Proceedings of the International Conference on Industrial Engineering and Operations Management

Publisher: IEOM Society International, USA DOI: 10.46254/AN16.20260038
Published: June 23, 2026

Differentiating PMF–ET/PV using Gene Expression Data: A Pipeline Approach

Yesim Kayabas Alkan
Research Assistant
Industrial Engineering Department
Tokat Gaziosmanpasa University, Tokat, Türkiye
yesim.alkan@gop.edu.tr

Mehmet Kabak
Professor
Industrial Engineering Department
Gazi University, Ankara, Türkiye
mkabak@gazi.edu.tr

Abstract

Within the scope of Philadelphia chromosome-negative myeloproliferative neoplasms, the effective differentiation of Primary Myelofibrosis from Essential Thrombocythaemia and Polycythaemia Vera is of vital importance for prognosis estimation and treatment approach planning. This study aimed to distinguish Primary Myelofibrosis samples from the Essential Thrombocythaemia - Polycythaemia Vera group using two gene expression datasets obtained from the Gene Expression Omnibus repository. A dataset of 69 samples for training and data from 33 untreated samples for independent external validation were converted to the gene level via GPL570 annotation and matched in a common feature space. Data preprocessing, feature selection, and modelling procedures were designed within a pipeline to minimize the risk of data leakage. Hyperparameter optimization was performed using the GridSearchCV technique with iterative stratified cross-validation. Similar Average Precision and Balanced Accuracy values were obtained during the validation phase performed on the training data. In the independent external validation phase, the Linear Support Vector Machine method stood out with higher Average Precision and Balanced Accuracy values, while the Elastic Net Logistic Regression method provided higher values in terms of the Receiver Operating Characteristic - Area Under Curve metric. The findings reveal that a pipeline design using open-source data from the Gene Expression Omnibus archive can be used to distinguish Primary Myelofibrosis samples.

Keywords
Primary Myelofibrosis, Gene Expression, Linear SVM, Elastic Net Logistic Regression, and PR-AUC.

1. Introduction
Philadelphia chromosome (BCR::ABL1)-negative Myeloproliferative Neoplasms (MPN), including Essential Thrombocythaemia (ET) and Polycythaemia Vera (PV), and Myelofibrosis (PMF), are chronic conditions with similar biological characteristics; however, they are serious diseases with significant morbidity and disease burden (Pescia et al., 2024; Thiele et al., 2023). When examining studies and trends specific to the population from the US Surveillance, Epidemiology, and End Results (SEER) data, the annual incidence rates per 100,000 people are 1.8 for ET, 1.7 for PV, and 0.5 for PMF. Considering the 5-year average mortality rates, they are 19.2% for ET, 19% for PV, and 51% for PMF. When examining the 5-year relative survival rates, they are 88.7% for ET, 88.3% for PV, and 44.9% for PMF (Smith et al., 2021; Singhal et al., 2025; Verstovsek et al., 2022). The similarity of these biologically related

diseases in laboratory and clinical data makes it challenging to differentiate between ET, PV, and PMF and to accurately diagnose them. In particular, the inability to distinguish between PMF and ET diseases increases the risk of misdiagnosis and incorrect treatment (Griesshammer et al., 2025; Thiele et al., 2025).

BCR::ABL1-negative MPNs, encompassing ET, PV, and PMF, are characterized by driver mutations in Thrombopoietin Receptor (MPL), Janus kinase 2 (JAK2), and Calreticulin (CALR), which continuously and uncontrollably affect JAK signaling (Ling et al., 2025; Loghavi et al., 2024). Besides these mutations, distinguishing between ET, PV, and PMF for diagnosis is challenging due to similarities among the subtype mutations (Gianelli et al., 2023). Primary Myelofibrosis (PMF) is a type of MPN caused by clonal abnormalities in stem cells that produce blood cells. The defining feature of PMF is the formation of fibrosis in bone marrow cells and, consequently, the production of blood in different organs of the body (extramedullary haematopoiesis).

