Artificial intelligence-based quantitative bone marrow pathology analysis for myeloproliferative neoplasms

Dandan Yu,[1-3]* Hongju Zhang,[1,2]* Yanyan Song,[1,2] Yuan Tao,[1,2] Fengyuan Zhou,[4] Ziyi Wang,[4] Rongfeng Fu,[1-3] Ting Sun,[1-3] Huan Dong,[1-3] Wenjing Gu,[1-3] Renchi Yang,[1-3] Zhijian Xiao,[1,2] Qi Sun[1,2] and Lei Zhang[1-3]

[1]State Key Laboratory of Experimental Hematology, National Clinical Research Center for Blood Diseases, Haihe Laboratory of Cell Ecosystem, Tianjin Key Laboratory of Gene Therapy for Blood Diseases, CAMS Key Laboratory of Gene Therapy for Blood Diseases, Institute of Hematology & Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Tianjin; [2]Tianjin Institutes of Health Science, Tianjin; [3]School of Population Medicine and Public Health, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing and [4]XY AI Technologies (Su Zhou) Limited, Jiangsu, China

*DY and HZ contributed equally as first authors.

Abstract

The evaluation of bone marrow pathology is essential for diagnosing and classifying myeloproliferative neoplasms (MPN). However, morphological assessments of bone marrow trephine (BMT) sections by hematopathologists are inherently subjective; thus, an accurate and objective diagnostic system is needed. Based on U²-Net, UNeXt, and ResNet, we developed an automated quantitative analysis platform for BMT sections from MPN patients and non-neoplastic cases (N=342 total) to enhance the accuracy of diagnosis and classification of MPN. Bone marrow metrics, including marrow cellularity, the myeloid-to-erythroid ratio, megakaryocyte morphology and distribution, and the grading of marrow fibrosis, were quantitatively analyzed (with an accuracy of approximately 0.9) based on the accuracy segmentation and identification of various cells and tissues (with an intersection over union of roughly 0.8). A bone marrow classification model incorporating bone marrow metrics, a clinical classification model utilizing clinical features, and a comprehensive classification model that includes both bone marrow metrics and clinical features were developed using random forest classifiers to differentiate MPN subtypes and non-neoplastic conditions. The bone marrow and comprehensive classification models reached a macro-average area under the curve (AUC) of 0.96 for differentiating MPN subtypes and non-neoplastic cases. The clinical classification model attained a macro-average AUC of 0.92. This platform is highly accurate for quantitatively analyzing bone marrow pathology and classifying MPN subtypes and non-neoplastic cases. It can be a potentially auxiliary diagnostic tool for hematopathologists when dealing with patients with suspected MPN.

Introduction

Myeloproliferative neoplasms (MPN) are a group of clonal hematopoietic stem cell disorders characterized by the excessive proliferation of one or more myeloid cell lineages.[1] Philadelphia chromosome-negative MPN consist of polycythemia vera (PV), essential thrombocythemia (ET), and primary myelofibrosis (PMF). PMF can be classified as either prefibrotic/early PMF (pre-PMF) or overt PMF.[2] Clear-cut differentiation among MPN is crucial for effective management since these entities differ significantly in terms of complications, risk of progression, and overall survival.[3-6] The integration of clinical, laboratory, and histopathological

characteristics is required to diagnose and categorize MPN accurately. MPN patients share clinical characteristics and laboratory features such as genetic mutations (generally *JAK2*, *CALR*, and *MPL*), especially ET and pre-PMF patients. However, they exhibit different morphological features in bone marrow trephines (BMT); therefore, careful histological examination of the bone marrow is critical for identifying and classifying MPN subtypes.[7,8] The pivotal aspects for accurate pathological differentiation of MPN are bone marrow cellularity, the increase of granulocytes and nucleated erythroid cells, the cytological and topographic features of megakaryocytes, and the severity of marrow fibrosis (MF). Currently, the pathological interpretation of BMT from suspected MPN patients assessed by hematopathologists is

highly demanding and subjective, and the reproducibility is controversial.[9]

Recent advancements in digital pathology have made computer image analysis a powerful supplement to traditional histopathology.[10,11] Using deep machine learning and image analysis to examine whole-slide images, these methods can help to diagnose diseases, predict progression, and assess prognosis by identifying and quantifying specific cell populations systematically.[12,13] In medicine, deep machine learning is currently mainly being applied in diagnosing cancers based on laboratory features, histopathological analyses, other biological factors, etc., and its accuracy and effectiveness have been verified.[14-17] In this study, we developed a set of advanced algorithms involving U^2-Net, UNeXt, and ResNet to aim to assist hematopathologists in the detailed quantitative analysis of BMT samples for classifying Philadelphia chromosome-negative MPN. Additionally, we constructed random forest classifiers that integrate bone marrow metrics and clinical data to distinguish between non-neoplastic conditions and MPN subtypes with greater accuracy, objectivity, and efficiency.

Methods

This study was approved by the Human Research Ethics Committee of the Institute of Hematology and Blood Diseases Hospital, CAMS-PUMC (QTJC2023037-EC-1).

The overall study workflow comprised four major steps (Figure 1).

Data preparation

Overall, 309 MPN patients (78 ET, 37 pre-PMF, 27 PV, and 167 PMF) and 33 non-neoplastic cases (25 with iron-deficiency anemia and 8 healthy donors) from the Blood Disease Hospital, Chinese Academy of Medical Sciences were enrolled retrospectively. Diagnoses were made following the 2016 World Health Organization (WHO) criteria through a uniform review process.[18] Eligible hematoxylin & eosin (H&E)- and Gomori-stained BMT obtained at diagnosis were collected (*Online Supplementary Figure S1*). All samples were scanned with Aperio GT 450 Dx at 40x magnification and saved in the SVS format. Peripheral blood counts, spleen sizes, lactate dehydrogenase levels, and gene mutation status were collected (*Online Supplementary Table S1*). In addition, eligible BMT samples from 96 MPN patients (19 ET, 20 pre-PMF, 38 PV, and 19 PMF) and ten healthy donors from Tianjin Union Bio-Precision Medical Diagnostics Technology Co., Ltd. served as an external test set.

Automated detection and segmentation of cells and tissues in bone marrow trephine samples

Three hematopathologists meticulously annotated 89,449 granulocytes/erythroid cells, 3,217 megakaryocytes, and a series of tissues by delineating contours using a custom annotation platform, Digital Pathology Analysis System (version 2.0.30). All annotations underwent cross-review by three hematopathologists (*Online Supplementary Table S2*). Target segmentations in H&E-stained and Gomori-stained sections were performed using U^2-Net and UNeXt, respectively, while ResNet-18 differentiated granulocytes and erythroid cells.[19-21] The algorithms underwent iterative learning cycles by hematopathologists who reviewed the computer-assisted annotations to rectify misidentifications (Figure 1, *Online Supplementary Methods*: *Detection and delineation of cells and tissues*).

