

Review

Applications of Artificial Intelligence in Philadelphia-Negative Myeloproliferative Neoplasms

Basel Elsayed ^{1,*}, Amgad M. Elshoeibi ¹, Mohamed Elhadary ¹, Khaled Ferih ¹, Ahmed Adel Elsabagh ¹, Alaa Rahhal ², Mohammad Abu-Tineh ³, Mohammad S. Afana ³, Mohammed Abdulgayoom ³ and Mohamed Yassin ^{3,*}

¹ College of Medicine, QU Health, Qatar University, Doha 2713, Qatar

² Pharmacy Department, Heart Hospital, Hamad Medical Corporation (HMC), Doha 3050, Qatar

³ Hematology Section, Medical Oncology, National Center for Cancer Care and Research (NCCCR), Hamad Medical Corporation (HMC), Doha 3050, Qatar

* Correspondence: be1905231@qu.edu.qa (B.E.); yassinmoha@gmail.com (M.Y.)

Abstract: Philadelphia-negative (Ph-) myeloproliferative neoplasms (MPNs) are a group of hematopoietic malignancies identified by clonal proliferation of blood cell lineages and encompasses polycythemia vera (PV), essential thrombocythemia (ET), and primary myelofibrosis (PMF). The clinical and laboratory features of Philadelphia-negative MPNs are similar, making them difficult to diagnose, especially in the preliminary stages. Because treatment goals and progression risk differ amongst MPNs, accurate classification and prognostication are critical for optimal management. Artificial intelligence (AI) and machine learning (ML) algorithms provide a plethora of possible tools to clinicians in general, and particularly in the field of malignant hematology, to better improve diagnosis, prognosis, therapy planning, and fundamental knowledge. In this review, we summarize the literature discussing the application of AI and ML algorithms in patients with diagnosed or suspected Philadelphia-negative MPNs. A literature search was conducted on PubMed/MEDLINE, Embase, Scopus, and Web of Science databases and yielded 125 studies, out of which 17 studies were included after screening. The included studies demonstrated the potential for the practical use of ML and AI in the diagnosis, prognosis, and genomic landscaping of patients with Philadelphia-negative MPNs.

Keywords: artificial intelligence; deep learning; machine learning; convolutional neural networks; clinical decision support system; myeloproliferative neoplasms; diagnosis; prognosis; genomics



Citation: Elsayed, B.; Elshoeibi, A.M.; Elhadary, M.; Ferih, K.; Elsabagh, A.A.; Rahhal, A.; Abu-Tineh, M.; Afana, M.S.; Abdulgayoom, M.; Yassin, M. Applications of Artificial Intelligence in Philadelphia-Negative Myeloproliferative Neoplasms.

Diagnostics **2023**, *13*, 1123.

<https://doi.org/10.3390/diagnostics13061123>

Academic Editors: Jae-Ho Han and Ahsan Khandoker

Received: 25 December 2022

Revised: 14 February 2023

Accepted: 13 March 2023

Published: 16 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Philadelphia chromosome-negative (Ph-) myeloproliferative neoplasms (MPNs) are a group of disorders characterized by acquired mutations within the JAK-STAT signaling pathways of bone marrow stem cells, resulting in the excessive proliferation of red blood cells, white blood cells, or platelets [1,2]. According to the World Health Organization (WHO) classifications, essential thrombocythemia (ET), polycythemia vera (PV), and primary myelofibrosis (PMF) are the three classical Philadelphia-negative MPNs [3,4]. They share clinical and laboratory characteristics that can make them difficult to distinguish, especially in the initial stages of the disease [4,5]. Mutations in one of three genes, JAK2, CALR, and MPL, are found in more than 90% of MPNs, with the remaining 10% of cases identified as triple-negative MPNs [6]. JAK2 mutations, most notably JAK2V617F, are found in almost all cases of PV and more than 60% of ET and PMF, whereas CALR and MPL mutations are found in ET and PMF [7–9]. Ph-negative MPNs are associated with high morbidity, reduced quality of life, and decreased overall survival [10]. Accurate categorization of MPNs is essential for effective management since treatment goals and progression risk vary across the diseases [9,11]. Major cardiovascular adverse events and thrombosis are significant contributors to death in patients with Ph-negative MPNs [12,13]. Additionally, patients incur the risk of developing leukemia and fibrosis [14,15].

Artificial intelligence (AI) refers to computer programs that simulate and imitate human intellect, including learning and problem-solving [16]. Machine learning is a branch of artificial intelligence that involves automatically identifying patterns in data [17]. As opposed to having the behavior explicitly written, AI learns to do a task automatically from experience (i.e., data) [18]. Since the turn of the century, advances in computer power and access to ever-growing data repositories have allowed artificial intelligence (AI) to advance rapidly [19]. AI provides numerous potential tools to doctors in general and especially in the field of hematology to better inform diagnosis, prognosis, therapy planning, and basic knowledge relating to malignant hematology [20]. Moreover, subjective histological assessments are currently a significant component of the categorization scheme crucial to the diagnosis of MPNs and various other human cancers [3,4,7]. Recent advancements in computer image analysis have the potential to revolutionize the conventional morphological evaluation of human tissues and can replace or supplement the categorization systems now used in cancers [21,22].

This review summarizes the current literature on the use of artificial intelligence (AI) and machine learning (ML) algorithms in the diagnosis, prognosis, and genomics of Philadelphia-negative myeloproliferative neoplasms (MPNs). The results section is divided into three sub-sections, each exploring a different aspect of MPN research. Section 3.1 discusses the use of AI and ML in MPN diagnosis, including detection, classification, and differentiation between MPN subtypes as well as the role of genomics. Section 3.2 examines the potential of AI and ML in predicting disease outcomes, such as myelofibrosis progression risk and thrombosis, as well as response to therapy using clinical criteria and genomics. This review aims to present an overview of the outcomes, limitations, and future research needs for each of the proposed models.

2. Methods

Our search strategy (Figure 1) was developed in PubMed/MEDLINE using title/abstract keywords. For Philadelphia-negative MPNs, we included terms such as “myeloproliferative”, “Polycythemia”, “Myelofibrosis”, and “Thrombocythemia”. To review the use of AI in MPNs, we also included terms such as “AI”, “deep-learning”, and “machine-learning”. The combined search was as follows: (“polycythemia”[tiab] OR “myelofibrosis”[tiab] OR “thrombocythemia”[tiab] OR “myeloproliferative”[tiab]) AND (“AI”[tiab] OR “machine learning”[tiab] OR “deep learning”[tiab]). The search was not restricted by language or timeframe. The Polyglot translator was used to convert the initial search strategy to Embase, Web of Science, and Scopus [23]. All the studies identified were moved into EndNote 20 and, subsequently, Rayyan to remove any duplicates [24,25].

This review included research articles that discuss the use of AI and ML algorithms in Philadelphia-negative MPNs in humans. Studies were excluded from our study if they had the following attributes: (1) animal studies, (2) reviews, and/or (3) different outcomes.

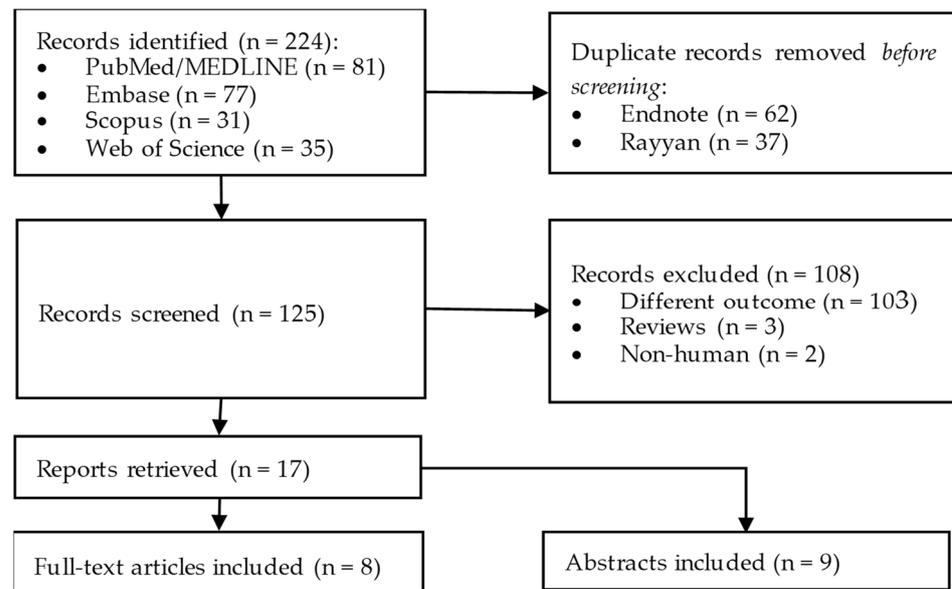


Figure 1. Schematic representation of the literature review process.

