Machine Learning Applications in Predicting the Onset of Psoriatic Arthritis: A Systematic Review

Borate SN, M Zuber, Gokhale P, Villa Zapata L

Department of Clinical and Administrative Pharmacy, College of Pharmacy, University of Georgia, Athens, GA, USA

MSR39

INTRODUCTION

- Around 25% of psoriasis patients develop psoriatic arthritis, necessitating early prediction to prevent joint damage and disability
- Machine learning models have recently emerged as significant tools in predicting complex conditions such as psoriatic arthritis
- This study focuses on evaluating and comparing the performance of various ML models to determine which models and types of predictors most effectively predict the onset of psoriatic arthritis (PsA) in psoriasis (PsO) patients

METHODS

- A systematic review of relevant databases (Pubmed, Embase, Web of Science, and Epistemonikos) was conducted from 1 January 1982 to 1 October 2023
- Studies published in the English language, reporting outcomes related to predicting PsA in PsO patients using machine learning were included
- The extraction was conducted using the CHARMS checklist and the risk of bias was assessed by the PROBAST checklist
- The discriminative ability of models was compared using the AUROC measure

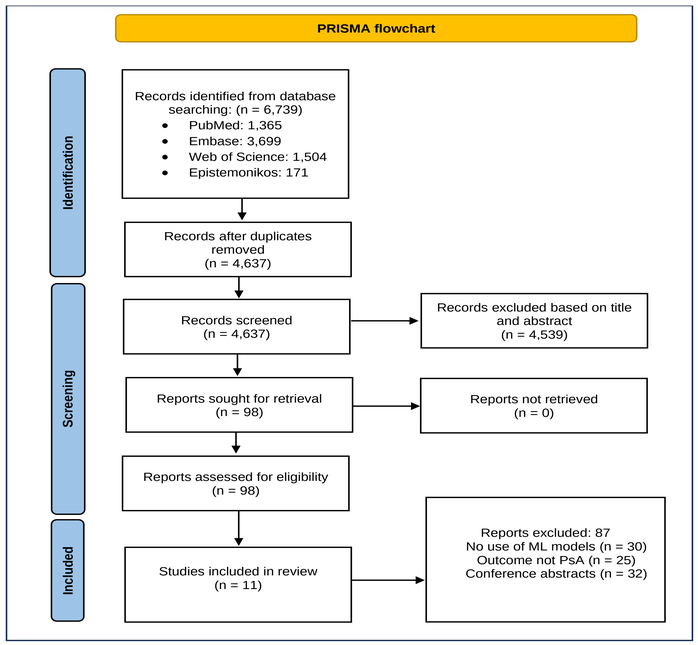

Fig. 1: PRISMA flowchart

Study	Participants	Predictors	Outcomes	Analysis	Overall ROB
Patrick 2018	--	--	--	+	+
Eder 2021	--	--	--	?	--
Farideh 2021	--	--	--	--	--
Lejla 2021	+	--	+	+	+
Mulder 2021	--	--	--	+	+
Liu 2022	--	--	--	?	--
Cruz--Correa 2023	--	--	--	--	--
Lee 2023	--	--	--	--	--
Nielsen 2023	--	--	--	--	--
Shapiro 2023	--	--	--	--	--
Xu 2023	--	--	--	--	--