Quantitative analysis of various bone marrow metrics

Bone marrow cellularity is determined by the average ratio of the area occupied by hematopoietic cells to the combined area of hematopoietic and fat cells (Figure 2A). The myeloid-to-erythroid (M:E) ratio is the ratio of granulocytes to erythropoietic cell numbers.

Megakaryocyte morphology includes size, nuclear-cytoplasmic ratio, and the presence of naked nuclei. Those characteristics were quantitatively defined based on megakaryocytes from healthy donors (*Online Supplementary Figures S8-S10*). Megakaryocyte distribution was assessed by centroid distance (*Online Supplementary Figure S11*) and categorized as non-clustered or clustered, with loose or dense clusters. Clustering was identified using the density-based spatial clustering of applications with noise (DBSCAN) algorithm based on manual annotations.[22]

For fibrosis grading, slides were segmented into small patches (*Online Supplementary Figure S12*), and fibrosis grades (MF 0 to 3) were predicted based on those patches labeled by hematopathologists (*Online Supplementary Figure S13*). Then, the number of patches corresponding to each grading (MF 0 to 3) of the entire slide was counted. According to the 2022 WHO diagnostic criteria, the final slide grade was determined by the highest grade present in at least 30% of the marrow area (*Online Supplementary Figure S14*).

Classification models

The bone marrow, clinical, and comprehensive classification models were constructed using random forest classifiers based on histopathological features, clinical features, and their combined integration, respectively (see *Online Supplementary Methods*: *Classification*).