3. Results

Eight full-text articles and nine conference abstracts were included in our study. Studies were summarized and categorized under two categories: Diagnosis and Prognosis. A summary of outcomes, advantages, disadvantages, AI/ML models and their uses for the included full-text articles can be viewed in Table 1.

Table 1. Summary of included full-text articles.

Study	Outcome	Advantages	Limitations	Model(s)	Model(s) Uses
Sirinukunwattana et al. [26]	Automated analysis of megakaryocytes can categorize MPNs and differentiate them from reactive BM samples	Fast assessment of sequential BM samples Comprehensive summary of megakaryocytic cells	Marrow cellularity, lineage maturation, degree of fibrosis, and blast cell estimation are required for MPNs WHO classification	Unsupervised Learning: Principal Component Analysis (PCA)	Reduction of high dimensional data Exploratory data analysis and visualization of complex datasets
Kimura et al. [27]	Automated diagnostic support system for MPNs using peripheral blood (PB) specimen	Fast assessment of PB specimens Accurate differentiation of PV, ET, and MF	Single-center study Small number of cases	Deep Learning: Convolutional Neural Network (CNN)	Image recognition and classification Learns features automatically from raw data
Asaulenko et al. [28]	Histotopographical features of megakaryocytes allowed correct differentiation between ET and PMF in 71.6% of cases	Patterns of megakaryocyte distribution in BM of ET and prePMF patients with JAK2/CALR mutations can be revealed only using ML	The percentage of correct diagnostic predictions for PMF was only 40%	Unsupervised Learning: Density-Based Spatial Clustering of Applications with Noise (DBSCAN)	Clustering and anomaly detection in high-dimensional data Automatically detect clusters of arbitrary shape and size
Kantardzic et al. [29]	Extraction of new decision rules for PV diagnosis, based on a reduced and optimized set of lab parameters	Reducing the original parameters of diagnosis to only four while still obtaining good classification results	Not diagnostic by its own Only complements the standard PVSG criteria	Supervised Learning: Artificial neural networks (ANNs) and Support vector machines (SVMs)	Classification and regression tasks Both require large amounts of labeled data for training

Table 1. Cont.

Study	Outcome	Advantages	Limitations	Model(s)	Model(s) Uses
Shen et al. [30]	Progressive platelet transcriptomic markers, enable an externally validated prediction for advanced MPNs	Comprehensive catalog of platelet transcriptome in chronic MPNs Accurate prediction of MF using <5 candidate markers	Only analyzed platelet-derived molecular alterations Biological and computational validations are needed for decision making	Supervised Learning: Multiple LASSO (Least Absolute Shrinkage and Selection Operator) penalized regression classifiers	Classification and prediction tasks, especially when there are more features than observations
Ryou et al. [31]	Continuous Indexing of Fibrosis (CIF) enhances the detection and monitoring of fibrosis within BMTs and aids MPN subtyping	Accurate discrimination between ET and Pre-PMF Identification of MPN patients at risk of progression	Invasive technique Requires BM trephines	Ranking-CNN: Learning to Rank (LTR) and Convolutional Neural Network (CNN)	Ranking and recommendation tasks LTR models rank items based on relevance, while CNNs extract features from raw data
Angelopoulos et al. [32]	Discovered genomic sets and their relationships in MPNs using Bayesian networks (BNs)	BNs allow correlations among driver events in large genomic cohorts Graphical illustrations	Robustness of the networks is important for BN learning Not suitable for small datasets	Probabilistic Graphical: Bayesian networks	Probabilistic inference and decision-making under uncertainty Models' complex relationships between variables
Li et al. [33]	EPB42, CALR, SLC4A1 and MPL are candidate prognostic biomarkers and potential therapeutic targets for early PMF	WGCNA, a powerful global research tool for data mining from multiple genes in large-scale datasets was used	Molecular biological studies and larger samples are needed to further validate these hub genes	Supervised Learning: Support vector machines (SVMs)	Classification and prediction tasks, especially when there are more features than observations

3.1. Diagnosis

Section 3.1 will focus on the use of AI and ML in the diagnosis of MPNs, including the detection, classification, and differentiation between subtypes of MPNs. This section will explore the potential of these technologies in improving the accuracy and efficiency of MPN diagnosis.

3.1.1. Diagnosis of MPNs Using Bone Marrow and Peripheral Blood Specimens

Sirinukunwattana et al. [26] created a machine-learning method for automated identification, quantitative measurement, and abstract representation of megakaryocyte characteristics utilizing reactive/nonneoplastic bone marrow trephines (BMT) and those from patients with confirmed MPN diagnosis. They discuss the use of an automated method for identifying and delineating key histological characteristics in routinely generated BMTs. Following analysis, tissue diagnosis of MPN was possible with high predictive accuracy (AUC = 0.95), and significant evidence of the ability to distinguish between major MPN subtypes was shown. The machine-learning algorithms described have numerous significant advantages over traditional histology analysis. To begin with, a totally automated model can deliver a rapid and reliable preliminary diagnostic assessment of specimens before official pathology reporting, which is likely to be valuable if access to hematopathology expertise is restricted (for example, in low-resource health care systems). Second, a detailed and interpretable description of the megakaryocytic population will enable the pathologist to concentrate on integrating broader pathological elements with clinical and laboratory data. Third, this approach is well-suited for a more accurate evaluation of consecutive specimens from patients undergoing therapy and/or recurrent examination.

Kimura et al. [27] established a clinical decision support system for Ph-negative MPNs to minimize workload and inter- and intra-personal discrepancies (Figure 2). The technique involved combining the complete blood counts (CBCs) and research data collected by an automated hematological analyzer (Sysmex XN-9000) with peripheral blood (PB) smear

morphological features retrieved using a recently developed Convolutional Neural Network (CNN) coupled with an Extreme Gradient Boosting (XGBoost)-based decision-making algorithm [34,35]. In brief, a deep learning system (DLS) was trained using 695,030 normal and pathological cell pictures, 174 XN-9000 parameters, and 114 cell morphological parameters through a CNN-based image-recognition system. The DLS could successfully categorize 17 cell subtypes and detect 97 aberrant morphological characteristics after training. The diagnostic-aid algorithm was also trained to distinguish between 23 polycythemia vera (PV), 101 essential thrombocythosis (ET), and 36 myelofibrosis (MF) cases. The system was evaluated utilizing samples from 9 PV, 53 ET, and 12 MF patients and demonstrated reliable performance in distinguishing PV, ET, and MF with high accuracy when compared to human diagnoses, specifically with >90% sensitivity and >90% specificity. The computed area under the curve of the ROC curves for PV, ET, and MF, respectively, were 0.990, 0.967, and 0.974. This work adds to previous AI-based studies utilizing mutational data, peripheral blood specimens, and bone marrow trephines for a more accurate diagnosis of MPNs [36–38].

