

## Machine Learning for Feature Selection and Classification of Systemic Lupus Erythematosus

**Authors :** H. Zidoum, A. AlShareedah, S. Al Sawafi, A. Al-Ansari, B. Al Lawati

**Abstract :** Systemic lupus erythematosus (SLE) is an autoimmune disease with genetic and environmental components. SLE is characterized by a wide variability of clinical manifestations and a course frequently subject to unpredictable flares. Despite recent progress in classification tools, the early diagnosis of SLE is still an unmet need for many patients. This study proposes an interpretable disease classification model that combines the high and efficient predictive performance of CatBoost and the model-agnostic interpretation tools of Shapley Additive exPlanations (SHAP). The CatBoost model was trained on a local cohort of 219 Omani patients with SLE as well as other control diseases. Furthermore, the SHAP library was used to generate individual explanations of the model's decisions as well as rank clinical features by contribution. Overall, we achieved an AUC score of 0.945, F1-score of 0.92 and identified four clinical features (alopecia, renal disorders, cutaneous lupus, and hemolytic anemia) along with the patient's age that was shown to have the greatest contribution on the prediction.

**Keywords :** feature selection, classification, systemic lupus erythematosus, model interpretation, SHAP, Catboost

**Conference Title :** ICBAHS 2023 : International Conference on Biomedical and Health Sciences

**Conference Location :** Barcelona, Spain

**Conference Dates :** May 22-23, 2023