ANISHA RAJ

(22051837)

RESEARCH PAPER-1

We developed and independently validated a rheumatoid arthritis (RA) mortality prediction model using the machine learning method Random Survival Forests (RSF). Two independent cohorts from Madrid (Spain) were used: the Hospital Clínico San Carlos RA Cohort (HCSC-RAC; training; 1,461 patients), and the Hospital Universitario de La Princesa Early Arthritis Register Longitudinal study (PEARL; validation; 280 patients). Demographic and clinical-related variables collected during the frst two years after disease diagnosis were used. 148 and 21 patients from HCSC-RAC and PEARL died during a median follow-up time of 4.3 and 5.0 years, respectively. Age at diagnosis, median erythrocyte sedimentation rate, and number of hospital admissions showed the higher predictive capacity. Prediction errors in the training and validation cohorts were 0.187 and 0.233, respectively. A survival tree identifed fve mortality risk groups using the predicted ensemble mortality. After 1 and 7 years of follow-up, time-dependent specifcity and sensitivity in the validation cohort were 0.79–0.80 and 0.43–0.48, respectively, using the cut-of value dividing the two lower risk categories.

Modeling Method	Random Survival Forests (RSF)
Training Cohort	HCSC-RAC (Hospital Clínico San Carlos RA Cohort)
Training Cohort Size	1,461 patients
Validation Cohort	PEARL (Hospital Universitario de La Princesa Early Arthritis Register)
Validation Cohort Size	280 patients
Outcome	Mortality Prediction in Rheumatoid Arthritis
Follow-Up (Median)	HCSC-RAC: 4.3 years; PEARL: 5.0 years
Number of Deaths	HCSC-RAC: 148 patients; PEARL: 21 patients
Key Predictors	1. Age at diagnosis 2. Median erythrocyte sedimentation rate 3. Number of hospital admissions
Prediction Errors	Training: 0.187; Validation: 0.233
Time-Dependent Sensitivity	1 year: 0.43; 7 years: 0.48
Time-Dependent Specificity	1 year: 0.79; 7 years: 0.80
Risk Groups Identified	5 mortality risk groups using survival tree