

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

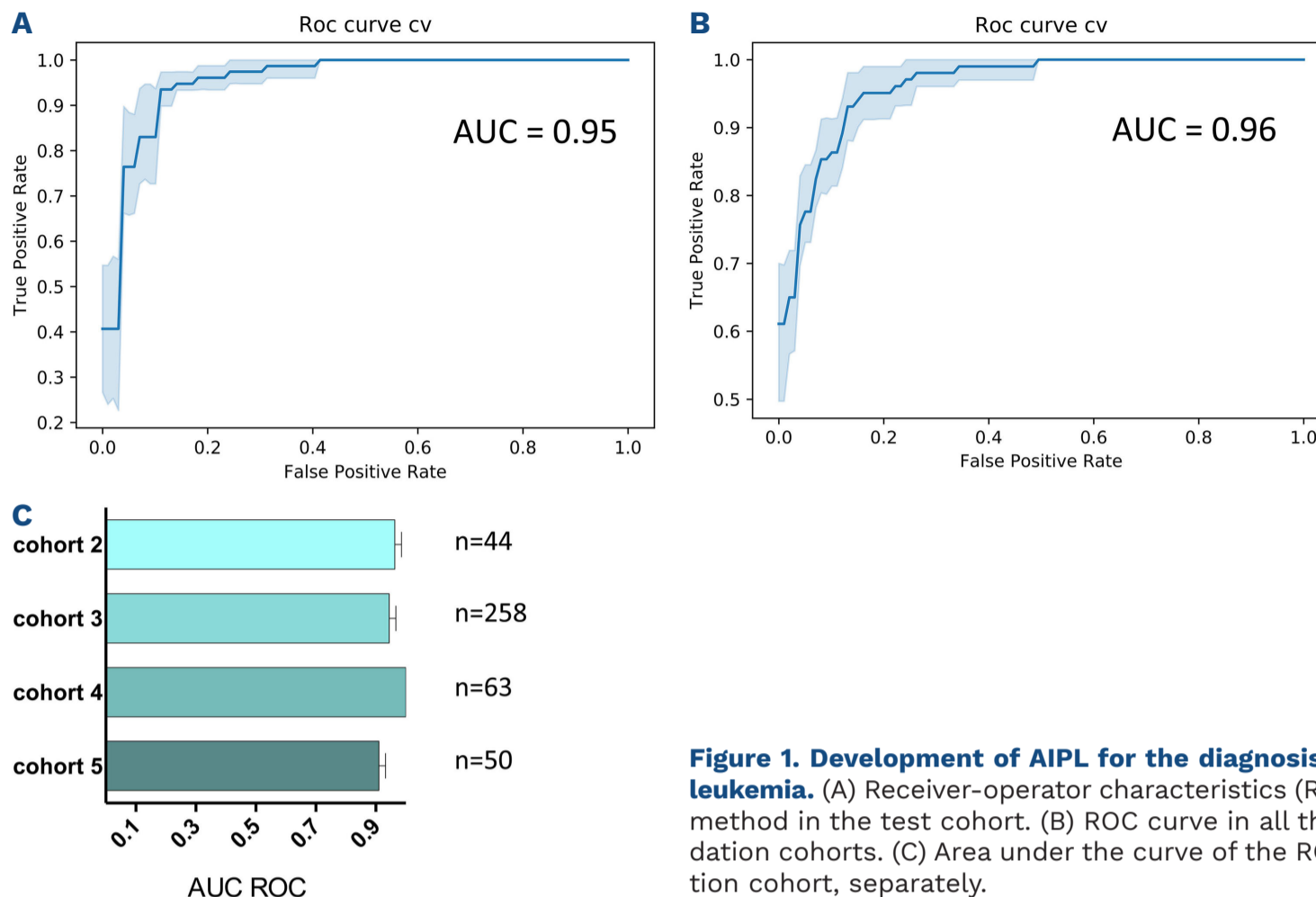
Acute promyelocytic leukemia (APL) has become the most curable subtype of acute myeloid leukemia (AML) since the introduction of all-trans retinoic acid and arsenic trioxide.<sup>1</sup> However, this disease is still characterized by a high rate of early death (10-17%), mainly due to severe coagulopathy.<sup>2,3</sup> In order to avoid these early deaths, immediate treatment initiation is recommended, either with all-trans retinoic acid or with chemotherapy in case of hyperleucocytosis.<sup>4</sup> Hence, a fast and accurate diagnosis is mandatory to allow early recognition and treatment of APL.