Results

Quantitative analysis of various metrics related to bone marrow histopathology

Bone marrow cellularity

Bone marrow cellularity offers insight into the patient's current hematopoietic status. The bone marrow cellularity was calculated based on accurate segmentation of
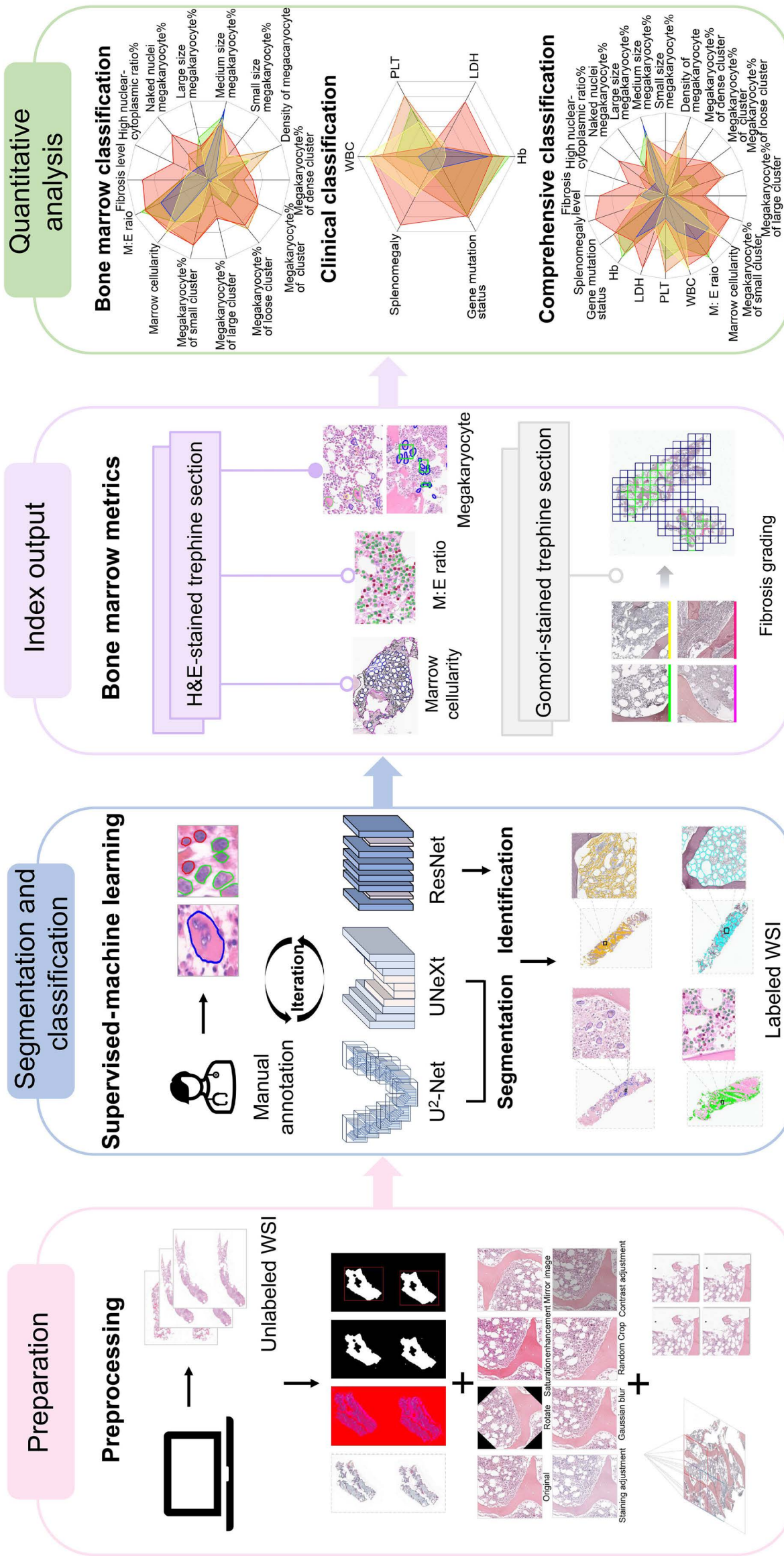
Figure 1. The overall workflow of the computer-assisted analysis for myeloproliferative neoplasms. *Preparation.* Eligible biopsy samples were scanned and stored as whole-slide images, followed by preprocessing steps, including extraction of the region of interest, image enhancement, and misaligned splitting. *Segmentation and classification.* An annotate-predict-correct cycle involving the UNeXt, U²-Net, and ResNet-18 algorithms was employed for iterative optimization and training of the models to segment and identify cells and tissues accurately. *Index output.* Bone marrow metrics were the output, based on the accurate segmentation and identification of various cells and tissues, including bone marrow cellularity, myeloid-to-erythroid ratio, morphology and distribution of megakaryocytes, and fibrosis severity. *Quantitative analysis.* The bone marrow indicators were analyzed quantitatively and incorporated to construct a bone marrow classification model based on the random forest classifier. Clinical indicators were included to establish a clinical classification model. A comprehensive classification model was also trained by fully incorporating bone marrow metrics and clinical indicators. WSI: whole-slide images; H&E: hematoxylin & eosin; M:E: myeloid-to-erythroid ratio; WBC: white blood cell count; PLT: platelet count; LDH: lactate dehydrogenase; Hb: hemoglobin.
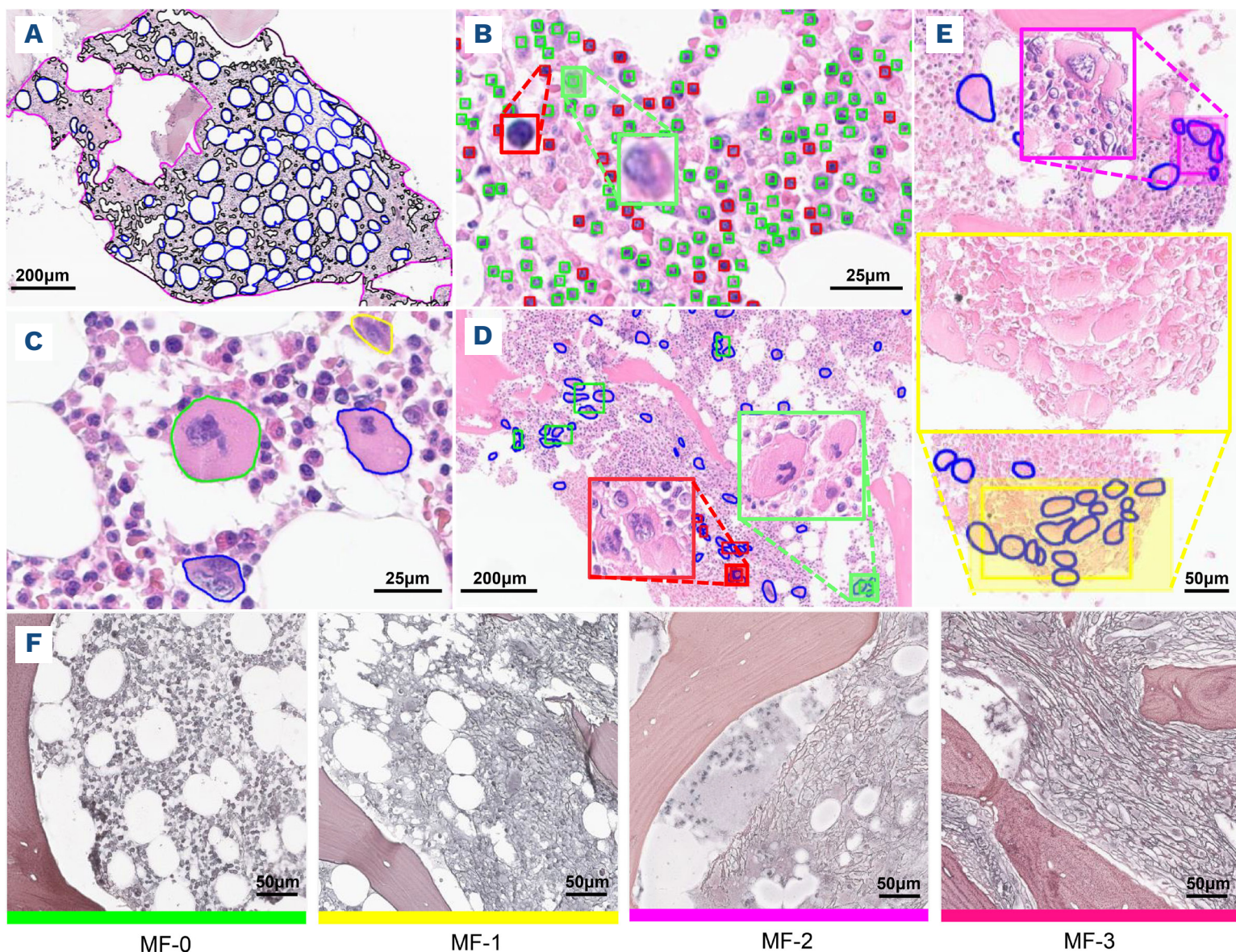
Figure 2. Identification of various cells and tissues in hematoxylin & eosin-stained biopsy specimens. (A) Recognition of structures related to hematopoietic tissue at 5X magnification. The purple contours represent the hematopoietic regions segmented by the model, with the blue contours indicating fat cells within the hematopoietic area. The black areas are blank areas within the hematopoietic region obtained through thresholding, morphological transformations, and filtering for smoothing. The actual hematopoietic tissue was derived by subtracting these blank areas from the hematopoietic region. (B) The identification of granulocytes and nucleated erythroid cells at 40X magnification, with nucleated erythroid cells marked in red and granulocytes in green. (C) The recognition of megakaryocytes at 40X magnification: large megakaryocytes are outlined in green, medium megakaryocytes in blue, and small megakaryocytes in yellow. (D) The identification of megakaryocyte clustering at 5X magnification. Dense clusters are highlighted in red boxes, and loose clusters in green. (E) The identification of megakaryocyte clustering at 10X magnification. Large clusters are highlighted in yellow, and small clusters in purple. (F) Images of patches corresponding to each grading of marrow fibrosis (0 to 3) at 10X magnification. MF: marrow fibrosis.

hematopoietic, trabecula, and fat areas in the images (the intersection over union (IoU) ranged from 0.73 to 0.87) (Figure 2A, *Online Supplementary Figure S5A-C, Online Supplementary Table S10*). Our model predicted a concordance rate of 83.3% (with a margin of error of ±15% considered concordant) for bone marrow cellularity compared to assessments by hematopathologists. The bone marrow cellularity of patients was compared with age-adjusted bone marrow cellularity reported in the literature[23] to determine the actual degree of increase. The marrow cellularity in ET patients was comparable to that in healthy donors (68% *vs*. 64%, *P*>0.05). Patients with pre-PMF, PV, and PMF exhibited more evident hypercellularity than healthy donors (77%, 88% and 94% *vs*. 64%, respectively, all *P*<0.05) (Figure 3A).