It is most commonly triggered by MPL, JAK2, or CALR mutations and is frequently seen in conjunction with the disease. Splenomegaly and hepatomegaly, cachexia, anaemia, and fibrosis formation are common symptoms that occur when the bone marrow fails to produce sufficient blood cells. Bone marrow biopsies of patients show megakaryocytic hyperplasia and fibrosis together. As the disease progresses, the risk of leukaemic transformation increases (Shao et al., 2025; Tefferi, 2023). Because of the difficulty of integrating molecular, clinical, and histological findings in the diagnosis of MPN diseases and the complexity and variability of bone marrow morphology, misdiagnosis is common (Schmidt et al., 2023). Therefore, for the detection and accurate diagnosis of these diseases, laboratory results must be considered in conjunction with bone marrow histopathology (McLornan et al. 2024). Accurate diagnosis is also critical for estimating patient survival and calculating the likelihood of progression to leukaemia using prognostic scoring methods (Duminuco et al., 2023).

Due to the difficulties encountered in diagnosis, data-driven strategies that can conclude pathological and clinical investigations are of significant importance (Srisuwananukorn et al., 2024; Yu et al., 2025). Large-scale molecular data and gene expression profiles can provide supportive information in distinguishing diseases from one another (Bassan et al., 2024). The data can be obtained through publicly available platforms (Clough and Barrett, 2016). In this study, an open-source gene expression dataset was utilized to present a classification approach based on distinguishing ET–PV–PMF diseases from one another, and performance evaluation was conducted.

1.1 Objectives

The aim of this study is to develop a machine learning approach that uses gene expression data obtained through the publicly available Gene Expression Omnibus (GEO) platform to distinguish PMF, a rare and prognostically unfavorable form of MPN that is BCR::ABL1 negative, from ET and PV diseases that are biologically similar to this disease. (Tefferi and Vannucchi, 2024). In this context, microarray data were downsampled to the gene level, and data cleaning and standardization procedures were performed. Linear Support Vector Machines (SVMs) and Elastic Net Logistic Regression methods were applied. The method selection was made based on the PR-AUC (Average Precision) metric, which is an explanatory feature in imbalanced datasets (Saito and Rehmsmeier, 2015). In the final stage, the preferred method was applied to a different dataset, and the validity of the model was evaluated. This study aims to contribute to the literature by providing an integrated analytical approach based on GEO datasets and their validation.

2. Literature Review

Within BCR::ABL1-negative MPNs, genetic factors are utilized in addition to clinical parameters in the literature for disease differentiation and clinical course prediction (Arber et al., 2022). In addition to the driver mutations commonly seen in PMF, DNMT3A, TET2, and ASXL1 mutations are also frequently observed in MPN patients. Mutations affecting epigenetic and splicing regulators (e.g., TP53, EZH2, STAG2, KRAS, IDH1–2, CBL, NRAS, ZRSR2, SRSF2, U2AF1, and SF3B1) are less common. These additional mutations are more commonly encountered in advanced stages of PMF than in ET and PV. Mutations such as SRSF2, EZH2, IDH1–2, ASXL1, and U2AF1 are known to increase the risk of poor survival. These mutations are identified by applying Next-Generation Sequencing (NGS) methods to determine the risk status and prognosis of the disease (Khoury et al., 2022; Ling et al., 2025).

In PMF, patients are assessed and grouped into risk categories according to various prognostic scoring methods to predict clinical course and survival (Tefferi and Vannucchi, 2024). The International Prognostic Scoring System (IPSS) was introduced into the literature in 2009. The criteria evaluated in this calculation include haemoglobin and