Cytology is the fastest technique for the diagnosis of APL, while the definitive confirmation requires the observation of the t(15;17) translocation or the *PML-RARA* fusion mRNA amplification, which induces further delay. Cytogenetic and molecular confirmation can be more difficult when other partners of *RARA* are implicated,<sup>5</sup> and even more challenging in exceptional cases of viral insertion in the *RARA* gene, as recently described.<sup>6</sup> The cytological diagnosis of APL is usually straightforward, when multiple bundles of Auer rods are observed in the blasts cells. However, the microgranular variant might be more difficult to diagnose, even for experienced hematopathologists. In some cases, myeloperoxidase deficiency in the blast cells further complicates the recognition of APL.<sup>7,8</sup> Moreover, cytology requires a long training to recognize rare diseases such as APL, and this expertise is not always available.

We hypothesized that routine biological parameters might fuel an artificial intelligence to identify APL without a high level of cytological expertise. We collected 34 basic biological parameters in all the APL patients diagnosed in Lyon University Hospital during the period from 2013 to 2020 (n=76), and in patients with non-promyelocytic AML matched according to the year of diagnosis (n=146). Altogether, these patients constituted the cohort 1 (n=222). All the APL cases were confirmed by cytogenetic and/or quantitative reverse transcription polymerase chain reaction (RT-qPCR) amplification of the *PML-RARA* fusion transcript. The biological parameters were measured during the first 2 days of hospital referral, before any treatment initiation. Missing data were imputed by the variable's median value.<sup>9</sup> The basic hematology and hemostasis parameters were available for most of the patients, but there were more missing data concerning the biochemical parameters (*Online Supplementary Table S1*). The cohort was randomly split into a training (80%, n=177) and a test cohort (20%, n=45). Different classification algorithms were then compared (XGBoost, random forest, gradient boosting classifier, adaboost classifier, decision tree, logistic regression and support vector machine), considering APL

diagnosis as a binary outcome and using 5X cross-validation to select the more stable models. No normalization of the data was used, because both strategies tested (StandardScaler, MinMaxScaler) had a negative impact on the performances of the algorithm. Hyperparameter tuning was performed using GridSearchCV. All analyses were performed using Python v3.7. Among the different artificial intelligence strategies tested, XGBoost's gradient boosting algorithm achieved the highest performances in the test cohort, with an area under the receiver operator curve (ROC) of 0.95 (Figure 1A, see the *Online Supplementary Appendix* for methodological details). Of note, learning curves reached a plateau with 80-100 patients, meaning that no major refinements is expected with an increase in the size of the cohort (data not shown). Using this model, we established artificial intelligence for promyelocytic leukemia (AIPL), an open-source tool with a graphical user interface to evaluate the probability of APL diagnosis (<https://github.com/Nico-Facto/Leukemia-Apl-Classification>) and propose a ready-to-use web application ([https://share.streamlit.io/nico-facto/leukemia-apl-classification/main/Leucemie\\_app.py](https://share.streamlit.io/nico-facto/leukemia-apl-classification/main/Leucemie_app.py)). The eight parameters required to run AIPL are the following: age, white blood cells (absolute value), lymphocytes (% of total leukocytes), neutrophil polynuclear count (absolute value), mean corpuscular volume (MCV), mean corpuscular hemoglobin concentration (MCHC), prothrombin time ratio, and fibrinogen concentration.

