

proteinuria level. Three additional features were important for two models—whether the patient is taking prednisolone; time adjusted mean (TAM) SLEDAI score; and SLEDAI score.

Conclusions The study showed and compared various machine learning models on their predictive power in determining whether a patient will achieve LLDAS on their next visit. The results determined that the current LLDAS, proteinuria levels, SLEDAI score (and TAM SLEDAI),

LSO-081 GENOMIC PREDICTION MODEL USING MACHINE LEARNING TECHNIQUES THAT CAN DISTINGUISH AUTOIMMUNE DISEASES (RA OR SLE) FROM HEALTHY CONTROLS

^{1,2}Young Bin Joo, ³Youngho Park, ^{1,2}So-Young Bang, ^{1,2}Sang-Cheol Bae, ^{1,2}Hye-Soon Lee*.

¹Department of rheumatology, Hanyang University Hospital for Rheumatic Diseases, Seoul, Republic of, Republic of Korea; ²Hanyang University Institute for Rheumatology Research, Seoul, Republic of Korea, Republic of Korea; ³Department of Big Data Application, Hannam University, Daejeon, Republic of Korea., Republic of Korea

10.1136/lupus-2023-KCR.123

Background Rheumatoid arthritis (RA) and systemic lupus erythematosus (SLE) are the prototypes of autoimmune diseases, for which many genetic loci have been identified using genome-wide association studies (GWAS) for recent decades. This study aims to establish genomic prediction models for autoimmune diseases using machine learning (ML) techniques, which can classify the patients with RA or SLE from healthy controls.

Methods We obtained SNPs data (N=446,097) from Korean Chip (Affymetrix Axiom™ KOR) for 3,145 RA, 1,870 SLE, and 3,635 controls, enrolled from Hanyang University Hospital for Rheumatic Diseases. After quality control (N=428,380), we selected the significant associated SNPs with each disease (RA or SLE) using univariate test ($p < 5.0e-8$)

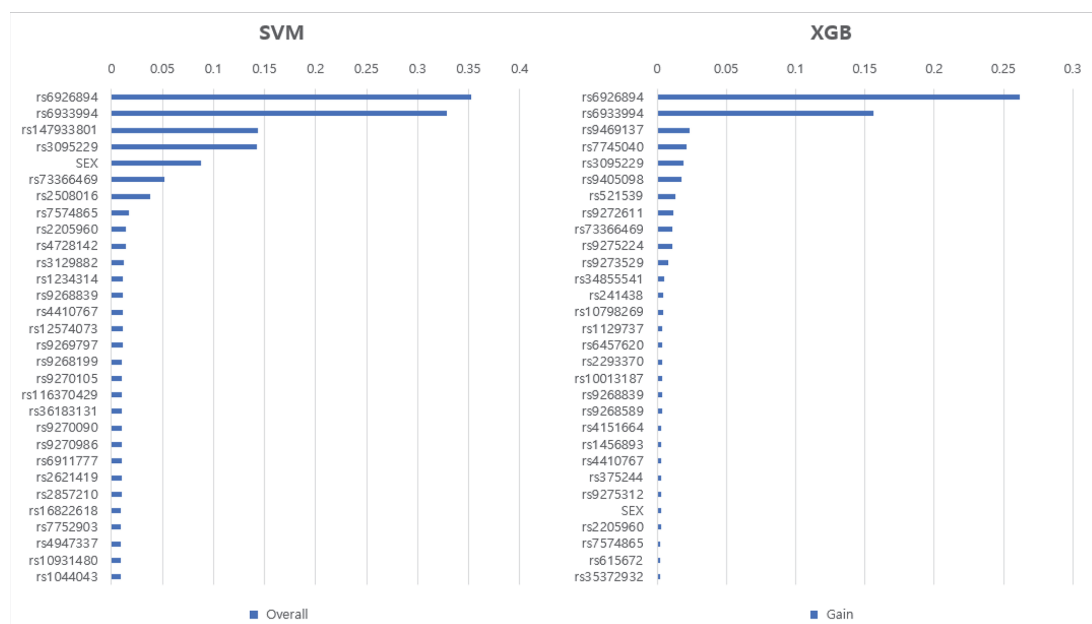
and imputed by SHAPEIT2, IMPUTE4, Eagle v2, Minimac3. We conducted ML, support vector machine (SVM) and extreme gradient boosting (XGB). Contribution of SNPs for classification was calculated with importance (or gain) values.

Results A total of 2,458 SNPs were selected and used as inputs for ML models. Each group of RA, SLE, and controls was randomly divided into training (70%) and testing (30%) subsets. Both SVM and XGB approaches showed high prediction performance to discriminate autoimmune diseases of RA or SLE from controls (table 1). However, these ML models showed lower prediction performance to discriminate RA from SLE (accuracy 0.7229 – 0.7368). Among top 30 SNPs showing high feature importance for classification in each ML model (figure 1), 8 SNPs were overlapped, which were annotated as HLA-DQA1, HLA-DOA, MICB, GTF2IRD1, HLA-DRA, TNFSF4 and STAT4.

Abstract LSO-081 Table 1 Statistics of machine learning models

	SVM	XGB
Accuracy	0.9946	0.9946
Sensitivity (Recall)	0.9898	0.9880
Specificity	0.9980	0.9993
Positive predictive value (Precision)	0.9972	0.9991
Negative predictive value	0.9928	0.9915
F1	0.9935	0.9935

Conclusions We identified ML based genomic prediction models that could distinguish between autoimmune diseases (RA or SLE) and healthy controls in the Korean population, while suggesting the need for more research for differentiating RA from SLE. These results suggest that ML approach using genomic data might be useful to predict the risk of autoimmune diseases.



Abstract LSO-081 Figure 1 Feature importance of top 30 SNPs of machine learning models.