of PCA, feature selection, and l_1-regularization

3. Repeat for feature subsets: non-invasive clinical, invasive, combined

4. Compare performance of models

Results

Feature impact on model performance*	Feature Description
3.91%	Itching (4)
3.07%	PNL Infiltrate (14)
2.51%	Koebner Phenomenon (5)
1.40%	Spongiosis (28)
.84%	Papillary Dermis Fibrosis (15)
.56%	Eosinophils in Infiltrate (13)
.28%	Scaling (2)
...	
-.28%	Definite borders (3)
-.28%	Acanthosis (17)
-.28%	Inflammatory monoluclear infiltrate (32)
-6.42%	Age (34)

Table 2. Feature impact on model performance (descending order). Clinical and histological features indicated by blue and green, respectively. Top box: features whose inclusion positively benefits the model; bottom box: features whose inclusion is detrimental to the model.

In our approach, our optimal classifier incorporated an SVM (trained using LibSVM[4]) with a Guassian kernel ($\gamma = 2^{-8}$) combined with principal component analysis (PCA), achieving a maximum classification accuracy of 97.8%, with average sensitivity, specificity, PPV, and NPV of 97.6%, 99.6%, 97.5% and 99.6% respectively (see Table 3).

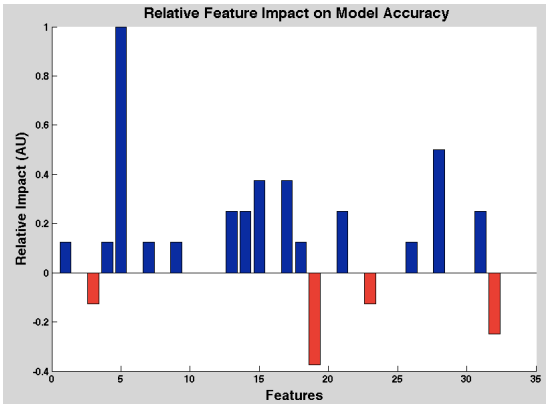

Figure 3. Relative feature impact on model accuracy. Feature numbers represent indices to features in the data set.

Table 2 and Figure 3 illustrate the results of feature selection, stratifying features by their relative contributions to the overall model accuracy. Of the total 34 features, 14 "positive-impact" features and 4 "negative-impact" features were identified.

* Error contribution of individual features from sequential feature selection analysis.

To improve the SVM model performance, we compared the impact of four kernels classes (polynomial, linear, sigmoid, Gaussian), optimizing respective parameters for each class. Optimal performance was achieved using a Gaussian kernel ($\gamma=2^{-8}$) (Figure 5), boasting a classification error of 2.23%.

Feature reduction techniques, including feature selection and PCA were compared; PCA feature reduction produced superior model performance, demonstrated in Figure 5.

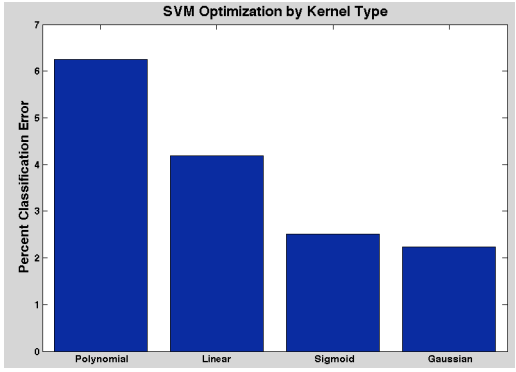

Figure 4. SVM Optimization by Kernel Type. Respective kernel parameters were optimized iteratively. Optimal values: Polynomial($\gamma=2^{-4}$,d=2), Linear, Sigmoid($\gamma=2^{-5}$), Gaussian($\gamma=2^{-8}$).