Myeloid-to-erythroid ratio

The M:E ratio is the ratio between the quantity of granulocytes and erythroid cells and reflects the relative proliferation of granulocytes and erythroid lineage cells. Our model achieved high recognition accuracy between the two cell types, with an area under the curve (AUC) of 0.96 after 5-fold cross-validation (Figures 2B and 7, *Online Supplementary Figure S5D*, *Online Supplementary Table S11*). Meanwhile, the evaluation of the inter-observer baseline illustrated the model's reliability and confirmed that it was not overfitting to any specific pathologist (*Online Supplementary Table S13*). Except for PV patients, who have a relatively low M:E ratio (mean, 1.71), the mean M:E ratio for patients with ET, pre-PMF, and PMF, as well as healthy donors, ranged between 2:1 and 4:1 (Figure 3B).
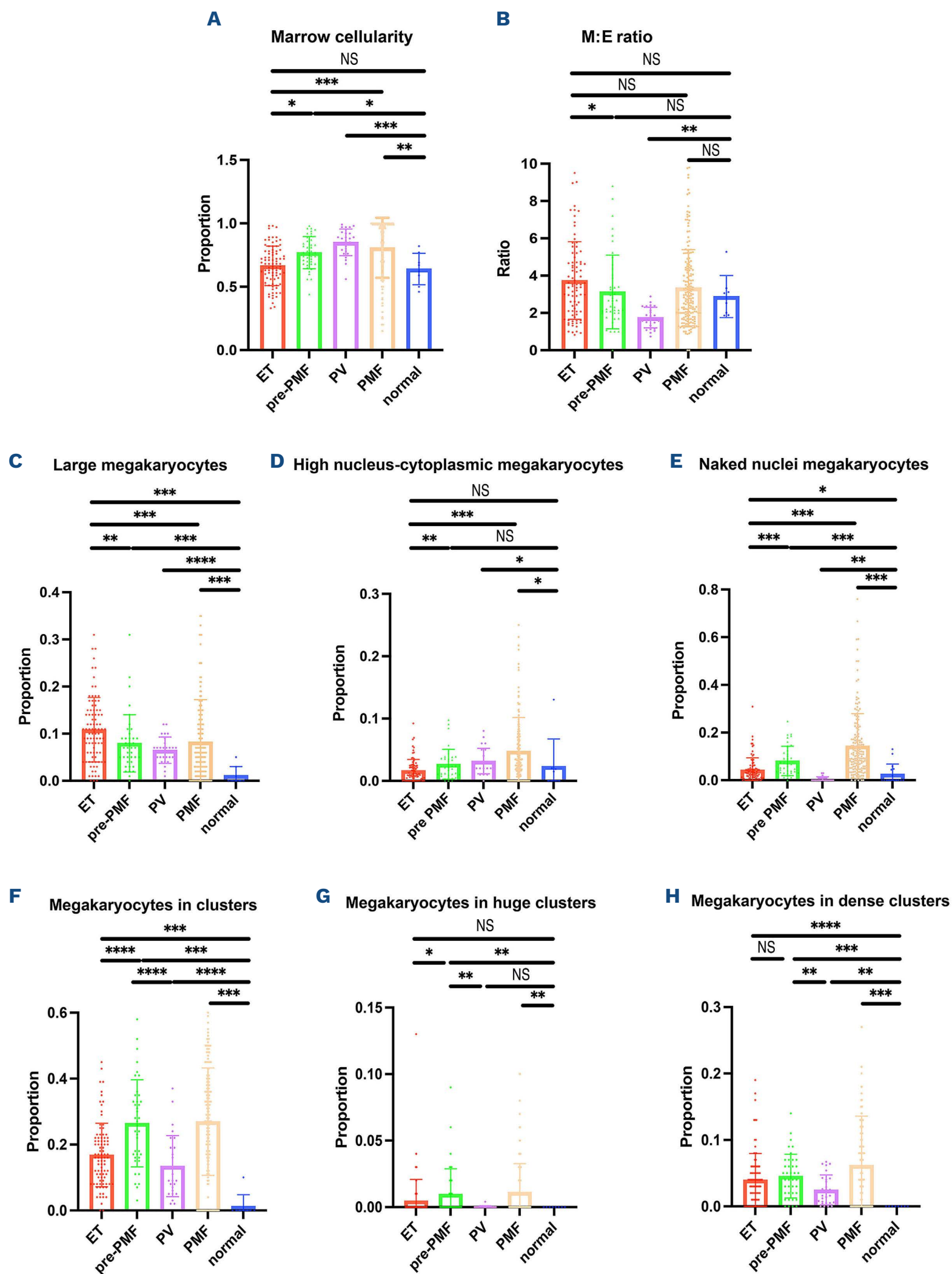
Figure 3. Quantitative analysis of various bone marrow metrics in hematoxylin & eosin-stained biopsies. (A) Comparison of bone marrow cellularity among patients with different diseases. (B) Comparison of the myeloid-to-erythroid ratios among patients with different diseases. (C-E) The differences in megakaryocyte morphology between different diseases were analyzed. (C) Comparison

of the proportion of large megakaryocytes between patients with different diseases. (D) Comparison of the proportion of mega-karyocytes with a high nucleocytoplasmic ratio (>2) between patients with different diseases. (E) Comparison of the proportion of naked megakaryocytes between patients with different diseases. (F-H) The differences in megakaryocyte distribution between different diseases were analyzed. (F) Comparison of the proportion of clustered megakaryocytes to all megakaryocytes between patients with different diseases. (G) Comparison of the proportion of megakaryocytes clustered into large clusters between patients with different diseases. (H) Comparison of the proportion of megakaryocytes clustered into dense clusters between patients with different diseases. ET: essential thrombocythemia; PMF: primary myelofibrosis; PV: polycythemia vera; M:E: myeloid-to-eryth-roid ratio. *$P<0.05$, **$P<0.01$, ***$P<0.001$, ****$P<0.0001$; NS: not significant.

Megakaryocyte morphology and distribution

Megakaryocytes play a crucial role in the examination of bone marrow due to their distinctive features between non-neoplastic conditions and MPN subtypes. Our model achieved high recognition accuracy for megakaryocytes after three training iterations (IoU=0.82) (*Online Supplementary Table S10, Online Supplementary Figure S5E*), which laid the foundation for subsequent analysis.

Based on the size range of megakaryocytes in H&E-stained sections from healthy donors (N=8) (Q2.5-Q97.5 as the normal range), we categorized megakaryocytes into small (<145.12 μm^2), medium (145.12-626.80 μm^2), and large (>626.80 μm^2) (Figure 2C, *Online Supplementary Figure S8*). Compared to healthy donors, MPN patients exhibited a significantly greater proportion of large megakaryocytes (Figure 3C). Compared to patients with PV, pre-PMF, and PMF, those with ET exhibited more large megakaryocytes (0.11 *vs.* 0.06, $P<0.05$; 0.11 *vs.* 0.07, $P<0.05$; 0.11 *vs.* 0.08, $P<0.05$, respectively).