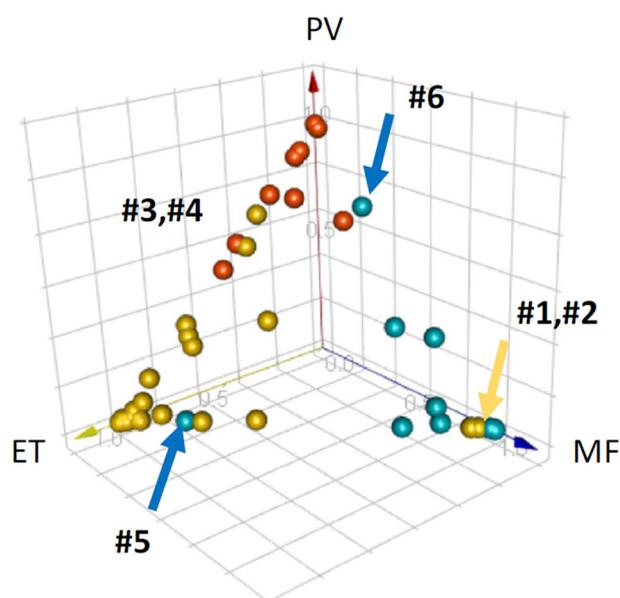


Figure 2. 3D plot of the predicted likelihood of Kimura et al.'s automated diagnostic support system with deep learning algorithms for the distinction of myeloproliferative neoplasms. Six misclassified cases are labeled #1–#6.

3.1.2. Differential Diagnosis of PMF and ET Using Megakaryocytic Lineage

Asaulenko et al. [28] created machine learning algorithms capable of distinguishing morphological and histotopographical aspects of bone marrow samples in order to increase the accuracy of essential thrombocythemia (ET) and prefibrotic primary myelofibrosis (prePMF) differential diagnosis. The Density-Based Spatial Clustering of Applications with Noise (DBSCAN) clustering technique was developed in-house and was used to investigate particular histotopographical aspects of the megakaryocytic lineage in bone marrow samples from 95 patients with JAK2- or CALR-mutated ET and prePMF. The percentage of correct predictions for ET (90%) was much greater than prePMF (40%). The overall correct diagnostic probability was 71.6%. By building logistic regression models with 75% and 25% of the source dataset, the model was verified in a training and testing sample. The ROC was constructed, and the area under the curve was 72.5%. The discrepancies found by the DBSCAN clustering technique suggest that there may be a connection between the MPN subtype and the development of the megakaryocytic lineage. The results of this study agree with a recent study by Yassin et al. [39] assessing bone marrow activity using 3'-18Fluoro-3'-deoxy-L-thymidine (18F-FLT) with positron emission tomography (PET) to differentiate between patients with essential thrombocythemia and

prefibrotic primary myelofibrosis. The findings demonstrate that the characteristics of the bone marrow in MPNs differ in a manner that cannot be visually observed during microscopic analysis and can only be detected using machine learning and advanced automated imaging techniques.

3.1.3. Reduced and Optimized PV Diagnosis Rules

In 2002, polycythemia vera was often diagnosed using the widely used Polycythemia Vera Study Group (PVSG) diagnostic criteria [40–42]. Kantardzic et al. [29] introduced a data-mining method to derive new decision criteria for polycythemia vera diagnosis based on a reduced and optimized subset of lab characteristics. This study included several laboratory and clinical findings from 431 polycythemia vera patients and 91 patients with other myeloproliferative disorders within the original dataset used in the PVSG study. They demonstrated that it is feasible to reach a diagnostic conclusion about polycythemia vera with the same degree of classification quality while utilizing only four parameters: hematocrit, platelet count, splenomegaly, and white blood cell count, using typical data-mining approaches such as artificial neural networks, support vector machines, and n-dimensional visualization. The categorization of polycythemia vera patients using a trained artificial neural network or a support vector machine compared to PVSG diagnostic criteria, the gold standard for polycythemia vera diagnosis at the time, did not show any significant differences. Therefore, these reduced and optimized diagnostic criteria could be utilized in addition to the conventional PVSG criteria, notably in differentiating polycythemia vera from other myeloproliferative syndromes.

3.1.4. Diagnosis of JAK2 V617F Negative Patients with WHO-Defined ET

Smart Blood Analytics (SBA) is a machine learning-based application that examines complete blood counts (CBCs) to generate a list of probable diagnoses. It is primarily proposed for medical professionals who lack a thorough grasp of hematology and has been proven to be useful in diagnosing hematological illnesses [43]. Belcic et al. [44] further evaluated the SBA algorithm's applicability in 237 patients referred for suspected essential thrombocythemia (ET). They classified individuals as having ET according to WHO 2008 or modified WHO criteria. All 237 patients were then assessed using the SBA algorithm. The SBA algorithm detected ET patients with 100% sensitivity and 50% specificity. The SBA machine-learning method may be effective in situations where expert consultation is not accessible.

3.1.5. Supervised Classification of MPNs

Skov et al. [45] studied the possibility of categorizing ET, PV, and PMF into independent categories using supervised classification techniques and looked into the possibility of overlap between the three entities. Blood specimens from 19 patients with ET, 18 with PV, and 9 with PMF were used in gene expression microarray investigations. In order to select the best multi-class classification technique, 11 support vector machine-learning algorithms were used in the voting process. Five gene sets were utilized to train the models, with one set left out for cross-validation. The categorization techniques were established to have optimal accuracy and significance level. In summary, 15 patients (79%) with ET, 28 patients (68%) with PV, and 8 patients (89%) with PMF were accurately predicted by the support vector algorithms. The categorization models have a 79% overall accuracy and performed quite well statistically ($p = 1 \times 10^{-6}$). These results suggested that ET, PV, and PMF might be classified as discrete disease entities, with overlap between ET and PV providing support to a biological continuum from ET through PV to PMF. Furthermore, this supervised multi-classifier might help with the differential diagnosis of MPNs.

3.1.6. Bayesian Networks Elucidate Complex Genomic Landscapes in MPNs

Bayesian networks (BNs) are artificial intelligence (AI) models that can define complex joint probability regions. Angelopoulos et al. [32] used the BN model to analyze genetic

data from 2035 individuals with myeloproliferative neoplasms and discovered a number of genomic groupings in myeloproliferative neoplasms (Figure 3). For example, mutations in TP53 co-occurred with chromosomes 17p and 5q29 deletions. Another grouping, characterized by LOH at chromosome 4q, abnormalities in chromosomes 7 and 7q occurring in conjunction with mutations in at least 14 myeloid malignancies, accounting for the largest co-mutation pattern in the Bayesian network, was enriched for patients with PMF. Moreover, JAK2 was identified as the most mutated gene in the network, with CALR and MPL mutations exhibiting reciprocal exclusivity patterns, demonstrating functional redundancy in their pathogenic processes. The three genes (JAK2, CARL, and MPL) accounted for most of the dataset's total driver mutations. The BN model beautifully depicts the major players' three-way mutual exclusivity. Moreover, the genomic landscapes elucidated by the BNs are supported by clinical exome sequencing studies assessing driver mutations in MPNs [46,47]. Understanding the genomic landscape of MPNs using AI may prove useful in developing diagnostic panels that complement and confirm clinical diagnoses.

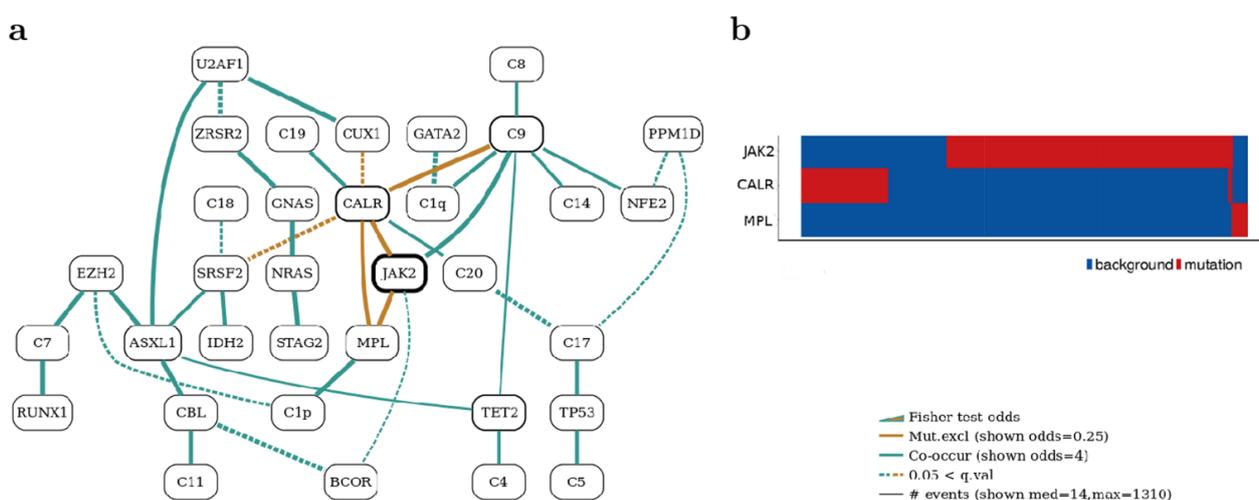


Figure 3. (a) Angelopoulos et al.'s Bayesian Network of the myeloproliferative neoplasms dataset; (b) Heatmap plots for presence (red) or absence (blue) of driver mutations in CALR, JAK2, and MPL genes on the *y*-axis against patients on the *x*-axis.