Table 2: Risk of bias based on PROBAST checklist

Study	Data source	Sample size	Type of predictors	Discrimination measure	Findings
1) Patrick 2018, Global	• Genetic data: - GWAS datasets: CASP, Exomechip, Genizon, Kiel, and PsAGWAS - Immunochip dataset: PAGE	N = 34,600 • PsA: 3,674 • Skin PsO (including a majority of PsO vulgaris): 14,590 • Control: 16,336	• Molecular: HLA allele, SNP and INDEL genetic markers	• AUROC = 0.78; Random forest • AUROC = 0.82; Conditional inference forest	• Excellent model performance was achieved with 200 genetic markers
2) Eder 2021, Canada	• Genetic data: Custom multi-SNP genetic assay • Medical records: Medical academic center in Toronto, Canada	N = 328; • PsA: 78 • Skin PsO: 250	• Molecular: SNP genetic markers • Clinical: Age, sex, race, PsO duration, a history of uveitis, nail psoriasis, flexural psoriasis, PASI, BMI, patient pain, HAQ-DI, FACIT, hsCRP and family history of PsA	-For genetic markers: • AUROC = 0.60; Naive bayes • AUROC = 0.56; Logistic regression • AUROC = 0.55; Random forest -For genetic markers and clinical variables: • AUROC = 0.69; Logistic regression • AUROC = 0.68; Random forest • AUROC = 0.65; Naive bayes	• The combination of genetic and clinical data resulted in an improvement in the performance of the models • The variables that contributed most information to the model were rs146571698 (TNIP1), HAQ and hsCRP)
3) Farideh 2021, UK	• Genetic data: - Skin PsO patients: BSTOP dataset from BADBIR registry - PsA patients: Rheumatology centres within the UK	N = 2,594 • PsA: 1,462 • Skin PsO: 1,132	• Molecular: HLA alleles genetic markers	- The single most informative genetic features identified after applying feature selection • AUROC = 0.54 in internal cross-validation, • AUROC = 0.53 in the internal hold-out set, • AUROC = 0.55 in an external dataset (UK Biobank) - Improved model performance with more features (Random Forest): • AUROC = 0.61 in internal cross-validation, • AUROC = 0.57 in the internal hold-out set, • AUROC = 0.58 in the external dataset (UK Biobank)	• The most informative genetic variant was HLA_C_*06 before controlling for confounding factors • After controlling for confounding factors, HLA_B_*27 was identified as the single most important genetic feature • Sequentially adding additional HLA features based on rank improved the performance of the Random Forest classification model leading to higher AUC
4) Lejla 2021, Bosnia and Herzegovina	• Clinical data - Skin PsO patients: A dataset created for the purpose of the study	N = 450	• Clinical: Age, sex, smoking status, alcohol consumption, nail psoriasis, PASI, BMI, and family history of PsA	NR	• The Random Forest model demonstrated the highest accuracy and validation performance
5) Mulder 2021, Netherlands	• Immunological and clinical data -PsA and skin PsO patients: Radboud University Medical Center, Netherlands	N = 86 • PsA: 45 • Skin PsO: 41	• Molecular: immune cell subsets	• AUROC = 0.95; Random forest	• The most relevant cell subsets contributing to the classification of PsA or Pso were memory T-cells, Th17-like cells, Th2-like cells and Treg cells • The Random forest classification model capable of distinguishing PsA from Pso achieved outstanding performcance
6) Liu 2022, United States	• Genetic, immunological and clinical data: dermatology clinics at UCSF	N = 95 • PsA: 28 • Skin PsO: 24 • Healthy controls: 29 • PsO patients with unclear PsA diagnosis: 14	• Molecular: DEGs and DEPs	-For DEGs • AUROC = 1; Random forest, SVMRadial, and NNET -For DEPs • AUROC = 0.96; Random forest • AUROC = 0.95; SVMRadial • AUROC = 0.96; NNET -For combined feature • AUROC = 1; Random forest • AUROC = 0.96; SVMRadial • AUROC = 0.7; NNET	• The models showed higher performance with DEGs compared to DEPs, with some models achieving outstanding performance using DEGs • Combining DEPs and DEGs led to intermediate results
7) Cruz-Correa 2023, Canada	• Genetic and clinical data: University of Toronto Psoriatic Disease Program	N = 117; • Converters to PsA from PsO: 58 • Non-converters to PsA from PsO: 59	• Molecular: DNA methylation markers • Clinical: Age, sex, BMI, smoking status, nail psoriasis, PsO duration, plague psoriasis, use of NSAIDs and DMARDs, and family history of PsO	-All methylation markers • AUROC: 0.9467; SVMRadial -FDR adjusted significant methylation markers • AUROC: 0.9867; SVMRadial -Most relevant methylation markers • AUROC: 0.9644; SVMRadial	• The model relying on the DNA methylation markers identified converters and nonconverters with outstanding performance