white blood cell counts, age, blast cell percentage, unexplained weight loss, fever, and night sweats (Cervantes et al., 2009). DIPSS (Dynamic International Prognostic Scoring System) was discovered in 2010 and considers the same criteria as the IPSS calculation, but the scoring system is different (Passamonti et al., 2010). DIPSS-plus was introduced in 2011 and, in addition to the criteria in the DIPSS calculation, also examines platelet count, red blood cell transfusion requirement, and cytogenetic abnormalities (Gangat et al., 2011). MIPSS70 (Mutation-Enhanced International Prognostic Scoring System for patients aged ≤ 70 years). Introduced in 2018 for patients aged 70 years and younger, it considers haemoglobin, leukocyte and platelet counts, blast cell percentage, the presence of high-risk mutations, fibrosis level, and the presence or absence of CALR Type 1 mutation in the blood (Guglielmelli et al., 2018). GIPSS (the Genetically-Inspired IPSS, which depends exclusively on mutations and karyotype). This method, introduced in 2018, evaluates genetic characteristics such as cytogenetic abnormalities and the presence or absence of ASXL1, CALR Type 1, U2AF1Q157, and SRSF2 mutations in the blood, rather than clinical parameters (Tefferi et al., 2018a). MIPSS70+ Version 2.0, introduced into the literature in 2018, incorporates cytogenetic data in addition to the MIPSS70 calculation (Tefferi et al., 2018b). In addition to these methods, the machine learning-based Artificial Intelligence Prognostic Scoring System for Myelofibrosis (AIPSS-MF)method proposed in 2023 predicts survival and the transformation of the disease into leukaemia (Duminuco et al., 2023).

Although prognostic scoring methods are effective, distinguishing between these three diseases requires extensive molecular data due to their similar biological structures (Grinfeld et al., 2018). The literature indicates that gene expression information supports the differentiation between ET, PV, and PMF by providing a more detailed picture of the molecular pathogenesis of the diseases (Schischlik, 2022; Skov et al., 2012). As gene expression information is shared in public archives such as GEO, it enables comparison and external validation in the context of disease differentiation (Clough and Barrett, 2016; Rung and Brazma, 2013). In recent academic studies, prognostic and diagnostic marker models based on machine learning have been developed using gene expression data, and the model validation process has been carried out using independent data sets (Li et al., 2025; Medeiros et al., 2026).

In the literature, unsupervised learning-based studies utilizing gene expression data to distinguish between ET-PV-PMF diseases are limited. Using the GSE26049 data set, Principal Component Analysis (PCA) and DNA-Chip Analyzer (dChip) methods were utilized to analyze the elements supporting the biological transformation phase between ET-PV-PMF at the molecular level and to reveal patterns between diseases (Skov et al., 2012). Using the GSE26049 data set, hierarchical clustering was utilized to identify patient groups at risk of transitioning from ET-PV diseases to PMF disease (Hasselbalch et al., 2014).

Studies utilizing deep learning and machine learning without using gene expression data for the diseases in question also exist. In order to differentiate between ET-PV-PMF diseases, SVM and 3-fold CV (Cross-Validation) methods were applied to patient groups of 243 and 183 individuals, taking into account driver mutations as well as 15 additional mutations, cytogenetic and blood results (Meggendorfer et al., 2017). Bayesian clustering, Cox models, and CV methods were applied to a sample of 2035 patients to develop the ET-PV-PMF taxonomy, predict prognosis, and perform validation (Grinfeld et al., 2018). The differentiation of ET-PV-PMF diseases was performed using the Random Forest method by converting the microscopic image and physical structure of bone marrow cells into numerical features (Sirinukunwattana et al., 2020). Convolutional Neural Network (CNN) and XGBoost methods applied to differentiate ET-PV-PMF diseases using blood results, peripheral smear, and cytology results from a sample of 234 individuals (Kimura et al., 2021). In this context, performance metric selection plays a critical role in measuring the success of models and methods used in the medical classification studies evaluated (Salmi et al., 2024).

