**Title: Machine learning for prediction and risk stratification of lupus nephritis renal flare**

**Supplementary Material**

**Supplementary figures: Figure S1-S3**

**Supplementary tables: Table S1-S5**

**Table S1. Demographic and clinical characteristics and laboratory data concerning patients with LN at the point of remission**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Feature** | **All** | **Non-relapse patients** | **Relapse patients** | **P-value** |
| Female, n (%) | 1482 (87.5) | 922 (89.4) | 560 (84.5) | 0.003 |
| Age (year) | 32.4 (9.9) | 33.7 (9.8) | 30.3 (9.6) | <0.001 |
| Time to remission (month) | 12.1 (12.3) | 12.6 (12.9) | 11.3 (11.4) | 0.007 |
| Partial remission, n (%) | 389 (23.0) | 108 (10.5) | 281 (42.4) | <0.001 |
| UPro (g/24 hours) | 0.5 (0.5) | 0.3 (0.4) | 0.6 (0.6) | <0.001 |
| Urine sediment red blood cell count (×104/ml) | 12.4 (43.5) | 6.3 (22.1) | 21.8 (62.7) | <0.001 |
| SCr (mg/dl) | 0.8 (0.3) | 0.8 (0.3) | 0.81 (0.4) | <0.001 |
| Serum UA (μmol/l) | 336.7 (109.6) | 319.1 (100.3) | 365.4 (117.7) | <0.001 |
| Serum Alb (g/l) | 41.6 (4.6) | 42.4 (4.2) | 40.5 (4.9) | <0.001 |
| Serum globulin (g/l) | 27.2 (4.9) | 27.4 (4.6) | 26.9 (5.3) | 0.004 |
| Hb (g/l) | 121.5 (16.6) | 122.8 (16.3) | 119.3 (16.9) | <0.001 |
| WBC count (/ul) | 6739.1 (2519.2) | 6764.9 (2519.8) | 6697.8 (2519.7) | 0.532 |
| Anti-dsDNA positive, n (%) | 324 (19.1) | 158 (15.3) | 166 (25.0) | <0.001 |
| ANA positive, n (%) | 1433 (84.6) | 878 (85.2) | 555 (83.7) | 0.847 |
| Serum C3 (g/l) | 0.8 (0.3) | 0.9 (0.3) | 0.8 (0.3) | <0.001 |
| Serum C4 (g/l) | 0.2 (0.1) | 0.2(0.1) | 0.2 (0.1) | 0.127 |

**Abbreviations:** Alb, albumin; ANA, antinuclear antibodies; Hb, hemoglobin; SCr, serum creatinine; UA, uric acid; UPro, urinary protein, WBC, white blood cell

Data are summarized as mean ± standard deviation or n (%).