Similarly, megakaryocytes with a nuclear-cytoplasmic ratio greater than the 97.5th percentile of the normal reference range (>2) were defined as megakaryocytes with a high nuclear-cytoplasmic ratio (*Online Supplementary Figure S10*). Furthermore, megakaryocytes with a nuclear ratio greater than 0.9 were determined to be megakaryocytes with naked nuclei. The proportion of megakaryocytes with a high nuclear-cytoplasmic ratio was comparable between ET patients and healthy donors (0.016 *vs.* 0.023, $P>0.05$); however, compared with that in ET patients, the proportion of megakaryocytes with a high nuclear-cytoplasmic ratio in pre-PMF and PMF patients was increased (0.016 *vs.* 0.026, $P<0.05$; 0.016 *vs.* 0.048, $P<0.05$, respectively) (Figure 3D). Similarly, more megakaryocytes with naked nuclei were observed in pre-PMF patients than in ET patients and healthy donors (0.081 *vs.* 0.043, $P<0.05$; 0.081 *vs.* 0.026, $P<0.05$) (Figure 3E). This tendency was more evident in PMF patients (0.144 *vs.* 0.043, $P<0.05$; 0.144 *vs.* 0.026, $P<0.05$). Clusters were defined as three or more megakaryocyte aggregations (inter-megakaryocyte distance <20 μm). Dense clusters (≥3 megakaryocytes; <5 μm spacing) and loose clusters (≥3 megakaryocytes; 5-20 μm spacing) were also defined (Figure 2D). Large clusters were defined as seven or more megakaryocyte aggregations with <20 μm spacing (Figure 2E).[2] The model achieved high accuracy in identifying megakaryocyte clusters (82.4%), with a sensitivity of 100.0%. Compared to non-neoplastic cases, MPN patients showed a clear tendency for megakaryocyte clustering, with PV, ET, pre-PMF, and PMF patients showing an increasing trend (Figure 3F). Compared to ET and PV patients, pre-PMF and PMF patients tended to demonstrate large and dense clusters of megakaryocytes in the bone marrow (Figure 3G, H).

Fibrosis grading

The grading of MF is a crucial component of the diagnosis and classification of MPN, particularly the distinction between grades MF 0/1 and MF 2/3. The differentiation of MF 2/3 grades forms the basis for diagnosing PMF. The predicted fibrosis grading of each patch (1,536x1,536 pixels) is shown in Figure 2F. Based on the evaluation of the entire BMT section (*Online Supplementary Figure S14*), the accuracy in differentiating between MF 0/1 and MF 2/3 grades by the model compared to manual assessment was 0.916 (N=250).

Classification

We trained a random forest classifier based on 14 bone marrow pathological metrics to distinguish between non-neoplastic conditions and MPN subtypes (Figure 4A). This classification model reached a macro-average AUC of 0.96 (Figure 5A), and the three-dimensional principal component analysis plot showed slight differentiation among various diseases (*Online Supplementary Figure S16A*). The AUC was 0.91 for discriminating between ET and other samples, 0.92 for discriminating between pre-PMF and other samples, 1 for discriminating between PMF and other samples, and 1 for discriminating between non-neoplastic and MPN samples (*Online Supplementary Figure S15*). The clinical classification model, which incorporates six clinical features (Figure 4B), demonstrated a macro-average AUC of 0.92 in identifying MPN subtypes (Figure 5B). The three-dimensional principal component analysis did not demonstrate a clear separation between various diseases (*Online Supplementary Figure S16B*). The comprehensive classification model, which integrates 14 pathological features and six clinical characteristics (Figure 4C), achieved a macro-average AUC of 0.96 in categorizing MPN subtypes and non-neoplastic cases (Figure 5C). The average precisions of the bone marrow, clinical, and comprehensive classification models in categorizing MPN subtypes and non-neoplastic cases in the internal test set were 0.75, 0.67, and 0.81, respectively (*Online Supplementary Figure S17A-C*).

Given that clinical information was not available for the

external set samples, we evaluated only the performance of the bone marrow classification model for these samples. The model demonstrated a macro-average AUC of 0.94 and an average precision of 0.84 in distinguishing non-neoplastic and MPN subtypes within external samples (Figure 5D, *Online Supplementary Figure S17D*).

Discussion

We developed an artificial intelligence-based system for automated analysis of BMT sections quantitatively related to Philadelphia chromosome-negative MPN, involving marrow cellularity, M:E ratio, megakaryocytes, and MF grading. This included developing new, precise ways to measure the size
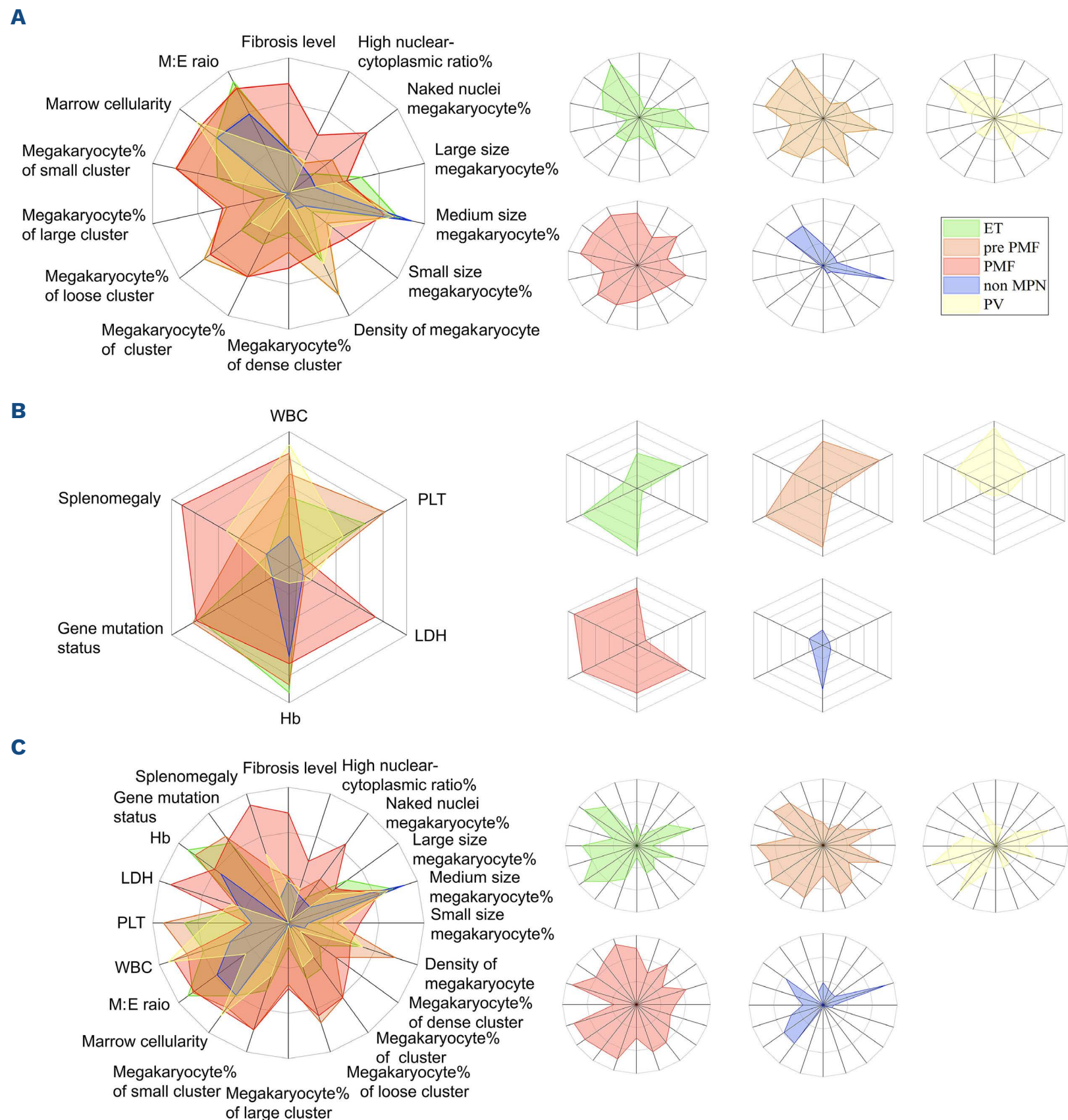

