

---

Paper ID: 1570940595

Paper Title: Feature Selection in the Classification of Erythematous-Squamous Diseases using Machine Learning Models and Principal Component Analysis

Authors: Thumpussorn Akarajarasroj, Ohmthong Wattanapermpool, Piyaon Sapphaphab and Oraya Rinthon (Satriwithaya School, Bangkok, Thailand); Suejit Pechprasarn and Pichit Boonkrong (Rangsit University, Thailand)

Email: 48904@satriwit.ac.th

---

## Abstract

Erythematous-squamous diseases (ESD) are dermatological diseases that significantly impact the quality of life of an increasing number of patients worldwide. This study used a publicly available clinical dataset of 366 patients from the Department of Computer Engineering and Information Science, Bilkent University with 34 predictors contributing to the classification of ESDs. The data was curated to ensure unbiasedness and accuracy before applying principal component analysis and machine learning models to identify crucial factors in classifying the six main types of ESDs. 31 different machine learning models, including Tree, Linear Discriminant, Quadratic Discriminant, Naïve Bayes, SVM, KNN, Ensemble, Neural Network, and Kernel were trained, validated, and the classification accuracy was compared. The model that is the most adequate is the Fine KNN, which has the highest cross-validation classification accuracy at 100%. This model requires only eight predictors: itching, the Koebner phenomenon, follicular papules, fibrosis of the papillary dermis, spongiosis, inflammatory mononuclear infiltrate, band-like infiltrate, and age.

---