The correct selection of performance metrics in medical classification problems plays a significant role in interpreting results and demonstrating discriminatory power. In this context, the widely used PR (Precision–Recall) and ROC-AUC (Receiver Operating Characteristic - Area Under Curve) curve models reflect model performance from different perspectives. PR Precision – sensitivity, while ROC-AUC presents the false positive rate – sensitivity interaction (Collins et al., 2024; Martin et al., 2025; Van Calster et al., 2025). When there is an imbalanced distribution in the classes of the dataset, it is recommended to evaluate PR-AUC (AUPRC) precision-recall-based metrics together with the ROC-AUC metric, as they complement each other (Salmi et al., 2024; Van Calster et al., 2025). Furthermore, it is recommended to perform both internal validation and external validation with a different dataset to prove the model's performance and validity (Collins et al., 2024; Riley et al., 2024). Open-access gene expression archives such as GEO provide advantages for model development, comparison, and validation (Clough et al., 2024).

In this study, two datasets obtained from the GEO archive were converted to the gene level using the GPL570 platform annotation to distinguish PMF from ET-PV diseases, with one selected for training and the other for external validation. Following data cleaning and standardization, Linear SVM and Elastic Net logistic regression methods were applied. Validation results were evaluated using Balanced Accuracy (BA), Average Precision (AP), ROC-AUC, and Confusion Matrix.

3. Methods

The working methodology utilizes one dataset for model construction and training stages, while the second dataset is used to analyze model performance through independent external validation. To enable genetic data tables to be evaluated in the same space, probe values were aligned by converting them to the gene level using the GPL570 platform annotation. To represent multiple probe values corresponding to a single gene, values were singularized using the median (NCBI GEO, 2026a).

All steps within the analysis process have been designed within a pipeline structure to minimize the risk of data leakage. As the model input data must contain numerical values, the input table values have been converted to numerical values; cells that do not contain numerical values or are found to be erroneous due to typing or reading errors have been defined as Not a Number (NaN). Missing data points in the input table were filled using the median approach with the SimpleImputer tool to reduce deviations caused by outliers and extreme values. Subsequently, z-score standardization was applied using StandardScaler to reduce the effect of variability that could arise from the measurement values of the features. Features with zero variance were removed from the dataset using the VarianceThreshold method because they would not contribute to the modelling stage. After the preprocessing and removal of features with zero variance, the SelectKBest method based on the ANOVA F-test was added to the pipeline for the analysis-ready data. Different numbers of genes (k = 200, 500, and 1000) were tested within the GridSearchCV framework to determine the feature set with the highest performance (Pedregosa et al., 2011).

3.1 Linear SVM Method

The method, introduced in 1992, was developed to select the correct hyperplane for separating classes in classification problems and to maximize the margin between classes at this stage. The method took its modern form in 1995 under the name 'Support Vector Networks'. Data that cannot be separated by the correct hyperplane is accepted with a soft-margin error. One of the advantages of the method is that it provides the possibility of reaching the global optimum without falling into local minima (Boser et al., 1992; Cortes and Vapnik, 1995). Variants of the method are used in studies using gene expression data to separate classes in high dimensions (Brown et al., 2000). The linear SVM method is fast and computationally straightforward for high-dimensional data sets and is based on the LIBLINEAR library (Fan et al., 2008). In this study, margin maximization and penalty for margin violations were targeted, and the soft-margin SVM optimization model given in Equation 1 was used.

$$\begin{aligned} &\min_{w,b,\xi} \frac{1}{2}\,||w||^2 + C\sum_{i=1}^{n}\xi_i \\ &\text{s.t. } y^{(i)}(w^T x^{(i)} + b) \geq 1 - \xi_i, \qquad i = 1,\ldots,n \\ &\xi_i \geq 0, \qquad i = 1,\ldots,n \end{aligned} \tag{1}$$

In Equation 1, the weight w belongs to the hyperplane separating the classes, and b represents the deviation, determining the position and direction of the separating plane. The input variables $y^{(i)}$ and $x^{(i)}$ in the model represent, respectively, the label of the class to which the $i.$ observation belongs and its feature vector. $y^{(i)}$ is defined as being in the set $\{-1, +1\}$. The term $\frac{1}{2}\,||w||^2$ in the objective function aims to widen the margin of the hyperplane, while the term $C\sum_{i=1}^{n}\xi_i$ aims to penalize the total margin violations. The first constraint aims to ensure that the functional margin is at least $1 - \xi_i$. The second constraint ξ_i parameter (violation level) indicates the non-negativity condition (Cortes and Vapnik, 1995; Ng and Ma, 2023).