3.1.7. Distinction of MPNs Using Genetic Markers

Meggendorfer et al. [48] combined genetic profiling using AI-augmented next-generation sequencing (NGS) with clinical parameters to differentiate and classify MPNs. Firstly, deep-learning algorithms were applied to NGS and allowed the exploration of 15 different genetic markers in addition to JAK2, CALR, and MP. Secondly, clinical parameters were combined with the 18 genes and applied to a training cohort of 243 Ph-negative MPNs using support vector machines (SVM). Individual models for PMF, PV, and ET were created using threefold cross-validation. Then, the trained models were utilized to estimate the most likely diagnosis of patients from a testing cohort, including 183 MPN patients, with the class probabilities for each patient retrieved. Overall, an accuracy of 96% was achieved, demonstrating the potential of deep learning to assist physicians by making the diagnosis based on the patient's genetic background, even if certain clinical criteria were absent.

3.1.8. Random Forest Classifier for Predicting MPN Subtype Using Genomics

Jabalameli et al. [49] used machine-learning methods to analyze germline genetic variations and built a robust prediction model for ET and PV patients. Using a patient's genetic data; a Random Forest Classifier was used to predict MPN subtypes. The Illumina Human OmniExpressExome-v1.2 was used to genotype 499 JAK2+ve ET cases and 505 JAK2+ve PV cases. Single nucleotide polymorphisms (SNPs) with a minor allele frequency (MAF) of <0.01 , SNPs and samples with more than 10% missing genotypes, and SNPs deviating from

Hardy–Weinberg Equilibrium (HWE) proportions were all removed for quality control. SNPs with $r^2 > 0.2$ within a 50 kb frame were eliminated to account for multi-collinearity and avoid overfitting. For data visualization and categorization, Principal component analysis (PCA) was employed as an unsupervised technique. To build the classifier model, 7144 SNPs with significant alternative allele relationships ($p > 0.05$) between ET and PV were employed. The MPN ensemble classifier model proved extremely accurate in predicting MPN subtypes. Using fivefold cross-validation, the Random Forest model (AUC = 90%) outperformed the Decision Tree model (AUC = 87%) in terms of accuracy. The findings show that germline variation information may be used to predict the MPN subtype using machine-learning modeling.

3.2. Prognosis

Section 3.2 will examine the potential of AI and ML in predicting disease outcomes, such as the risk of myelofibrosis progression and thrombosis, as well as response to therapy. This section will explore the potential of these technologies in guiding personalized treatment decisions for MPN patients.

3.2.1. Myelofibrosis Prediction Using Platelet Transcriptome of MPNs

The capacity to anticipate how a patient's condition will develop or change over time, from essential thrombocythemia (ET) or polycythemia vera (PV) diagnoses to secondary myelofibrosis (SMF), is currently restricted by an understanding of genetic, cytogenetic, or epigenetic variances in myeloproliferative neoplasms. [50,51]. Shen et al. [30] built multiple LASSO (Least Absolute Shrinkage and Selection Operator) penalized regression classifiers [52] that use machine learning to evaluate platelet transcriptome in order to distinguish MPN subtypes and to allow MF prediction (Figure 4). Aside from a training set, two robust external validation criteria [53] were used: temporal validation with two independent cohorts and geographical validation using independently published datasets on healthy donors, and patients with MF [54,55]. The best-performing model employed a collection of progressively differentiated MPN genes and identified five potential genetic markers as the top predictors of disease development. Essentially, LASSO regression and progressive transcriptome studies yielded a compact gene signature that might be used in a PCR-based predictive assessment and performed in clinical laboratories.

3.2.2. Continuous Index of Fibrosis in MPNs

In MPNs, fibrosis grading is a significant component of disease categorization, prognosis, and surveillance [3]. Ryou et al. [31] devised a machine-learning (ML) approach to objectively measure bone marrow reticulin fibrosis and enhance fibrosis grading (Figure 5). Manually labeled areas of fibrosis were utilized to train the model, yielding a continuous index of fibrosis (CIF) ranging from 0 to 1. The projected scores of new samples were then translated into a quantitative fibrosis map using artificial intelligence, which was subsequently superimposed over entire sample images. The whole spectrum and variety of fibrosis among established MPN and normal/reactive BMT samples are captured in analyses of MPN sample cohorts. Additionally, when paired with megakaryocyte analysis, CIF has good predictive accuracy (AUC = 0.94) in distinguishing between the typically difficult differential diagnosis of ET and pre-fibrotic MF. CIF also showed promise in identifying MPN patients at risk of disease progression. When CIF was tested on a study of 35 ET patients engaged in the Primary Thrombocythemia-1 trial, the model discovered characteristics predictive of post-ET MF with relatively high accuracy (AUC = 0.77) [56].

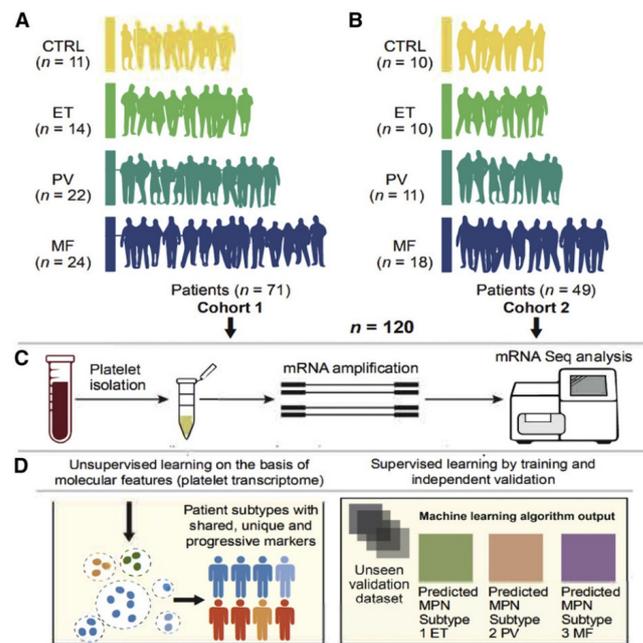


Figure 4. Graphical abstract of Shen et al.'s machine learning algorithm for myelofibrosis prediction using platelet transcriptome of myeloproliferative neoplasms. (A,B) Two independent cohorts of MPN patients and healthy controls were included. (C) Platelets from peripheral blood samples of the two cohorts were isolated and sequenced to yield a platelet transcriptome. (D) Machine learning approaches, including unsupervised and supervised techniques, were utilized for myelofibrosis prediction using the platelet transcriptome.

3.2.3. Predicting Risk of Thrombosis in PV Using Clinicopathologic Features

Using the most significant of 60 clinicopathologic variables, Abu-Zeinah et al. [57] developed a machine-learning algorithm-based clinical decision support system that predicts the incidence of thrombosis in polycythemia vera patients. Prediction of arterial or venous thrombosis in the following 3 months was performed using a random survival forest (RSF) model. A cohort of 470 patients with polycythemia vera was obtained from Weill Cornell Medicine's database and divided into training (75%) and testing (25%). The grading system was created using the most important ML-derived characteristics and validated using multivariable logistic regression analysis. Using the Fine-Gray model, the cumulative incidence (CI) of thrombosis was compared between risk categories. Five of the top ten variables that independently predicted thrombosis within the RSF model were included in a clinical decision support system that quantified thrombosis risk. A point was allocated for the presence of each of the five variables: age, old thrombosis, recent thrombosis, leukocytosis, and recent diagnosis. The cohort was then labeled as low, intermediate, and high-risk if their score in the grading system was 1/5, 2/5, or $\geq 3/5$, respectively. The grading method revealed that high-risk and intermediate-risk individuals were 6.5 and 2.3 times more likely to develop thrombosis, respectively, than low-risk patients, and the likelihood of thrombosis was substantially different after 1, 2, and 5 years.