Figure 4. Metrics incorporated into classifiers. (A) Radar plots displaying the 14 indicators incorporated in the bone marrow classification model and the differences in these indicators between non-neoplastic cases and various myeloproliferative neoplasm (MPN) subtypes. (B) Radar plots displaying the six indicators incorporated in the clinical classification model and the differences among non-neoplastic conditions and various MPN subtypes in these indicators. (C) Radar plots displaying the 20 indicators incorporated in the comprehensive classification model and the differences among non-neoplastic conditions and various MPN subtypes in these indicators. M:E: myeloid-to-erythroid ratio; ET: essential thrombocythemia; PMF: primary myelofibrosis; PV: polycythemia vera; WBC: white blood cell count; PLT: platelet count; LDH: lactate dehydrogenase; Hb: hemoglobin.

and distribution of megakaryocytes, which are important steps toward increasing the objectivity of bone marrow pathology evaluations. By integrating clinical information, we also constructed a comprehensive classification model that shows promise for accurately classifying non-neoplastic cases and MPN subtypes.

In this study, all BMT sections were scanned and stored as high-resolution whole-slide images, which encompassed all morphological features present in the sections. This provided a reliable foundation for precise quantitative analysis in digital pathology.[24] Our segmentation models were iteratively optimized to achieve accurate segmentation of various tissues (with an IoU of approximately 0.8) after three rounds, indicating that the complex morphological
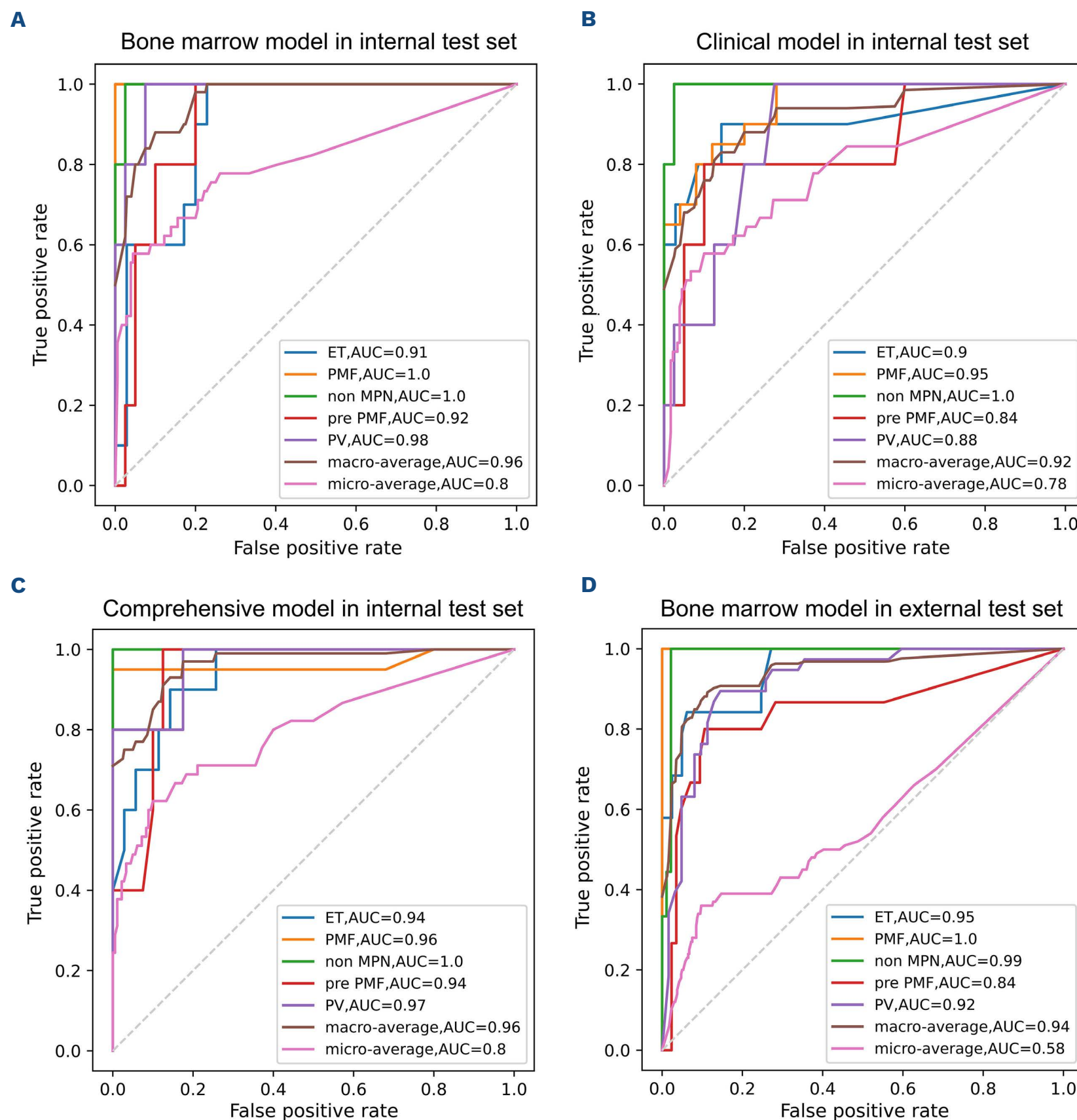

Figure 5. Categorization performances of classification models applied in differentiating non-neoplastic cases and myeloproliferative neoplasm subtypes. (A-C) Classification performances of classification models were applied to differentiate non-neoplastic cases and myeloproliferative neoplasm (MPN) subtypes within the internal test set. (A) The bone marrow classification model reached a macro-average area under the curve (AUC) of 0.96. (B) The clinical classification model reached a macro-average AUC of 0.92. (C) The comprehensive classification model reached a macro-average AUC of 0.96. (D) The bone marrow classification model reached a macro-average AUC of 0.94 in differentiating non-neoplastic cases and MPN subtypes within the external test set. ET: essential thrombocythemia; PMF: primary myelofibrosis; PV: polycythemia vera.

features of bone marrow can be accurately captured and visualized through image analysis algorithms on digital images. This provides a basis for quantitatively analyzing various indicators in BMT sections.