3.2 Elastic Net Logistic Regression Method

The method introduced in 2005 is a regularization method that integrates L1 (Lasso) and L2 (Ridge) penalties (Zou and Hastie 2005). When the number of variables exceeds the number of samples, and there is correlation between variables, L1 is used to set some coefficients to zero to improve discriminability, while L2 is used to support a stable structure by only reducing coefficients rather than setting them to zero. The method's grouping effect feature provides the advantage of selecting groups of variables with high correlation. This mechanism is a preferred approach in

Generalized Linear Models (GLM) methods such as logistic regression (Tay et al., 2023). The GLMnet framework, published in 2010, enabled system optimization by integrating elastic net penalties (L1 – L2) into logistic regression (Friedman et al., 2010). Equation 2 presents the mathematical formulation of the method (Torang et al., 2019).

$$\min_{\beta,\beta_0} \frac{1}{n} \sum_{i=1}^{n} log(1 + exp(-y_i(\beta_0 + x_i^T \beta)) \; + \; \lambda((1-\alpha)\frac{||\beta||_2^2}{2} \; + \; \alpha||\beta||_1 \qquad (2)$$

The β coefficient vector given in Equation 2, the β_0 constant term, n total sample count, y_i and x_i respectively denote the label of the class to which the i. observation belongs to and the feature vector, represented by $\frac{1}{n} \sum_{i=1}^{n} log(1 + exp(-y_i(\beta_0 + x_i^T \beta))$ represents the mean log loss function associated with the logistic regression method, and the expression $\beta_0 + x_i^T \beta$ calculates the fit score of the samples. $\lambda((1-\alpha)\frac{||\beta||_2^2}{2} + \alpha||\beta||_1$ is the regularization part of the formulation. The $||\beta||_2^2$ term reduces the coefficients to minimise overfitting, while the $||\beta||_1$ term sets some coefficients to zero. λ denotes the penalty coefficient, while α denotes the mixing parameters between the L1 and L2 methods. The α value of 1 represents L1, while a value of 0 represents L2. y_i is defined as being in the set $\{-1, +1\}$, $\lambda \geq 0$, and α is in the interval [0,1] (Bühlmann and Van De Geer, 2011; Münch et al., 2021).

3.3 Hyperparameter Optimization and Internal Validation

The hyperparameters of the established models were identified by scanning different parameters using the GridSearchCV method. The Repeated Stratified K-Fold method was applied during the internal validation process of the models to ensure consistent class balance; 3-fold cross-validation was repeated 10 times. Furthermore, a fixed random state value was defined to minimize deviations in the results that could arise from random variability. To ensure that the data preprocessing and modelling stages of the analysis process were repeated at each fold of the cross-validation, all analysis steps were configured through a pipeline, minimizing the risk of data leakage in both the training and testing phases. To analyze model performance, multi-faceted performance metrics such as Average Precision (AP/PR-AUC), BA, and ROC-AUC were evaluated together (Efthimiou et al., 2024; Yates et al., 2023).

Model performance has been measured using various metrics, as these metrics complement each other by calculating ranking performance without requiring class-based metrics or classification thresholds. Specifically, in cases of data imbalance, the PR curve and AP criteria were selected as the final model based on the AP criterion because they more clearly represent the ability to distinguish the positive class (Davis and Goadrich, 2006; Saito and Rehmsmeier, 2015). The BA metric provides a more reliable result in distributions with class imbalance by equalizing the effect between classes (Brodersen et al., 2010). Therefore, the model that yielded the best results was completed using the AP metric as a basis through refit.

