Diagnosis of acute promyelocytic leukemia based on routine biological parameters using machine learning

Estelle Cheli, Simon Chevalier, Olivier Kosmider, Marion Eveillard, Nicolas Chapuis, Adriana Plesa, Maël Heiblig, Lydie Andre, Jenny Pouget, Pascal Mossuz, Olivier Theisen, Vincent Alcazer, Dan Gugenheim, Nicolas Autexier and Pierre Sujobert

1 Hospices Civils de Lyon, Hôpital Lyon Sud, Service d'Hématologie Biologique, Lyon; 2 CHU Grenoble Alpes, Service d'Hématologie Biologique, Grenoble; 3 APHP, Hôpital Cochin, Service d'Hématologie Biologique, Cochin; 4 CHU Nantes, Service d'Hématologie Biologique, Nantes; 5 Hospices Civils de Lyon, Hôpital Lyon Sud, Service d'Hématologie Clinique, Lyon; 6 Groupe onepoint, Bordeaux and 7 Université Claude Bernard Lyon 1, Faculté de Médecine et de Maïeutique Lyon Sud, Charles Mérieux, Lymphoma Immunobiology Team, Lyon, France.

Correspondence:
PIERRE SUJOBERT - pierre.sujobert@chu-lyon.fr
https://doi.org/10.3324/haematol.2021.280406
### Supplemental table 1: Missing data in the training cohort

| Features                                           | Missing data, n (%) |
|----------------------------------------------------|---------------------|
| **Demographic characteristics**                   |                     |
| Sex                                                | 0 (0)               |
| Age at diagnosis                                   | 0 (0)               |
| **Hematology**                                     |                     |
| Leucocytes (x10⁹/L)                                | 1 (0.5)             |
| Red Blood Cell (x10¹²/L)                           | 6 (2.7)             |
| Hemoglobin (g/L)                                   | 1 (0.5)             |
| Hematocrit (%)                                     | 5 (2.25)            |
| Mean Corpuscular Volume (fL)                       | 4 (1.8)             |
| Mean Corpuscular Hemoglobin (pg)                   | 3 (1.3)             |
| Mean Corpuscular Hemoglobin Concentration (g/L)    | 5 (2.25)            |
| Red blood cell Distribution Width-CV (%)           | 6 (2.7)             |
| Neutrophils (x10⁹/L)                               | 7 (3.15)            |
| Eosinophils (x10⁹/L)                               | 9 (4.05)            |
| Basophils (x10⁹/L)                                 | 9 (4.05)            |
| Lymphocytes (x10⁹/L)                               | 9 (4.05)            |
| Monocytes (x10⁹/L)                                 | 9 (4.05)            |
| Blasts (%)                                         | 12 (5.4)            |
| Platelets (x10⁹/L)                                 | 2 (0.9)             |
| Mean Platelet Volume (fL)                          | 35 (15.7)           |
| Reticulocytes (x10⁹/L)                             | 108 (48.6)          |
| **Hemostasis**                                     |                     |
| Prothrombin ratio (%)                              | 3 (1.3)             |
| Activated Partial Thromboplastin Time              | 2 (0.9)             |
| Fibrinogen (g/L)                                   | 8 (3.6)             |
| Fibrin Degradation Products (µg/mL)                | 194 (87.4)          |
| **Biochemistry**                                   |                     |
| LDH (U/L)                                          | 30 (13.5)           |
| Calcium (mmol/L)                                   | 4 (1.8)             |
| Phosphorus (mmol/L)                                | 23 (10.3)           |
| Ferritin (µg/L)                                    | 83 (37.4)           |
| C Reactive Protein (mg/L)                          | 21 (9.5)            |
| Urea (mmol/L)                                      | 3 (1.3)             |
| Creatinine (µmol/L)                                | 2 (0.9)             |
| Uric Acid (µmol/L)                                 | 22 (9.9)            |
Supplemental figure 1: Value of the confidence score for each patient according to the time of diagnosis in the validation cohort n°3. The analyzers automat used for hematology, hemostasis, and biochemistry were the following: period 1: Siemens Dimension Vista, StaR, Beckman Coulter LH780; period 2: Siemens Dimension Vista, StaRevolution, Sysmex XE5000 and XE2100 D; period 3: Siemens Atellica, StaRMax3 and Sysmex XN. The differences are non-significant (Kruskal-Wallis test).

Supplemental figure 2: Variations in AIPL confidence scores according to the cytological subtype of APL.