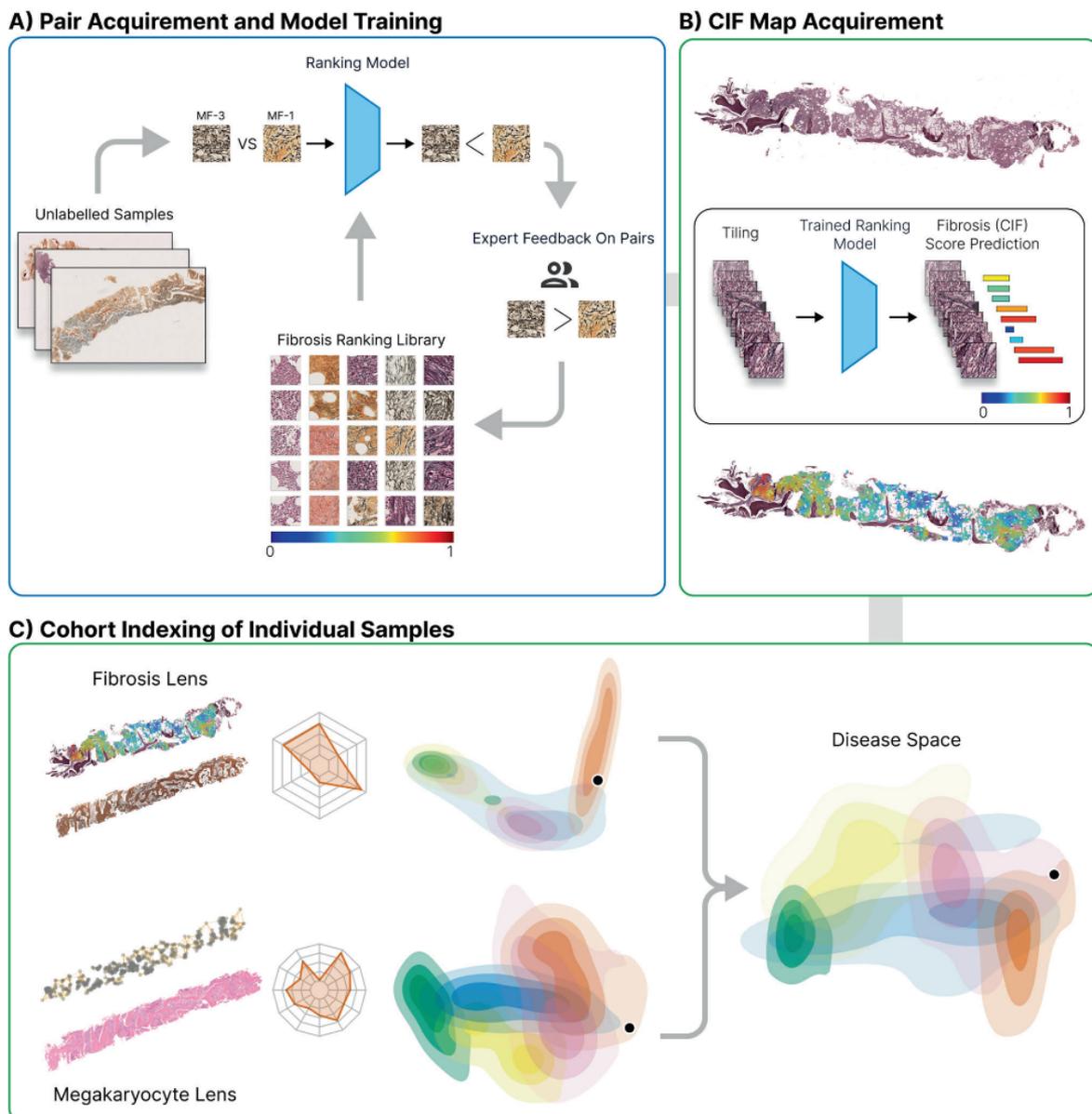


Figure 5. Ryou et al.’s computational steps for detection, quantitation, and visualization of reticulin fibrosis in BMTs. **(A)** Image tiles are labeled with fibrosis grade and used to train a ranking-CNN model that was refined using a human-in-the-loop approach. **(B)** The finalized model generates CIF maps by predicting scores and overlaying them on reticulin-stained images. **(C)** Fibrosis features are represented in a two-dimensional disease space, enabling comparison of MPN subtypes and individual patient BMTs against a sample library.

3.2.4. Predicting Fibrosis in PV Using Accessible Baseline Characteristics

By using easily accessible baseline parameters, Srisuwananukorn et al. [58] created a machine-learning algorithm that predicts the development of myelofibrosis in polycythemia vera (PV) at the time of diagnosis using characteristics such as demographics, physical exam findings, and laboratory values. A total of 35 models were created after dividing a cohort of 527 PV patients into a development set (90%) and a hold-out set (10%) based on disease progression status. A set of biologically plausible characteristics were chosen using minimum depth analysis across all 35 models. The models were compared to well-known prediction models such as the European LeukemiaNet (ELN) score and the International Working Group MPN Research and Treatment (IWG-MRT) score. Age and past thrombosis

define the ELN score, while age, WBC, and prior thrombosis determine the IWG-MRT score. Three variables—age, neutrophilia, and leukocytosis—were utilized as continuous values in a final random survival forest (RSF) model. The model outperformed the ELN and the IWG-MRT scores ($p = 0.001$) and is undergoing development to become an interactive online interface.

3.2.5. Predicting Hydroxyurea Failure and Thromboembolism in PV

Verstovsek et al. [59] developed an accurate and reliable machine-learning model that forecasts the incidence of thromboembolism (TE) in polycythemia vera (PV) patients undergoing Hydroxyurea (HU) treatment. The model was trained using 69,464 polycythemia vera patients within the US OPTUM database. For prediction, a logistic regression model was constructed, and the results were transformed into therapeutically applicable decision-tree classification algorithms. Lymphocytosis and high red cell distribution width (RDW) were predictive for patients without a history of TEs, while thrombocytosis and lymphocytosis were predictors in patients with a history of TEs. Similarly, high RDW was predictive of HU failure in phlebotomy-dependent patients after three months of treatment. Additionally, PV patients who switched to Ruxolitinib had a lower incidence of TEs than those who stayed on HU.

3.2.6. Predicting Thrombosis Risk in Secondary Myelofibrosis

Mora et al. [60] examined thrombosis predictors in the MYSEC PM (Myelofibrosis Secondary to PV and ET-Prognostic Model) database. A machine-learning random survival forest (RSF) model that incorporated phenotypic and genotypic variables upon secondary myelofibrosis (SMF) diagnosis was utilized to find predictors of thrombosis incidence such as gender, age, RBC indices, WBC indices, bone marrow fibrosis, and prior thrombosis. Then, in order to assess the function of the ML algorithm in predicting thrombotic risk, a second model was developed, which aggregates MYSEC-PM components. The results have shown that the random forest model involving MYSEC-PM and past thrombosis predicted thrombotic risk following SMF evolution with great accuracy.