3.4 External Validation

During the internal validation process, all stages within the pipeline structure were retrained on the training dataset according to the best hyperparameters determined using the GridSearchCV method based on the AP criterion, and tested on the independent external dataset without any changes. During the external validation process, no hyperparameter changes or retraining were performed on the parameters learned from the training data. The independent external data was applied to the external data using the parameters learned from the training data, and prediction results were calculated. This prevented the possibility of data leakage. The model's performance on the external data set was presented using AP/PR-AUC, BA, ROC-AUC, and the confusion matrix (Collins et al., 2024; Efthimiou et al., 2024).

4. Data Collection

The gene expression datasets used in this study were obtained from the GEO archive. The untreated subsets GSE26049 were used during the training phase and GSE57793 during the independent external validation phase (NCBI GEO, 2026b; NCBI GEO, 2026c). In the training dataset, ET-PV-PMF diagnosed samples were included, while in the external validation dataset, pre-treatment individuals were included in the study. The study was designed within a binary classification framework, with class labels coded as PMF 1 and ET – PV 0. Genetic data tables were obtained via GEO and imported as compressed .csv.gz format files. The datasets used in the study are publicly available on GEO and do not contain patient identification information. Therefore, they are consistent with the institution's perspective on the reuse of targeted data.

5. Results and Discussion

The findings obtained are examined in the initial stage through numerical summaries; subsequently, the performance demonstrated by the methods is comprehensively addressed through performance curves and graphs. Additionally, potential development suggestions and an examination of the validation approach are shared.

5.1 Numerical Results

This section of the study presents the class distributions and common gene counts of the training and external validation datasets. Additionally, the internal and external validation performance values of the linear SVM and Elastic Net logistic regression methods considered are presented in Table 1.

Table 1. Characteristics of the data sets and model performance

Category	***Item***	***Result***
Datasets	Train (GSE26049)	ET = 19, PV = 41, PMF =9, n = 69
	External (GSE57793, untreated)	ET = 8, PV = 21, PMF =4, n = 33
	Number of shared genes	22,879 Features
Internal Validation	Elastic Net Logistic Regression (C=0.1, α =0.1, k=200)	AP: 0.830±0.154 ROC-AUC: 0.913±0.091 BA: 0.821±0.116
	Linear SVM (C=0.01, k=1000)	AP: 0.825±0.157 ROC-AUC: 0.903±0.102 BA: 0.823±0.120
External Validation	Elastic Net Logistic Regression	AP: 0.604 (baseline=0.121) ROC-AUC: 0.862 BA: 0.733 CM [[TN=28, FP=1], [FN=2, TP=2]] P=0.667, R=0.500
	Linear SVM	AP: 0.650 (baseline=0.121) ROC-AUC: 0.828 BA: 0.806 CM [[TN=25, FP=4], [FN=1, TP=3]] P=0.429, R=0.750

Among the hyperparameters that play a critical role during optimization, C represents the parameter that determines the regularization effect of the model. In the Elastic Net Logistic Regression model, α represents the mixing coefficient of the L1 and L2 techniques, and k, present in both models, represents the number of features to be included in the models during the feature (gene) selection process. GridSearchCV method results showed that C = 0.1, α = 0.1, and k = 200 for Elastic Net logistic regression, and C = 0.01 and k = 1000 for linear SVM are the hyperparameters that optimize model performance. In the independent external validation process, the baseline used as a reference point represents the proportion of positive class (PMF) in the relevant dataset (4/33 ≈ 0.121). In analyzing the region under the Sensitivity (PR) curve (AUPRC/AP), the reference baseline limit is directly equal to the percentage of positive classes in the dataset, so model success was evaluated based on this value (Ren et al., 2024). Confusion matrix outputs were presented in the format [[TN, FP], [FN, TP]]; in this representation, the resulting TN, FP, FN, and TP values represent the accuracy of negatives, false alarms, false rejections, and accurate diagnoses, respectively.