**Table S2. The duration of disease and pathologic data of patients with LN at baseline**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Feature** | **All** | **Non-relapse patients** | **Relapse patients** | **P-value** |
| Duration of SLE (month) | 40.8 (50.4) | 40.0 (50.9) | 41.2 (49.8) | 0.226 |
| Duration of LN (month) | 20.6 (33.7) | 20.1 (34.9) | 21.4 (31.8) | 0.013 |
| SLE-DAI | 12.4 (5.2) | 12.2 (5.2) | 12.7 (5.2) | 0.074 |
| Activity index | 6.7 (4.6) | 6.2 (4.6) | 7.4 (4.5) | <0.001 |
| Chronicity index | 2.1 (1.8) | 2.0 (1.8) | 2.2 (1.9) | 0.037 |
| Proportion of global sclerosis (%) | 5.6 (10.8) | 5.0 (9.8) | 6.4 (12.2) | 0.124 |
| Proportion of segmental sclerosis (%) | 2.1 (5.6) | 1.8 (5.1) | 2.5 (6.4) | 0.187 |
| Proportion of cellular crescents (%) | 11.8 (15.9) | 10.9 (15.6) | 13.3 (16.3) | <0.001 |
| Endocapillary hypercellularity, n (%) | 1188 (70.1) | 677 (65.7) | 511 (77.1) | <0.001 |
| Neutrophils/karyorrhexis, n (%) | 914 (54.0) | 512 (49.7) | 402 (60.6) | <0.001 |
| Fibrinoid necrosis, n (%) | 738 (43.6) | 414 (40.2) | 324 (48.9) | 0.001 |
| Subendothelial immune deposits/Hyaline deposits, n (%) | 769 (45.4) | 425 (41.2) | 344 (51.9) | <0.001 |
| Subepithelial immune deposits, n (%) | 1236 (73.0) | 755 (73.2) | 481 (72.6) | 1.000 |
| Interstitial inflammation, n (%) | 174 (16.9) | 148 (22.3) | 322 (19.0) | 0.006 |
| Acute tubulointerstitial injury, n (%) | 1068 (63.1) | 624 (60.5) | 444 (67.0) | 0.009 |
| Tubular atrophy, n (%) | 1026 (60.6) | 602 (58.4) | 424 (64.0) | 0.025 |
| Interstitial fibrosis, n (%) | 1024 (60.5) | 602 (58.4) | 422 (63.7) | 0.035 |
| Intensity of IgG staining |  |  |  | 0.047 |
| Negative, n (%) | 35 (2.1) | 19 (1.8) | 16 (2.4) |  |
| 1+, n (%) | 340 (20.1) | 188 (18.2) | 152 (22.9) |  |
| 2+, n (%) | 1242 (73.3) | 774 (75.1) | 468 (70.6) |  |
| Intensity of IgM staining |  |  |  | 0.004 |
| Negative, n (%) | 78 (4.6) | 53 (5.1) | 25 (3.8) |  |
| 1+, n(%) | 1111 (65.6) | 695 (67.4) | 416 (62.8) |  |
| 2+, n(%) | 420 (24.8) | 227 (22.0) | 193 (29.1) |  |
| Intensity of IgA staining |  |  |  | 0.819 |
| Negative, n (%) | 85 (5.0) | 50 (4.9) | 35 (5.3) |  |
| 1+, n (%) | 1126 (66.5) | 688 (66.7) | 438 (66.1) |  |
| 2+, n (%) | 406 (24.0) | 242 (23.5) | 164 (24.7) |  |
| Intensity of C3 staining |  |  |  | 0.047 |
| Negative, n (%) | 20 (1.2) | 8 (0.8) | 12 (1.8) |  |
| 1+, n (%) | 238 (14.1) | 156 (15.1) | 82 (12.4) |  |
| 2+, n (%) | 1360 (80.3) | 818 (79.3) | 542 (81.8) |  |

**Table S2 continue**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Feature** | **All** | **Non-relapse patients** | **Relapse patients** | **P-value** |
| Intensity of C1q staining |  |  |  | 0.204 |
| Negative, n (%) | 43 (2.5) | 29 (2.8) | 14 (2.1) |  |
| 1+, n (%) | 789 (46.6) | 492 (47.7) | 297 (44.8) |  |
| 2+, n (%) | 785 (46.3) | 460 (44.6) | 325 (49.0) |  |
| GCW deposition of IgG, n (%) | 1562 (92.2) | 949 (92.1) | 613 (92.5) | 0.729 |
| GCW deposition of IgM, n (%) | 1509 (89.1) | 908 (88.1) | 601 (90.7) | 0.187 |
| GCW deposition of IgA, n (%) | 1506 (88.9) | 914 (88.7) | 592 (89.3) | 0.962 |
| GCW deposition of C3, n (%) | 1573 (92.9) | 956 (92.7) | 617 (93.1) | 0.630 |
| GCW deposition of C1q, n (%) | 1550 (91.5) | 934 (90.6) | 616 (92.9) | 0.176 |
| Mesangial deposition of IgG, n (%) | 1227 (72.4) | 729 (70.7) | 498 (75.1) | 0.084 |
| Mesangial deposition of IgM, n (%) | 1237 (73.0) | 734 (71.2) | 503 (75.9) | 0.062 |
| Mesangial deposition of IgA, n (%) | 1210 (71.4) | 723 (70.1) | 487 (73.5) | 0.230 |
| Mesangial deposition of C3, n (%) | 1264 (74.6) | 753 (73.0) | 511 (77.1) | 0.110 |
| Mesangial deposition of C1q, n (%) | 1232 (72.7) | 735 (71.3) | 497 (75.0) | 0.167 |
| TBM deposition of IgG, n (%) | 94 (5.6) | 56 (5.4) | 38 (5.7) | 0.913 |
| TBM deposition of IgM, n (%) | 17 (1) | 9 (0.87) | 8 (1.21) | 0.686 |
| TBM deposition of IgA, n(%) | 7 (0.4) | 3 (0.3) | 4 (0.6) | 0.564 |
| TBM deposition of C3, n(%) | 238 (14.1) | 126 (12.2) | 112 (16.9) | 0.010 |
| TBM deposition of C1q, n(%) | 175 (10.3) | 94 (9.1) | 81 (12.2) | 0.055 |