3.2.7. Prediction of Primary Myelofibrosis Using Gene Expression

Li et al. [33] sought to uncover unique diagnostic and predictive gene expression profiles of primary myelofibrosis (PMF) from gene expression datasets. Weighted Gene Co-expression Network Analysis (WGCNA) methods were used to find the most associated genes to PMF. Following that, Gene Ontology (GO) and Kyoto Encyclopedia Genes and Genomes (KEGG) searches were conducted in order to properly comprehend the specific information of a module of interest labeled Green. Following WGCNA analysis, module Green was substantially associated with PMF disease. Twenty genes in module Green were identified as hub genes for PMF development. EPB42, CALR, SLC4A1, and MPL were the most highly expressed and correlated with PMF, and a machine-learning model was developed to demonstrate the reliability of these genes. The SVM (support vector machine) was chosen as the best-suited prediction model and was assessed in third-party datasets, obtaining AUCs of 0.922 and 0.875 within Gene Expression Omnibus (GEO) datasets GSE53482 and GSE61629, respectively.

3.2.8. Prediction of Myelofibrosis Progression in Health Records

Bejan et al. [61] devised a system to predict myelofibrosis using Natural language Processing (NLP) machine learning with negation detection of MF keywords, drugs, and ICD coding to construct phenotype-genotype connections. JAKextractor, an algorithm to find individuals clinically tested for JAK2V617F across 248,000 cases in BioVU, a bank of de-identified DNA samples, was developed to enhance the dataset. A supervised learning system was taught to learn decision rules that encode MF-specific ICD codes, drugs, the text mentions, and the assertion status of MF and JAK2 references in patient notes for MF identification. A 10-fold cross-validation technique was used to evaluate the

experiments. JAKextractor employed pattern matching to determine the state of each JAK2 text mention (wild-type versus mutated). Based on the information gathered from patient notes, machine learning identified JAK2V617F patients. The best-performing MF algorithm included all clinical data sources and attained an F1-measure (F1) of 96%, identifying 309 MF patients in BioVU. Moreover, 71/131 (54.2%) of MF patients were genotyped with JAK2V617F, compared to 66/131 (50.4%) identified using JAKextractor. There were only two false positives and four false negatives in the JAKextractor predictions. These findings confirmed the possibility of constructing an MPN database using retrospective genotyping of biobanked DNA and demonstrated the effective identification of MF and JAK2V617F within an electronic health record.

4. Discussion

While the use of AI and ML technologies in MPNs has shown promise, several limitations should be considered. Firstly, the use of these technologies may not be diagnostic by itself and only complements the standard diagnostic criteria for MPNs, such as the requirement for marrow cellularity, lineage maturation, degree of fibrosis, and blast cell estimation, which are necessary for MPNs WHO classification. Therefore, the use of AI and ML should not be relied upon solely for diagnosis but rather as a tool to assist physicians in making a diagnosis. Secondly, many of the reviewed studies were based on single-center and a small number of cases, which may limit the generalizability of the results. Additionally, some of the reviewed studies only analyzed cell-derived molecular alterations, indicating that biological and computational validations are necessary for decision-making. Furthermore, the application of AI and ML technologies in MPNs may be based on invasive techniques that require BM trephines, which may not be suitable for all patients and therefore limit the utility of these technologies. Finally, some of the reviewed studies focused on larger datasets and molecular biology, such as the identification of hub genes or the use of Bayesian networks, which may require larger samples to validate the proposed models. Therefore, future studies should focus on addressing these limitations to fully realize the potential of AI and ML technologies in MPNs research and clinical practice.

5. Conclusions

This literature review summarizes the recent advances in the applications of Artificial Intelligence (AI) within Philadelphia-negative (Ph-) myeloproliferative neoplasms (MPNs). Machine learning, deep learning and digital hematopathology models have shown promising results in diagnosing, prognosing and explaining the complex genomics of Ph-MPNs with remarkably high accuracy and reliability. As management of the complex and overlapping MPNs depends upon correct classification, AI models might prove useful in the treatment and monitoring of patient suffering from these rare disorders. Nevertheless, in order to assess the effectiveness of these models in diverse clinical scenarios, additional research studies utilizing larger samples and improved methods are still required.

Author Contributions: Conceptualization, M.Y.; writing—original draft preparation, B.E. and M.Y.; methodology, A.M.E. and M.E.; software, B.E.; validation, A.R., M.A.-T. and M.S.A.; investigation, A.A.E., B.E. and A.M.E.; resources, M.E. and K.F.; writing—review and editing, K.F, M.E. and M.A.; supervision, M.Y.; project administration, B.E., M.Y., A.R. and A.A.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Rampal, R.; Al-Shahrour, F.; Abdel-Wahab, O.; Patel, J.P.; Brunel, J.P.; Mermel, C.H.; Bass, A.J.; Pretz, J.; Ahn, J.; Hricik, T.; et al. Integrated genomic analysis illustrates the central role of JAK-STAT pathway activation in myeloproliferative neoplasm pathogenesis. *Blood* **2014**, *123*, e123–e133. [[CrossRef](#)]
2. O'Sullivan, J.; Mead, A.J. Heterogeneity in myeloproliferative neoplasms: Causes and consequences. *Adv. Biol. Regul.* **2019**, *71*, 55–68. [[CrossRef](#)]
3. Arber, D.A.; Orazi, A.; Hasserjian, R.; Thiele, J.; Borowitz, M.J.; Le Beau, M.M.; Bloomfield, C.D.; Cazzola, M.; Vardiman, J.W. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* **2016**, *127*, 2391–2405. [[CrossRef](#)] [[PubMed](#)]
4. Barbui, T.; Thiele, J.; Gisslinger, H.; Kvasnicka, H.M.; Vannucchi, A.M.; Guglielmelli, P.; Orazi, A.; Tefferi, A. The 2016 WHO classification and diagnostic criteria for myeloproliferative neoplasms: Document summary and in-depth discussion. *Blood Cancer J.* **2018**, *8*, 15. [[CrossRef](#)] [[PubMed](#)]
5. Harrison, C.N.; McMullin, M.F.; Green, A.R.; Mead, A.J. Equivalence of BCSH and WHO diagnostic criteria for ET. *Leukemia* **2017**, *31*, 1660. [[CrossRef](#)]
6. Xie, J.; Chen, X.; Gao, F.; Hou, R.; Tian, T.; Zhang, Y.; Fan, L.; Hu, J.; Zhu, G.; Yang, W.; et al. Two activating mutations of MPL in triple-negative myeloproliferative neoplasms. *Cancer Med.* **2019**, *8*, 5254–5263. [[CrossRef](#)] [[PubMed](#)]
7. Cabagnols, X.; Favale, F.; Pasquier, F.; Messaoudi, K.; Defour, J.P.; Ianotto, J.C.; Marzac, C.; Le Couedic, J.P.; Droin, N.; Chachoua, I.; et al. Presence of atypical thrombopoietin receptor (MPL) mutations in triple-negative essential thrombocythemia patients. *Blood* **2016**, *127*, 333–342. [[CrossRef](#)] [[PubMed](#)]
8. Rumi, E.; Pietra, D.; Pascutto, C.; Guglielmelli, P.; Martinez-Trillos, A.; Casetti, I.; Colomer, D.; Pieri, L.; Pratcorona, M.; Rotunno, G.; et al. Clinical effect of driver mutations of JAK2, CALR, or MPL in primary myelofibrosis. *Blood* **2014**, *124*, 1062–1069. [[CrossRef](#)]
9. Verger, E.; Cassinat, B.; Chauveau, A.; Dosquet, C.; Giraudier, S.; Schlageter, M.H.; Ianotto, J.C.; Yassin, M.A.; Al-Dewik, N.; Carillo, S.; et al. Clinical and molecular response to interferon-alpha therapy in essential thrombocythemia patients with CALR mutations. *Blood* **2015**, *126*, 2585–2591. [[CrossRef](#)]
10. Yassin, M.A.; Taher, A.; Mathews, V.; Hou, H.A.; Shamsi, T.; Tuglular, T.F.; Xiao, Z.; Kim, S.J.; Depei, W.; Li, J.; et al. MERGE: A Multinational, Multicenter Observational Registry for Myeloproliferative Neoplasms in Asia, including Middle East, Turkey, and Algeria. *Cancer Med.* **2020**, *9*, 4512–4526. [[CrossRef](#)]
11. Allahverdi, N.; Yassin, M.; Ibrahim, M. Environmental Factors, Lifestyle Risk Factors, and Host Characteristics Associated With Philadelphia Negative Myeloproliferative Neoplasm: A Systematic Review. *Cancer Control* **2021**, *28*, 10732748211046802. [[CrossRef](#)] [[PubMed](#)]
12. Hultcrantz, M.; Bjorkholm, M.; Dickman, P.W.; Landgren, O.; Derolf, A.R.; Kristinsson, S.Y.; Andersson, T.M.L. Risk for Arterial and Venous Thrombosis in Patients With Myeloproliferative Neoplasms: A Population-Based Cohort Study. *Ann. Intern. Med.* **2018**, *168*, 317–325. [[CrossRef](#)] [[PubMed](#)]
13. Marchioli, R.; Finazzi, G.; Landolfi, R.; Kutti, J.; Gisslinger, H.; Patrono, C.; Marilus, R.; Villegas, A.; Tognoni, G.; Barbui, T. Vascular and neoplastic risk in a large cohort of patients with polycythemia vera. *J. Clin. Oncol.* **2005**, *23*, 2224–2232. [[CrossRef](#)] [[PubMed](#)]
14. Dunbar, A.J.; Rampal, R.K.; Levine, R. Leukemia secondary to myeloproliferative neoplasms. *Blood* **2020**, *136*, 61–70. [[CrossRef](#)]
15. Baumeister, J.; Chatain, N.; Sofias, A.M.; Lammers, T.; Koschmieder, S. Progression of Myeloproliferative Neoplasms (MPN): Diagnostic and Therapeutic Perspectives. *Cells* **2021**, *10*, 3551. [[CrossRef](#)]
16. Topol, E.J. High-performance medicine: The convergence of human and artificial intelligence. *Nat. Med.* **2019**, *25*, 44–56. [[CrossRef](#)]
17. Beam, A.L.; Kohane, I.S. Big Data and Machine Learning in Health Care. *JAMA* **2018**, *319*, 1317–1318. [[CrossRef](#)]
18. Jordan, M.I.; Mitchell, T.M. Machine learning: Trends, perspectives, and prospects. *Science* **2015**, *349*, 255–260. [[CrossRef](#)]
19. Shamshirband, S.; Fathi, M.; Dehzangi, A.; Chronopoulos, A.T.; Alinejad-Rokny, H. A review on deep learning approaches in healthcare systems: Taxonomies, challenges, and open issues. *J. Biomed. Inform.* **2021**, *113*, 103627. [[CrossRef](#)]
20. Rajkomar, A.; Dean, J.; Kohane, I. Machine Learning in Medicine. *N. Engl. J. Med.* **2019**, *380*, 1347–1358. [[CrossRef](#)]
21. Khosravi, P.; Kazemi, E.; Imielinski, M.; Elemento, O.; Hajirasouliha, I. Deep Convolutional Neural Networks Enable Discrimination of Heterogeneous Digital Pathology Images. *EBioMedicine* **2018**, *27*, 317–328. [[CrossRef](#)] [[PubMed](#)]
22. Bulten, W.; Bandi, P.; Hoven, J.; Loo, R.V.; Lotz, J.; Weiss, N.; Laak, J.V.; Ginneken, B.V.; Hulsbergen-van de Kaa, C.; Litjens, G. Epithelium segmentation using deep learning in H&E-stained prostate specimens with immunohistochemistry as reference standard. *Sci. Rep.* **2019**, *9*, 864. [[CrossRef](#)] [[PubMed](#)]
23. Clark, J.M.; Sanders, S.; Carter, M.; Honeyman, D.; Cleo, G.; Auld, Y.; Booth, D.; Condron, P.; Dalais, C.; Bateup, S.; et al. Improving the translation of search strategies using the Polyglot Search Translator: A randomized controlled trial. *J. Med. Libr. Assoc.* **2020**, *108*, 195–207. [[CrossRef](#)]
24. Bramer, W.M.; Giustini, D.; de Jonge, G.B.; Holland, L.; Bekhuis, T. De-duplication of database search results for systematic reviews in EndNote. *J. Med. Libr. Assoc.* **2016**, *104*, 240–243. [[CrossRef](#)]
25. Ouzzani, M.; Hammady, H.; Fedorowicz, Z.; Elmagarmid, A. Rayyan—a web and mobile app for systematic reviews. *Syst. Rev.* **2016**, *5*, 210. [[CrossRef](#)] [[PubMed](#)]