While no significant difference was observed in the internal validation phase of the two methods' performance, method-based differences were observed in various metrics in the external validation phase. When compared, the highest ROC-AUC value was observed in the Elastic Net Logistic Regression method with a value of 0.862, while the highest AP value was observed in the Linear SVM method with a value of 0.650, and the highest BA value with a value of 0.806. Since the baseline value in the independent external dataset was 0.121, the AP metric results being above 0.121 indicate that the methods used have a higher power to discriminate PMF disease compared to a random method. In this part of the analysis, the confusion matrix results were examined in detail to analyze model

performance. In the Elastic Net Logistic Regression method, out of a total of 33 samples, 29 were diagnosed with ET-PV, and 4 with PMF. Twenty-eight samples were correctly identified as the ET-PV group, while one sample, although within the ET-PV group, was incorrectly classified as a PMF patient. Of the four PMF samples, two were classified as PMF and the other two as the ET-PV group. In the linear SVM method, out of a total of 33 samples, 25 were correctly identified as the ET-PV group, while four samples, although within the ET-PV group, were incorrectly classified as PMF patients. Of the four PMF samples, three were classified as PMF and the remaining one as the ET-PV group.

When the PMF class-based evaluation was performed, it was observed that the precision and recall values of the methods showed different performance. According to the Elastic Net logistic regression method results, the precision was obtained as 0.667 and the recall value as 0.500. This result shows that the method predicted fewer false positives in PMF samples, but failed to detect some PMF samples. According to the Linear SVM method results, the precision was obtained as 0.429 and the recall value as 0.750. This result shows that the method detected PMF samples at a higher rate compared to the Elastic Net logistic regression method, but predicted some ET-PV group samples as PMF. Since the main goal of this study is to distinguish PMF samples from ET-PV group samples, the Linear SVM method, which has a higher recall value that minimizes the risk of missing samples, is more successful in detecting PMF samples.

5.2 Graphical Results

In this part of the study, the estimation results obtained from independent external validation data are presented graphically. To analyze the data, which has an unbalanced distribution with a low positive class share, the AP metric and the PR curve were evaluated together with the ROC curve. In Figure 1, it is stated that the baseline value used for comparison based on the PR curve represents the positive class share (PMF frequency) in the independent external validation data.

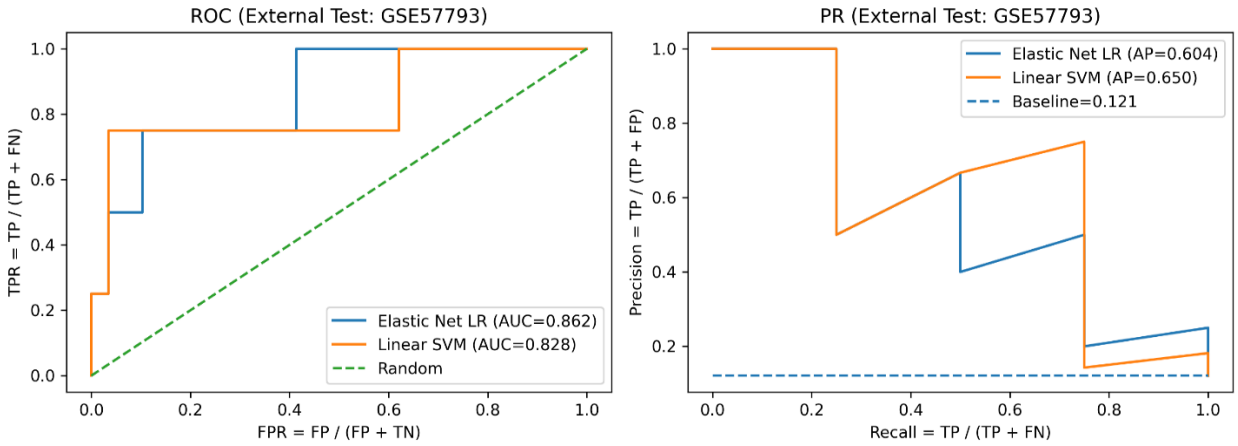

Figure 1. Comparison of ROC and PR (AP) curves in an independent external dataset

The ROC and PR (AP) curves of the Linear SVM and Elastic Net Logistic Regression methods in the independent external validation dataset are presented. The classification success of the comparison performed on the curves at the sample level is detailed through the confusion matrix shown in Figure 2.