8) Lee 2023, Taiwan	• Clinical data: Taiwan's National Health Insurance Research Database	N = 2,215 • PsA: 443 • PsO: 1,772	• Clinical: Age, sex, comorbidities (autoimmune connective tissue diseases, rheumatic disease, obesity and metabolic syndrome, cardiovascular disorders), drugs (DMARDs, corticosteroids, retinoids, immunosuppressants)	• AUROC: 0.70; CNN model	• This 6-month risk prediction model used sequential diagnostic and drug prescription information yielded a model with good performance
9) Nielsen 2023, Switzerland	• Clinical data: Swiss Dermatology Network on Targeted Therapies	N = 864 • PsA at Beginning: 257 • PsA Developed after Inclusion: 91 • No PsA: 516	• Clinical: Age, sex, weight, BMI, age at diagnosis, age at prescription, PASI, DLQI, EQ-5D, BSA, number of previous treatment series, number of previous biologics	-For model with number of previous treatments • AUROC: 0.722; Gradient Boosted Decision Trees and mixed models -For model without number of previous treatments • AUROC: 0.737; Gradient Boosted Decision Trees and mixed models	• The findings suggest that ML models can effectively differentiate between patients at risk for PsA and those not likely to develop the condition, indicating good model performance
10) Shapiro 2023, Israel	• Clinical data: Maccabi Healthcare Services	-PsO cohort: N = 39,509 • PsA: 1,528 • PsO: 37,825 -GP cohort: N = 7,60,709 • PsA: 2,096 • PsO: 42,457 • Patients without PsO: 7,16,156	• Clinical data: Age, sex, BMI, laboratory results (ALKP, CRP, ESR, HDL), physician encounters, specialist or hospital visits and referrals (any and X ray), arthritis diagnosis, prescribed and filled medications, specific anti-psoriasis medications (topical and systemic), family history of PsA	- PsO cohort at 90 and 95% specificity cutoff • AUROC at 1 year gap: 0.79; Gradient Boosted Decision Trees • Subsequent years (2, 3 and 4 years gap): 0.76, 0.73, 0.72 - GP cohort at 99.0 and 99.9% specificity cutoff • AUROC at 1 year gap: 0.9; Gradient Boosted Decision Trees 2 years gap: 0.87 3 years gap: 0.86 4 years gap: 0.84	• The model to predict undiagnosed PsA in patients, achieved good and outstanding performance with gradient boosted decision trees
11) Xu 2023, China	• Clinical and genetic data: Xiangya Hospital of Central South University	N = 3,961 • PsA: 265 • Non-PsA: 3,696	• Molecular: HLA-B27 genetic markers • Clinical: Age, sex, height, weight, waist-hip ratio, BMI, SBP, DBP, PASI, BSA, DLQI, laboratory tests (UrineRT, BloodrRT, lipid, ESR) and drugs (Acitretin, MTX, corticosteroids, IL-17 inhibitors, IL-23 inhibitors, TNF-α inhibitors)	• AUROC: 0.87 • AUPR: 0.89	• The model with a combination of genetic and clinical data achieved excellent performance

Table 1: Characteristics of included studies

RESULTS

- Five studies described various models achieving good discrimination performance (AUROC ≥ 0.7 to <0.8), three studies with excellent (AUROC ≥ 0.8 to <0.9), and four with outstanding (AUROC ≥ 0.9) performance
- Two studies described all models with poor discrimination
- Only one study conducted external validation of their proposed model
- Overall ROB was found to be low for eight studies, and high for three studies

CONCLUSION

- Although a majority of models displayed good discriminative ability, the necessity for external validation is underscored to enhance the clinical utility of these models

REFERENCES

1) Alinaghi F, Calov M, Kristensen LE, et al. Prevalence of psoriatic arthritis in patients with psoriasis: A systematic review and meta-analysis of observational and clinical studies. J Am Acad Dermatol. 2019;80:251-265.e19.
2) Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, Reitsma JB, Kleijnen J, Mallett S; PROBAST Group. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. Ann Intern Med. 2019 Jan 1;170(1):51-58. doi: 10.7326/M18-1376. PMID: 30596875.
3) Mandrekar JN. Receiver Operating Characteristic Curve in Diagnostic Test Assessment. Journal of Thoracic Oncology 2010;5(9):P1315-1316