26. Sirinukunwattana, K.; Aberdeen, A.; Theissen, H.; Sousos, N.; Psaila, B.; Mead, A.J.; Turner, G.D.H.; Rees, G.; Rittscher, J.; Royston, D. Artificial intelligence-based morphological fingerprinting of megakaryocytes: A new tool for assessing disease in MPN patients. *Blood Adv.* **2020**, *4*, 3284–3294. [[CrossRef](#)]
27. Kimura, K.; Ai, T.; Horiuchi, Y.; Matsuzaki, A.; Nishibe, K.; Marutani, S.; Saito, K.; Kaniyu, K.; Takehara, I.; Uchihashi, K.; 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*, 3367. [[CrossRef](#)]
28. Asaulenko, Z.P.; Polushkina, L.B.; Lepsky, A.I.; Krivolapov, Y.A. Morphological Differential Diagnosis of Primary Myelofibrosis and Essential Thrombocythemia with Computer Cluster Analysis of a Megakaryocytic Lineage in Myeloid Tissue. *Biophysica* **2020**, *65*, 676–680. [[CrossRef](#)]
29. Kantardzic, M.; Djulbegovic, B.; Hamdan, H. A data-mining approach to improving Polycythemia Vera diagnosis. *Comput. Ind. Eng.* **2002**, *43*, 765–773. [[CrossRef](#)]
30. Shen, Z.; Du, W.; Perkins, C.; Fechter, L.; Natu, V.; Maecker, H.; Rowley, J.; Gotlib, J.; Zehnder, J.; Krishnan, A. Platelet transcriptome identifies progressive markers and potential therapeutic targets in chronic myeloproliferative neoplasms. *Cell Rep. Med.* **2021**, *2*, 100425. [[CrossRef](#)]
31. Ryou, H.; Sirinukunwattana, K.; Aberdeen, A.; Grindstaff, G.; Stolz, B.J.; Byrne, H.; Harrington, H.A.; Sousos, N.; Godfrey, A.L.; Harrison, C.N.; et al. Continuous Indexing of Fibrosis (CIF): Improving the assessment and classification of MPN patients. *Leukemia* **2022**, *37*, 348–358. [[CrossRef](#)] [[PubMed](#)]
32. Angelopoulos, N.; Chatzipli, A.; Nangalia, J.; Maura, F.; Campbell, P.J. Bayesian networks elucidate complex genomic landscapes in cancer. *Commun. Biol.* **2022**, *5*, 306. [[CrossRef](#)] [[PubMed](#)]
33. Li, W.; Zhao, Y.; Wang, D.; Ding, Z.; Li, C.; Wang, B.; Xue, X.; Ma, J.; Deng, Y.; Liu, Q.; et al. Transcriptome research identifies four hub genes related to primary myelofibrosis: A holistic research by weighted gene co-expression network analysis. *Aging* **2021**, *13*, 23284–23307. [[CrossRef](#)] [[PubMed](#)]
34. Kimura, K.; Tabe, Y.; Ai, T.; Takehara, I.; Fukuda, H.; Takahashi, H.; Naito, T.; Komatsu, N.; Uchihashi, K.; Ohsaka, A. A novel automated image analysis system using deep convolutional neural networks can assist to differentiate MDS and AA. *Sci. Rep.* **2019**, *9*, 13385. [[CrossRef](#)] [[PubMed](#)]
35. Briggs, C.; Longair, I.; Kumar, P.; Singh, D.; Machin, S.J. Performance evaluation of the Sysmex haematology XN modular system. *J. Clin. Pathol.* **2012**, *65*, 1024–1030. [[CrossRef](#)]
36. Salah, H.T.; Muhsen, I.N.; Salama, M.E.; Owaidah, T.; Hashmi, S.K. Machine learning applications in the diagnosis of leukemia: Current trends and future directions. *Int. J. Lab. Hematol.* **2019**, *41*, 717–725. [[CrossRef](#)]
37. Wu, Y.Y.; Huang, T.C.; Ye, R.H.; Fang, W.H.; Lai, S.W.; Chang, P.Y.; Liu, W.N.; Kuo, T.Y.; Lee, C.H.; Tsai, W.C.; et al. A Hematologist-Level Deep Learning Algorithm (BMSNNet) for Assessing the Morphologies of Single Nuclear Balls in Bone Marrow Smears: Algorithm Development. *JMIR Med. Inform.* **2020**, *8*, e15963. [[CrossRef](#)]
38. Wang, Q.; Bi, S.; Sun, M.; Wang, Y.; Wang, D.; Yang, S. Deep learning approach to peripheral leukocyte recognition. *PLoS ONE* **2019**, *14*, e0218808. [[CrossRef](#)]
39. Yassin, M.A.; Nehmeh, S.A.; Nashwan, A.J.; Kohla, S.A.; Mohamed, S.F.; Ismail, O.M.; Al Sabbagh, A.; Soliman, D.S.; Szabados, L.; Fayad, H. Assessing Bone Marrow Activity with [18F]FLT PET in Patients with Essential Thrombocythemia and Prefibrotic Myelofibrosis: A Proof of Concept. *Technol. Cancer Res. Treat.* **2022**, *21*, 15330338221086396. [[CrossRef](#)]
40. Stuart, B.J.; Viera, A.J. Polycythemia vera. *Am. Fam. Physician* **2004**, *69*, 2139–2144.
41. Djulbegovic, B.; Hadley, T.; Joseph, G. A new algorithm for the diagnosis of polycythemia. *Am. Fam. Physician* **1991**, *44*, 113–120.
42. Djulbegovic, B.; Hozo, I. Computer program for the diagnosis and treatment of polycythemia rubra vera. *MD Comput.* **1999**, *16*, 83–89.
43. Guncar, G.; Kukar, M.; Notar, M.; Brvar, M.; Cernelc, P.; Notar, M.; Notar, M. An application of machine learning to haematological diagnosis. *Sci. Rep.* **2018**, *8*, 411. [[CrossRef](#)]
44. Belcic, T.; Cernelc, P.; Sever, M. Artificial intelligence aiding in diagnosis of JAK2 V617F negative patients with who defined essential thrombocythemia. *HemaSphere* **2019**, *3*, 998. [[CrossRef](#)]
45. Skov, V.; Burton, M.; Thomassen, M.; Stauffer Larsen, T.; Kjær, L.; Hasselbalch Riley, C.; Jensen, M.K.; Bjerrum, O.W.; Kruse, T.A.; Hasselbalch, H.C. Supervised multi-classifier separation of the Ph-negative myeloproliferative neoplasms. *Haematologica* **2015**, *100*, 522–523.
46. Al-Dewik, N.; Ben-Omran, T.; Zayed, H.; Trujillano, D.; Kishore, S.; Rolfs, A.; Yassin, M.A. Clinical Exome Sequencing unravels new disease-causing mutations in the myeloproliferative neoplasms: A pilot study in patients from the state of Qatar. *Gene* **2019**, *689*, 34–42. [[CrossRef](#)] [[PubMed](#)]
47. Al-Dewik, N.I.; Cassinat, B.; Kiladjian, J.-J.; Knuth, A.; Yassin, M.A. Targeted Exome Sequencing Identifies Novel Mutations in Familial Myeloproliferative Neoplasms Patients in the State of Qatar. *Blood* **2014**, *124*, 5570. [[CrossRef](#)]
48. 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*, 4223.
49. Jabalameli, M.R.; Ennis, S.; Collins, A.; Cross, N.C.P.; Tapper, W.J. Ensemble Random Forest Classifier for predicting myeloproliferative neoplasms subtype using patient’s genomic profile. *Eur. J. Hum. Genet.* **2019**, *26*, 690. [[CrossRef](#)]
50. Spivak, J.L. Myeloproliferative Neoplasms. *N. Engl. J. Med.* **2017**, *377*, 895–896. [[CrossRef](#)] [[PubMed](#)]