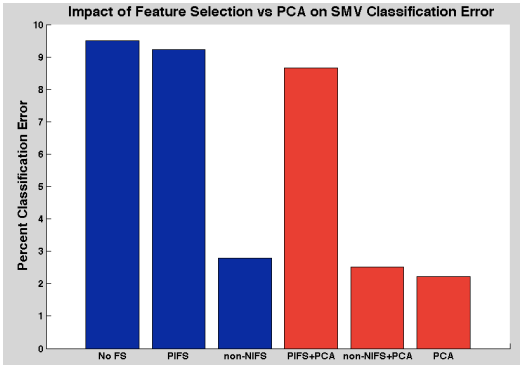

Figure 5. Impact of Feature Selection vs. PCA on SMV LOOCV Error. FS=Feature Selection, PIFS=Positive-impact FS, non-NIFS=non-negative-impact FS.

Figure 6 compares model performance when trained using one of four feature subsets: a non-invasive subset (comprised of 11 features), an invasive subset (22 features), a "positive-impact" subset (14 features), and a "non-negative-impact" subset (30 features) which were identified through feature selection. The non-invasive and invasive feature sets demonstrated 15.9% and 7.8%

classification error, respectively, when trained using LOOCV.

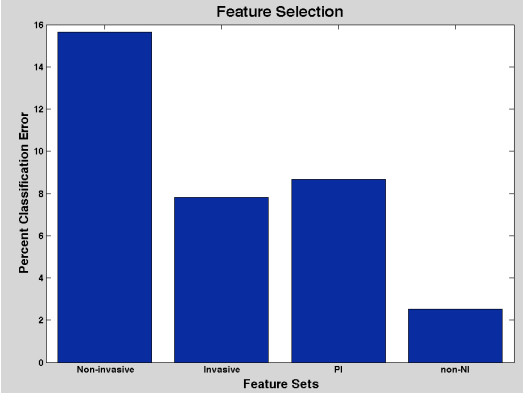

Figure 6. LOOCV classification error of models trained on total feature subsets: Non-invasive clinical features (1-11,34); invasive/biopsy-derived features (12-33); PI=positive-impact (features with positive impact on model error, positive-blue bars in figure 3), non-NI=non-negative-impact (features with non-negative impact, excluding negative-red bars in figure 3).

Feature Subset	Sens.	Spec.	PPV	NPV
All	97.6	99.6	97.5	99.6
Non-invasive	81.6	96.8	83.2	96.7
Invasive	90.6	98.5	91.5	98.5
Non-negative-impact	97.3	99.5	97.2	99.5

Table 3. Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of various feature subsets tested on the SVM model.

Discussion

CADx tools built using machine learning principles can assist physicians in more accurately diagnosing conditions that present with similar symptoms and features. In particular, if such tools can improve diagnostic accuracy using non-invasive features alone, the need for invasive biopsy procedures can be significantly reduced.

Previous approaches to machine-assisted classification of erythemato-squamous disease have included an SVM with a novel F-score feature selection approach3 and a voting feature interval (VFI5) classifier2. Classification accuracies greater than 98% were achieved in each of these studies. Although these classifiers have been built, no attempts have been made to assess and rank the relative contributions of the non-invasive and invasive feature subsets to the overall classification accuracy.

Through feature selection we identified non-invasive features, specifically "itching" and the presence of the "Koebner phenomenon", that had a larger impact on the predictive accuracy of our classifier compared to the majority of invasive features. Additionally, invasive features such as the presence of "acanthosis" and the degree of "inflammatory mononuclear infiltrates" were found to be detrimental to the accuracy of the model. Unexpectedly, age was shown to increase model error by 6.42% when included as a feature. While not only providing physicians with better insight into the disease process, these results suggest that fewer histological tests can be performed in the laboratory while still yielding comparable diagnostic accuracies and concurrently reduce costs of extraneous tests and ultimately the number of invasive procedures.

Our model, an optimized Gaussian-SVM, when combined with PCA demonstrated a classification performance of 97.8%. This model subsequently served as a basis to address the impact of training only on non-invasive clinical features. While model accuracy decreased with this feature reduction (classification accuracy of 84.1%), an 81.56% sensitivity and 96.78% specificity were still achieved. The exclusion of a disease can be just as effective as an absolute diagnosis; the utility of an absolute diagnosis is only valuable to a patient if it will result in a change in the therapeutic approach, significantly alter the patient's overall prognosis, or influence a significant change in behavior. The specificity achieved by our feature-reduced model could function as a screening tool to assist physicians in better risk-stratifying patients and tailoring subsequent diagnostic work-up. In certain more severe cases where stringency is desired, non-invasive features can serve to pre-screen patients for potential diagnoses and assess whether an invasive biopsy is necessary for confirmation.