In order to validate the AIPL tool, its performances were assessed in three independent retrospective validation cohorts from three other hospitals (cohorts 2, 3, and 4) which comprised 44 (including 15 APL), 258 (including 46 APL), and 63 (including 32 APL) patients, respectively. A prospective cohort (cohort 5) was also collected in the Lyon University Hospital with 50 (including 10 APL) new AML diagnoses referred during a 6-month period. AIPL showed a very high discrimination ability both in the merged (n=415 patients, AUC =0.96, Figure 1B) and in individual cohorts (Figure 1C). Importantly, AIPL output is not only a classification (APL vs. non-APL), but also a confidence score reflecting how much the conclusion can be trusted. As expected, the confidence score was significantly higher in cases where the prediction was correct compared to cases where the AIPL prediction failed (mean 95% vs. 85%, Mann-Whitney test  $P < 0.0001$ , Figure 2A). Hence, the AIPL confidence score could be used to determine for which patient the prediction of AIPL is reliable in routine use. For 244 (59%) patients with a high confidence score (above 99%), the accuracy of AIPL was



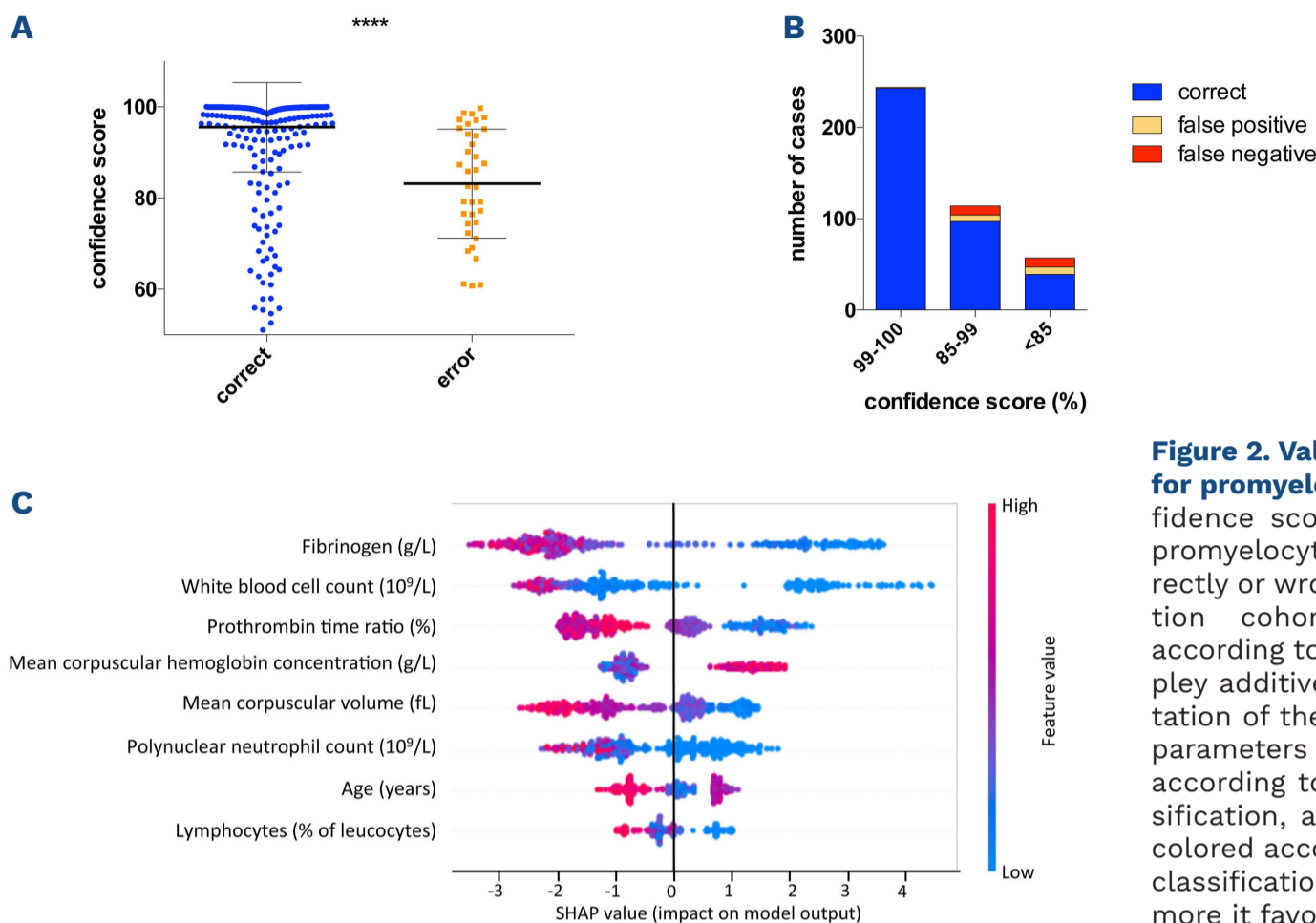
**Figure 1. Development of AIPL for the diagnosis of acute promyelocytic leukemia.** (A) Receiver-operator characteristics (ROC) curve of the XGBoost method in the test cohort. (B) ROC curve in all the patients from the validation cohorts. (C) Area under the curve of the ROC curves in each validation cohort, separately.