51. Spivak, J.L.; Considine, M.; Williams, D.M.; Talbot, C.C., Jr.; Rogers, O.; Moliterno, A.R.; Jie, C.; Ochs, M.F. Two clinical phenotypes in polycythemia vera. *N. Engl. J. Med.* **2014**, *371*, 808–817. [[CrossRef](#)] [[PubMed](#)]
52. Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Society Ser. B* **1996**, *58*, 267–288. [[CrossRef](#)]
53. Moons, K.G.; Kengne, A.P.; Grobbee, D.E.; Royston, P.; Vergouwe, Y.; Altman, D.G.; Woodward, M. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart* **2012**, *98*, 691–698. [[CrossRef](#)]
54. Guo, B.B.; Linden, M.D.; Fuller, K.A.; Phillips, M.; Mirzai, B.; Wilson, L.; Chuah, H.; Liang, J.; Howman, R.; Grove, C.S.; et al. Platelets in myeloproliferative neoplasms have a distinct transcript signature in the presence of marrow fibrosis. *Br. J. Haematol.* **2020**, *188*, 272–282. [[CrossRef](#)] [[PubMed](#)]
55. Rondina, M.T.; Voora, D.; Simon, L.M.; Schwertz, H.; Harper, J.F.; Lee, O.; Bhatlekar, S.C.; Li, Q.; Eustes, A.S.; Montenont, E.; et al. Longitudinal RNA-Seq Analysis of the Repeatability of Gene Expression and Splicing in Human Platelets Identifies a Platelet SELP Splice QTL. *Circ. Res.* **2020**, *126*, 501–516. [[CrossRef](#)]
56. Campbell, P.J.; MacLean, C.; Beer, P.A.; Buck, G.; Wheatley, K.; Kiladjian, J.J.; Forsyth, C.; Harrison, C.N.; Green, A.R. Correlation of blood counts with vascular complications in essential thrombocythemia: Analysis of the prospective PT1 cohort. *Blood* **2012**, *120*, 1409–1411. [[CrossRef](#)]
57. Abu-Zeinah, G.; Krichevsky, S.; Silver, R.T.; Taylor, E.; Tremblay, D.; Srisuwananukorn, A.; Mascarenhas, J.; Scandura, J. A Novel Machine Learning-Derived Dynamic Scoring System Predicts Risk of Thrombosis in Polycythemia Vera (PV) Patients. *Blood* **2021**, *138*, 3619. [[CrossRef](#)]
58. Srisuwananukorn, A.; Percha, B.; Abu-Zeinah, G.; Scandura, J.; Krichevsky, S.; Taylor, E.; Silver, R.T.; Thibaud, S.; Ronner, L.; Podoltsev, N.; et al. Novel Machine Learning Algorithm Predicts Disease Progression in Polycythemia Vera (PV) with Readily-Available Baseline Characteristics. *Blood* **2021**, *138*, 2583. [[CrossRef](#)]
59. Verstovsek, S.; De Stefano, V.; Heidel, F.H.; Zuurman, M.; Zaiac, M.; Bigan, E.; Ruhl, M.; Meier, C.; Beffy, M.; Kiladjian, J.J. Machine-learning to predict hydroxyurea (HU) failure and incidence of thromboembolic events (TES) with hu vs ruxolitinib switch therapy in polycythemia vera patients. *HemaSphere* **2020**, *4*, 517–518. [[CrossRef](#)]
60. Mora, B.; Kuykendall, A.; Guglielmelli, P.; Rumi, E.; Maffioli, M.; Komrokji, R.; Barraco, D.; Rambaldi, A.; Caramella, M.; Kiladjian, J.J.; et al. MYSEC-prognostic model and previous thrombotic events predict the risk of thrombosis in post polycythemia vera and post essential thrombocythemia myelofibrosis: A study of the MYSEC group. *Leukemia* **2022**, *36*, 2453–2460. [[CrossRef](#)]
61. Bejan, C.A.; Sochacki, A.; Zhao, S.; Xu, Y.; Savona, M. Identification of myelofibrosis from electronic health records with novel algorithms and JAKextractor. *Cancer Res.* **2018**, *78*, 5303. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.