While such pre-screening can be beneficial, determining appropriate thresholds to use can be challenging and context-dependent. Such thresholds are influenced by several factors such as the downstream benefits of a definitive diagnosis and the cost of additional workup. Further research must be done to assess the threshold at which a biopsy would be justified in the case of the erythemato-squamous diseases. An additional consideration for future work is to cluster diseases based on severity and procedural costs before running the diagnostic model. In our own ongoing research, we intend to use publicly available medical billing data to identify procedural costs associated with various features used in this model. With this added prognostic and economic context, we believe an optimal balance can be identified between maximizing predictive accuracy of the model and minimizing both cost and excessive intervention associated with feature collection and follow-up treatment, given these pre-specified objectives.

Summary

In this work we developed a classification algorithm to assist in the differential diagnosis of the erythemato-squamous diseases. The model was developed using previous dermatological data that included both clinical and histopathological features. The final embodiment of the model was an optimized SVM with a Guassian kernel ($\gamma = 2^{-8}$) combined with principal component analysis (PCA), achieving a maximum classification accuracy of 97.8%. After performing feature selection on clinical and histological subsets of the data, we found features 4, 14, and 5 were most important to model accuracy while feature 34 (age) had the greatest confounding effect on the model's accuracy. Using only non-invasive clinical features our model's performance decreased to a classification accuracy of 84.4%. Despite the decrease this result suggests the utility of such a model in minimizing number of biopsies performed. Future work will focus on developing reasonable clinical thresholds for biopsy and diagnosis decisions using additional considerations such as procedure costs.

References

[1] UC Irvine Machine Learning Repository. Dermatology Data Set. http://archive.ics.uci.edu /ml/datasets/Dermatology.

[2] H. A. Güvenir, G. Demiröz, N. İlter. "Learning differential diagnosis of erythemato-squamous diseases using voting feature intervals". Machine Learning in Medicine, 1998.

[3] J. Xie, C. Wang. "Using Support vector machines with a novel hybrid feature selection method for diagnosis of erythemato-squamous diseases." Expert Systems with Applications, 2010.

[4] LIBSVM: A Library for Support Vector Machines. http://www.csie.ntu.edu.tw/~ cjlin/libsvm/.

[5] E. Maglogiannis, C. Caroni, S. Pavlopoulos, V. Karioti. "Utilizing Artificial Intelligence for the Characterization of Dermatological Images." NNESMED, 2001.

[6] Nanni, L. (2006). An ensemble of classifiers for the diagnosis of erythemato-squamous diseases. Neurocomputing, 69, 842–845.

[7] H. A., & Emeksiz, N. (2000). An expert system for the differential diagnosis of erythemato-squamous diseases. Expert Systems with Applications, 18, 43–49.

[8] Karabatak, M., & Ince M. C. (2009). A new feature selection method based on association rules for diagnosis of erythemato-squamous diseases. Expert Systems with Applications, 36, 12500–12505.

Appendix

Clinical Features (non-invasive): erythema, scaling, definite borders, itching, koebner phenomenon, polygonal papules, follicular papules, oral mucosal involvement, knee and elbow involvement, scalp involvement, family history, age

Histopathological features (invasive): melanin incontinence, eosinophils in the infiltrate, PNL infiltrate, fibrosis of the papillary dermis, exocytosis, acanthosis, hyperkeratosis, parakeratosis, clubbing of the rete ridges, elongation of the rete ridges, thinning of the suprapapillary epidermis, spongiform pustule, munro microabcess, focal hypergranulosis, disappearance of the granular layer, vacuolisation and damage of basal layer, spongiosis, saw-tooth appearance of retes, follicular horn plug, perifollicular parakeratosis, inflammatory monoluclear infiltrate, 33: band-like infiltrate