99.5% (only one false negative case, i.e., an APL patient wrongly classified as non-APL AML). For 114 (27%) patients with an intermediate confidence score (between 85 and 99%), the accuracy was 85% (7 false negative and 10 false positive cases, i.e., a non-APL AML patient wrongly classified as APL), and for 57 (14%) patients with a low confidence score (below 85%), the accuracy dropped to 68% (8 false negative and 10 false positive cases) (Figure 2B). As the data from the retrospective cohorts were acquired on different analyzers, it was possible to assess the impact of the variability in analytical techniques on AIPL performances. When comparing different time periods determined by changes in analyzers, there was no significant variation in the level of confidence scores of AIPL, suggesting that this approach is robust to variations due to analytical processes (*Online Supplementary Figure S1*). Of note, 16 cases of the microgranular variant of APL were identified in the validation cohorts. Their AIPL confidence scores tended to be lower than the confidence scores of classical APL (86% vs. 92%, ns, *Online Supplementary Figure S2*), suggesting that further algorithm training using microgranular variants might be interesting. Importantly, AIPL correctly identified the six microgranular cases with a confidence score above 99%. We also assessed the performances of AIPL in patients with other differential diagnosis of APL: aplastic anemia (n=10), acute lymphoblastic leukemia (n=28) and AML with t(8;16), a rare subtype of AML with clinical presentation often resembling APL (n=9). AIPL classified only one case aplastic anemia and one case of acute lymphoblastic leukemia as APL, but with

confidence score below 87%. Hence, with the proposed threshold of confidence score of 99%, there was no false positive diagnostic of APL in these challenging differential diagnoses.

In order to further interpret the predictions from AIPL, Shapley additive explanation (SHAP) was used to illustrate the impact of the different parameters according to their value obtained from the individual cases of all the validation cohorts.<sup>10</sup> In Figure 2C, the parameters are ordered from top to bottom according to their importance in the classification, and each individual measure is colored according to its impact on the final classification. The high performances of AIPL rely on some expected parameters such as fibrinogen, prothrombin time ratio, or polynuclear neutrophil count (Figure 2C). Unexpectedly, two parameters of red blood cells (MCV and MCHC) were highly discriminant between APL and non-APL AML, even if the mean values of these parameters remained in the normal ranges (mean value of MCV 89 fL vs. 96 fL, mean value of MCHC 349 g/L vs. 334 g/L in APL and non-APL cases, respectively). This observation, together with a report of PML-RARA expression in burst forming unit-erythroid (BFU-E) derived from APL patients,<sup>11</sup> raises the hypothesis that PML-RARA cells contribute to erythropoiesis.