Our target identification algorithm models relied upon manual annotations, while existing guidelines offer only descriptive definitions of bone marrow metrics for diagnosing MPN, resulting in subjectivity among hematopathologists in assessing bone marrow morphology, especially in distinguishing immature myeloid cells and nucleated erythroid cells in H&E-stained trephine sections and fibrosis grading in Gomori-stained trephine sections. Therefore, consistent annotation data about granulocytes and nucleated erythroid cells and fibrosis grading were chosen as standard training and validation sets to weaken the subjectivity after being labeled manually by three hematopathologists independently. Furthermore, the inter-observer baseline analysis illustrated the model's reliability and confirmed that it was not overfitting to any specific pathologist. Immunohistochemistry would be an alternative method to validate the identification of granulocytes and nucleated erythroid cells in the future to enhance accuracy and reliability.

In our study, age-adjusted estimated bone marrow cellularity data from the U.S. population were used as the reference to assess bone marrow proliferation, as they are currently an internationally accepted reference. However, due to potential inter-population differences, further research utilizing population-specific reference data is warranted.

For the assessment of the morphology and distribution of megakaryocytes, we adopted a numerical quantitative way to classify and contribute to quantifying the size of megakaryocytes and the inter-megakaryocyte distance for clustering for the first time. These indicators can also be visualized, aiding hematopathologists in effectively identifying characteristic features. This enables accurate disease diagnosis and precise analysis of the differences between disease subtypes. The accurate recognition of megakaryocyte nuclear morphology is acknowledged as a crucial indicator in the diagnosis of MPN. However, we only described the nuclear-cytoplasmic ratio rather than providing a more detailed description of megakaryocyte nuclear morphology due to the complexity and heterogeneity of nuclear shapes. However, our classifiers still accurately categorized non-neoplastic cases and MPN subtypes. We speculate that the nuclear-cytoplasmic ratio, as a quantitative indicator, may provide a more objective alternative to the subjective morphological descriptions of the nucleus in disease classification.[25] Additionally, the atypia of megakaryocyte nuclei might correlate with other bone marrow metrics during disease progression, somewhat mitigating the impact of this limitation on the diagnosis.

The complete assessment of fibrosis severity according to WHO diagnostic criteria includes fiber thickness, intersections, and the presence of collagen deposition in Masson's trichrome-stained trephine biopsies, not only fiber quantity. However, we only assessed the fibrosis density in Gomori-stained trephine biopsies since Masson's trichrome-stained samples were unavailable in our center, and collagen fibers can also be observed in H&E- and Gomori-stained sections. However, the accuracy of the evaluation of fibrosis severity of our model compared to manual assessment was 0.916, which can predict the fibrosis severity to some extent, although it cannot represent the fibrosis grading completely equal to WHO MF 0-3. In future studies, Masson's trichrome-stained samples will be collected concurrently with an expanded sample cohort to enhance the comprehensiveness of the analysis. Several previous studies aimed to develop automated diagnosis systems for MPN by analyzing bone marrow pathology features. D'Abbronzo *et al.*[26] developed an artificial intelligence-based tool for evaluating cellularity in BMT sections and obtained an optimal concordance between their model and the expert pathologists' evaluation. Sirinukunwattana *et al.*[27] categorized megakaryocytes into nine subtypes using unsupervised learning and achieved an AUC of 0.95 for the classification of MPN. However, these complex subtypes of megakaryocytes may present challenges in interpreting model results and the overlapping application of algorithm models. Subsequently, they also conducted quantitative and distributional analysis of reticulin fibrosis to differentiate MPN, achieving an AUC of 0.82. However, when they combined fibrosis analysis with the analysis of megakaryocyte morphology, the accuracy of MPN detection decreased slightly compared to the analysis of megakaryocytes alone (AUC 0.94 *vs.* 0.96).[28] Compared to classifiers in other studies, our classifier incorporated more comprehensive bone marrow features and gave results in close agreement with the clinical observations made by hematopathologists; thus, this method exhibits high accuracy and clinical interpretability. Such advancements demonstrate that the model might assist hematopathologists in classifying suspected MPN in the future, particularly in regions in which experts are in short supply.

Recent studies have also attempted to recognize MPN using other indicators, such as complete blood counts, peripheral blood smears, and bone marrow smears, with AUC values between 0.8 and 0.9.[29-31] However, their clinical application requires further validation. Another study used platelet transcriptome data for MPN diagnosis, with an AUC of 0.95.[32] Nevertheless, diagnosing MPN based on transcriptome analysis is costly and offers limited clinical benefit. We also established a clinical classification model which was inferior to the bone marrow and a comprehensive classification model in distinguishing MPN subtypes and non-neoplastic samples, suggesting the necessity of bone marrow pathological features in diagnosis. Compared with the bone marrow model, the comprehensive model performed better in differentiating ET and pre-PMF from other samples but had a slightly weak performance in differentiating PV and PMF. The observed phenomenon may be attributed to the more distinctive bone marrow

characteristics of PV and PMF. Additionally, the absence of erythropoietin levels among the clinical indicators in our model could have affected its performance, as decreased serum erythropoietin is a significant marker in diagnosing PV. Our study has some limitations. The uncertainty in recognizing megakaryocyte nuclear morphology and the relatively uneven sample sizes across different MPN subtypes may impact our model in differentiating performance, especially for pre-PMF. New methods for precise recognition of megakaryocyte nuclear morphology need to be developed. Furthermore, this narrow model can possibly be applied when MPN is clinically or histologically suspected. We will expand samples to refine our model and integrate it within clinical settings to verify its practical utility.

In conclusion, this platform exhibits considerable proficiency in the quantitative analysis of BMT sections from patients with MPN and high accuracy in diagnosis and classification. It has the potential to be a valuable aid for hematopathologists who are evluating patients suspected of having an MPN.

Disclosures
No conflicts of interest to disclose.

Contributions
DY, LZ and QS designed this study. DY, HZ, YS and YT collected samples. DY, HZ, FZ and ZW performed statistical analyses and developed the machine learning algorithms. DY and HZ drafted the manuscript. All authors reviewed and approved the final manuscript.