Data are summarized as mean ± standard deviation or n (%)

**Abbreviations**: GCW, glomerular capillary wall; LN, lupus nephritis; SLE-DAI, systemic lupus erythematosus disease activity index; TBM, tubular basement membrane

**Table S3. Induction treatment for patients with LN**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Feature** | **All\*** | **Non-relapse patients** | **Relapse patients** | **P-value** |
| Glucocorticoids plus IV-CYC, n (%) | 411 (24.3%) | 255 (24.7%) | 156 (23.5%) | 0.613 |
| Glucocorticoids plus MMF, n (%) | 317 (18.7%) | 177 (17.2%) | 140 (21.1%) | 0.049 |
| Combination therapy, n (%) | 253 (14.9%) | 156 (15.1%) | 97 (14.6%) | 0.832 |
| Glucocorticoids plus CNI, n (%) | 142 (8.4%) | 73 (7.1%) | 69 (10.4%) | 0.020 |
| Glucocorticoids alone, n (%) | 571 (33.7%) | 370 (35.9%) | 201 (30.3%) | 0.021 |
| Intravenous MP pulse therapy, n (%) | 931 (55.0%) | 572 (55.5%) | 359 (54.2%) | 0.626 |

Data are summarized as n (%).

**Abbreviations:** CNI, calcineurin inhibitor; Combination therapy, combining glucocorticoids, MMF, and tacrolimus; IV-CYC, intravenous cyclophosphamide pulse; LN, lupus nephritis; MMF, mycophenolate mofetil; MP, methylprednisolone

**Table S4. Details of hyper-parameters in the XGBoost classifier**

|  |  |  |
| --- | --- | --- |
| **Parameters** | **Explanation** | **Values** |
| Eta | Boosting learning rate | 0.16 |
| Gamma | Minimum loss reduction required to make a further partition on a leaf node of the tree | 12 |
| max\_depth | Maximum tree depth | 1000 |
| min\_child\_weight | Minimum sum of instance weight needed in a child | 20 |
| subsample | Subsample ratio of the training instance | 0.5 |
| colsample\_bytree | Subsample ratio of columns when constructing each tree | 0.7 |
| Booster | Specify which booster to use: gbtree or gblinear | ‘gbtree’ |
| objective | Specify the learning task and the corresponding learning objective | ‘survival:cox’ |
| nrounds | Number of boosted trees to fit | 1000 |

**Table S5. A summary of the stepwise Cox regression model based on variables selected using the XGBoost model**

|  |  |  |  |
| --- | --- | --- | --- |
| **Feature** | **Beta** | **HR (95% CI)** | **P-value** |
| Partial remission | 1.342 | 3.826 (3.055–4.791) | <0.001 |
| Age at the point of remission (year) | -0.030 | 0.970 (0.960–0.980) | <0.001 |
| Serum Alb at the point of remission (g/l) | -0.027 | 0.973 (0.953–0.995) | 0.014 |
| anti-dsDNA positive at the point of remission | 0.256 | 1.292 (1.038–1.608) | 0.022 |
| Serum C3 at the point of remission (g/l) | -0.561 | 0.571 (0.390–0.836) | 0.004 |
| Endocapillary hypercellularity | 0.262 | 1.300 (1.039–1.625) | 0.022 |

**Abbreviations:** Alb, albumin; CI, confidence interval; HR, hazard ratio



**Figure S1. Feature contribution of age at the point of remission to the risk prediction.**

Each dot represents the SHAP value calculated for each patient. The x-axis is the value of age at the point of remission, and the y-axis is the SHAP value assigned to their age. A higher SHAP value indicates that the risk of relapse attributed to age at the point of remission rises.

Abbreviation: SHAP, Shapley Additive Explanations

****

**Figure S2. Feature contribution of serum Alb at the point of remission to the risk prediction.**

Each dot represents the SHAP value calculated for each patient. The x-axis is the value of serum Alb, and the y-axis is the SHAP value assigned to their measures of serum Alb. A higher SHAP value indicates that the risk of relapse attributed to serum Alb increases.

**Abbreviations:** Alb, albumin; SHAP, Shapley Additive Explanations



**Figure S3. Feature contribution of serum complement C3 at the point of remission to the risk prediction.**

Each dot represents the SHAP value calculated for each patient. The x-axis is the value of serum complement C3 at the point of remission, and the y-axis is the SHAP value assigned to their measure of serum complement C3. A higher SHAP value indicates that the risk of relapse attributed to serum complement C3 at the point of remission increases.

**Abbreviation:** SHAP, Shapley Additive Explanations