To conclude, this work demonstrates that machine learning based on routine biological parameters provides a fast and accurate help for the diagnosis of APL in the majority of cases. Given the unmet need to improve the reliability of APL diagnosis, other strategies based on incorporation of cell population data generated during complete blood



**Figure 2. Validation of artificial intelligence for promyelocytic leukemia.**

(A) Mean confidence score of artificial intelligence for promyelocytic leukemia (AIPL) in cases correctly or wrongly classified in all the validation cohorts. (B) AIPL performances according to the confidence score. (C) Shapley additive explanation (SHAP) representation of the parameters used by AIPL. The parameters are ordered from top to bottom according to their importance in the classification, and each individual measure is colored according to its impact on the final classification (the more a point is red, the more it favors the classification as APL)

cell count with cytometry based analyzers,<sup>12</sup> or deep learning analysis of blood smears, have also been developed.<sup>13</sup> As these approaches rely on parameters not used in AIPL, combining them with AIPL could help further increase diagnostic accuracy. Another interesting possibility could be the addition of other biological parameters such as the fibrinolysis marker D-Dimers, which were not included in this retrospective study due to excessive missing data.

AIPL might represent a very important complement to cytological expertise, and could allow early diagnosis of APL in settings where this expertise is not available on a 24/7 basis, or not available at all, such as in developing countries. The consequences of misclassification could be excessive treatment with ATRA in case of false positive result, or delay in ATRA initiation in case of false negative result. Using AIPL with the proposed threshold of 99% of confidence score, this risk is very low but should not be forgotten. All-trans retinoic acid could hence be initiated in patients with a high probability of APL according to AIPL prediction without waiting for diagnostic confirmation in specialized laboratories, thus preventing early death from coagulopathy. However, an important limitation of AIPL is that its ability to distinguish APL from other differential diagnoses such as acute lymphoblastic leukemia or aplastic anemia has not been formally assessed in this study. In order to make this tool available, a web user interface has been created (available at <https://share.streamlit.io/nico-facto/leukemia-apl-clas->

sification/main/Leucemie\_app.py) to use AIPL in the case of patients with myeloid blast on the blood smear. It allows to instantaneously classify patients as APL or non-APL and provides a confidence score. Of course, this result does not supplant the need to evaluate the bone marrow and to formally demonstrate the presence of the t(15;17) translocation or the *PML-RARA* fusion transcript. Given the importance of early treatment initiation in these patients, we hope that AIPL will contribute to decrease early mortality in APL patients.

## Authors

Estelle Cheli,<sup>1</sup> Simon Chevalier,<sup>2</sup> Olivier Kosmider,<sup>3</sup> Marion Eveillard,<sup>4</sup> Nicolas Chapuis,<sup>3</sup> Adriana Plesa,<sup>1</sup> Maël Heiblig,<sup>5</sup> Lydie Andre,<sup>2</sup> Jenny Pouget,<sup>4</sup> Pascal Mossuz,<sup>2</sup> Olivier Theisen,<sup>4</sup> Vincent Alcazer,<sup>5</sup> Dan Gugenheim,<sup>6</sup> Nicolas Autexier<sup>6</sup> and Pierre Sujobert<sup>1,7</sup>

<sup>1</sup>Hospices Civils de Lyon, Hôpital Lyon Sud, Service d'Hématologie Biologique, Lyon; <sup>2</sup>CHU Grenoble Alpes, Service d'Hématologie Biologique, Grenoble; <sup>3</sup>APHP, Hôpital Cochin, Service d'Hématologie Biologique, Cochin; <sup>4</sup>CHU Nantes, Service d'Hématologie Biologique, Nantes; <sup>5</sup>Hospices Civils de Lyon, Hôpital Lyon Sud, Service d'Hématologie Clinique, Lyon; <sup>6</sup>Groupe onepoint, Bordeaux and <sup>7</sup>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>

Received: November 23, 2021.

Accepted: February 11, 2022.

Prepublished: February 24, 2022.

### Disclosures

NA and DG were employed by Groupe onepoint when the work was conducted. All other authors have no conflicts of interest to declare.