Data-sharing statement
The datasets used in this study are available from the corresponding author upon reasonable request.

References

1. Sabattini E, Pizzi M, Agostinelli C, et al. Progression in Ph-chromosome-negative myeloproliferative neoplasms: an overview on pathologic issues and molecular determinants. Cancers (Basel). 2021;13(21):5531.
2. Khoury JD, Solary E, Abla O, et al. The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: myeloid and histiocytic/dendritic neoplasms. Leukemia. 2022;36(7):1703-1719.
3. Barbui T, Thiele J, Passamonti F, et al. Survival and disease progression in essential thrombocythemia are significantly influenced by accurate morphologic diagnosis: an international study. J Clin Oncol. 2011;29(23):3179-3184.
4. Szuber N, Mudireddy M, Nicolosi M, et al. 3023 Mayo Clinic patients with myeloproliferative neoplasms: risk-stratified comparison of survival and outcomes data among disease subgroups. Mayo Clin Proc. 2019;94(4):599-610.
5. Rumi E, Boveri E, Bellini M, et al. Clinical course and outcome of essential thrombocythemia and prefibrotic myelofibrosis according to the revised WHO 2016 diagnostic criteria. Oncotarget. 2017;8(60):101735-101744.
6. Guglielmelli P, Pacilli A, Rotunno G, et al. Presentation and outcome of patients with 2016 WHO diagnosis of prefibrotic and overt primary myelofibrosis. Blood. 2017;129(24):3227-3236.
7. Arber DA, Orazi A, Hasserjian RP, et al. International Consensus Classification of Myeloid Neoplasms and Acute Leukemias: integrating morphologic, clinical, and genomic data. Blood. 2022;140(11):1200-1228.
8. Tefferi A. Primary myelofibrosis: 2023 update on diagnosis, risk-stratification, and management. Am J Hematol. 2023;98(5):801-821.
9. Barbui T, Thiele J, Gisslinger H, et al. The 2016 WHO classification and diagnostic criteria for myeloproliferative neoplasms: document summary and in-depth discussion. Blood Cancer J. 2018;8(2):15.
10. Bhargava R, Madabhushi A. Emerging themes in image informatics and molecular analysis for digital pathology. Annu Rev Biomed Eng. 2016;18:387-412.
11. Pantanowitz L. Digital images and the future of digital pathology. J Pathol Inform. 2010;2:36.
12. Kumar N, Gupta R, Gupta S. Whole slide imaging (WSI) in pathology: current perspectives and future directions. J Digit Imaging. 2020;33(4):1034-1040.
13. Niazi MKK, Parwani AV, Gurcan MN. Digital pathology and artificial intelligence. Lancet Oncol. 2019;20(5):e253-e261.
14. Ghosh A, Sirinukunwattana K, Khalid Alham N, et al. The potential of artificial intelligence to detect lymphovascular invasion in testicular cancer. Cancers (Basel). 2021;13(6):1325.
15. Yaghjyan L, Austin-Datta RJ, Oh H, et al. Associations of reproductive breast cancer risk factors with breast tissue composition. Breast Cancer Res. 2021;23(1):70.
16. Haghighat M, Browning L, Sirinukunwattana K, et al. Automated quality assessment of large digitised histology cohorts by artificial intelligence. Sci Rep. 2022;12(1):5002.
17. Salah HT, Muhsen IN, Salama ME, Owaidah T, Hashmi SK.

Machine learning applications in the diagnosis of leukemia: current trends and future directions. Int J Lab Hematol. 2019;41(6):717-725.

18. Arber DA, Orazi A, Hasserjian R, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood. 2016;127(20):2391-2405.

19. Qin X, Zhang Z, Huang C, Dehghan M, Zaiane OR, Jagersand M. U2-Net: going deeper with nested U-structure for salient object detection. Pattern Recognition. 2020;106(11):107404.

20. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016. p. 770-778.

21. Valanarasu JMJ, Patel VM. UNeXt: MLP-based rapid medical image segmentation network. In: Wang L, Dou Q, Fletcher PT, Speidel S, Li S. (eds) Medical Image Computing and Computer Assisted Intervention – MICCAI 2022. Lecture Notes in Computer Science, vol 13435. Springer, Cham; 2022. p. 23-33.

22. Saha PK, Logofatu D. Efficient approaches for db. In: Maglogiannis I, Macintyre J, Iliadis L. (eds) Artificial Intelligence Applications and Innovations. AIAI 2021. IFIP Advances in Information and Communication Technology, vol 627. Springer, Cham; 2021. p. 184-195.

23. Wong J, Jackson R, Chen L, et al. Determination of age-dependent bone marrow normocellularity. Am J Clin Pathol. 2024;161(2):170-176.

24. Zarella MD, Rivera Alvarez K. High-throughput whole-slide scanning to enable large-scale data repository building. J Pathol. 2022;257(4):383-390.

25. Alvarez-Larrán A, Ancochea A, García M, et al. WHO-histological criteria for myeloproliferative neoplasms: reproducibility, diagnostic accuracy and correlation with gene mutations and clinical outcomes. Br J Haematol. 2014;166(6):911-919.

26. D'Abbronzo G, D'Antonio A, De Chiara A, et al. Development of an artificial-intelligence-based tool for automated assessment of cellularity in bone marrow biopsies in Ph-negative myeloproliferative neoplasms. Cancers (Basel). 2024;16(9):1687.

27. Sirinukunwattana K, Aberdeen A, Theissen H, et al. Artificial intelligence-based morphological fingerprinting of megakaryocytes: a new tool for assessing disease in MPN patients. Blood Adv. 2020;4(14):3284-3294.

28. Ryou H, Sirinukunwattana K, Aberdeen A, et al. Continuous indexing of fibrosis (CIF): improving the assessment and classification of MPN patients. Leukemia. 2023;37(2):348-358.

29. Kimura K, Ai T, Horiuchi Y, et al. Automated diagnostic support system with deep learning algorithms for distinction of Philadelphia chromosome-negative myeloproliferative neoplasms using peripheral blood specimen. Sci Rep. 2021;11(1):3367.

30. Gunčar G, Kukar M, Notar M, et al. An application of machine learning to haematological diagnosis. Sci Rep. 2018;8(1):411.

31. Meggendorfer M, Walter W, Haferlach C, Kern W, Haferlach T. Deep learning algorithms support distinction of PV, PMF, and ET based on clinical and genetic markers. Blood. 2017;130(Supplement 1):4223.

32. Shen Z, Du W, Perkins C, et al. Platelet transcriptome identifies progressive markers and potential therapeutic targets in chronic myeloproliferative neoplasms. Cell Rep Med. 2021;2(10):100425.