### Contributions

PS, AP, EC and DG designed the study; EC, SC, OK, ME, NC, AP, PM,

LA, JP, PM, OT and MH collected the data; NA, VA and DG developed the AI and the web interface; PS, EC, NA and DG analyzed the data; PS wrote the paper and all authors revised and approved the final manuscript.

### Acknowledgements

The authors thanks Mrs Verena Landel from Hospices Civils de Lyon for English editing, as well as Dr Nicolas Duployez, Dr Damien Bodet, Dr Thomas Boyer and Dr Pascale Flandrin for their help in assessing AIPL performances on patients with t(8;16) AML.

### Data sharing statement

The data that support the findings of this study are available on request from the corresponding author.

## References

---

1. Lo-Coco F, Avvisati G, Vignetti M, et al. Retinoic acid and arsenic trioxide for acute promyelocytic leukemia. *N Engl J Med*. 2013;369(2):111-121.
2. Park JH, Qiao B, Panageas KS, et al. Early death rate in acute promyelocytic leukemia remains high despite all-trans retinoic acid. *Blood*. 2011;118(5):1248-1254.
3. Rahmé R, Thomas X, Recher C, et al. Early death in acute promyelocytic leukemia (APL) in French centers: a multicenter study in 399 patients. *Leukemia*. 2014;28(12):2422-2424.
4. Sanz MA, Fenaux P, Tallman MS, et al. Management of acute promyelocytic leukemia: updated recommendations from an expert panel of the European LeukemiaNet. *Blood*. 2019;133(15):1630-1643.
5. Grimwade D, Biondi A, Mozziconacci MJ, et al. Characterization of acute promyelocytic leukemia cases lacking the classic t(15;17): results of the European Working Party. Groupe Français de Cytogénétique Hématologique, Groupe de Français d'Hématologie Cellulaire, UK Cancer Cytogenetics Group and BIOMED 1 European Community-Concerted Action "Molecular Cytogenetic Diagnosis in Haematological Malignancies." *Blood*. 2000;96(4):1297-1308.
6. Astolfi A, Masetti R, Indio V, et al. Torque teno mini virus as a cause of childhood acute promyelocytic leukemia lacking PML/RARA fusion. *Blood*. 2021;138(18):1773-1777.
7. Cui W, Qing S, Xu Y, Hao Y, Xue Y, He G. Negative stain for myeloid peroxidase and Sudan black B in acute promyelocytic leukemia (APL) cells: report of two patients with APL variant. *Haematologica*. 2002;87(5):ECR16.
8. Heiblig M, Paubelle E, Plesa A, et al. Comprehensive analysis of a myeloperoxidase-negative acute promyelocytic leukemia. *Blood*. 2017;129(1):128-131.
9. Acuña E, Rodriguez C. The Treatment of Missing Values and its Effect on Classifier Accuracy. In: Banks D, McMorris FR, Arabie P, Gaul W, editors. *Classification, Clustering, and Data Mining Applications*. Berlin, Heidelberg: Springer. 2004. p639-647.
10. Lundberg SM, Erion GG, Lee S-I. Consistent Individualized Feature Attribution for Tree Ensembles. *arXiv*. <https://doi.org/10.48550/arXiv.1802.03888> preprint, [not peer-reviewed]
11. Takatsuki H, Sadamura S, Umemura T, et al. PML/RAR alpha fusion gene is expressed in both granuloid/macrophage and erythroid colonies in acute promyelocytic leukaemia. *Br J Haematol*. 1993;85(3):477-482.
12. Haider RZ, Ujjan IU, Shamsi TS. Cell population data-driven acute promyelocytic leukemia flagging through artificial neural network predictive modeling. *Trans Oncol*. 2020;13(1):11-16.
13. Sidhom J-W, Siddarthan IJ, Lai B-S, et al. Deep Learning for distinguishing morphological features of acute promyelocytic leukemia. *Blood*. 2020;136(Suppl 1):S10-12.