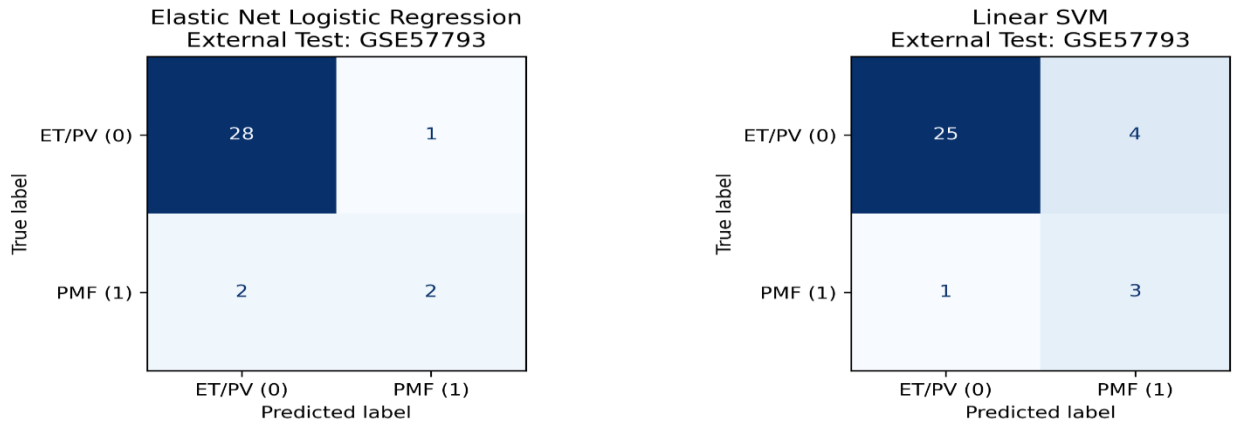

Figure 2. Classification results based on the confusion matrix in an independent external dataset

The Confusion Matrix in Figure 2 provides sample-based results of the graphical outputs presented in Figure 1, revealing which categories are predominantly error-prone. Because the independent external validation dataset does not have a homogeneous distribution, sample-level misclassifications can directly influence the evaluation of the findings.

5.3 Proposed Improvements

To support the conclusions of this study and increase its comprehensiveness, the success of the models can be re-evaluated with independent external validation datasets obtained from alternative sources and platforms; prediction and PMF capture results can be analyzed across different datasets. Due to the heterogeneous nature of the data, sensitivity tests to decision threshold values can be performed to evaluate potential impacts in the context of various scenarios. Performance metrics can be presented more clearly with validation of prediction probabilities and Bootstrap-based confidence interval analyses. In addition, comparative studies can be conducted with different feature selection methods and models for gene expression data with a large feature space; reporting steps can be detailed with current approaches.

5.4 Validation

In this study, the success rates in the independent external validation dataset were more cautious compared to the internal validation performed on the training dataset. This finding parallels the genetic-phenotypic overlaps between MPN subtypes and the non-homogeneous structure of the dataset (Elsayed et al., 2023; Jung et al., 2025). The fact that the AP value was higher than the baseline in the independent external validation phase, and that the ROC-AUC metrics were at an acceptable level in terms of success, indicates that the designed model was able to detect classifications related to the positive class and that the results were confirmed with independent test data (Elsayed et al., 2023; Saelmans et al., 2025). Considering the heterogeneous structure of the dataset, and since studies in the literature have addressed the fact that ROC-AUC and PR indicators provide information depending on the conditions, presenting these metrics together allows for the interpretation of the findings (McDermott et al., 2024). Additionally, since limited sample sizes and data from a single source can limit the validity of results in machine learning and artificial intelligence research, it is recommended to conduct further independent validation tests on larger datasets from different locations (Elsayed et al., 2023; Saelmans et al., 2025).

In approaches requiring parameter optimization, the test step can be designed using an internal cross-validation architecture on the training data to minimize the risk of inflating model performance. This allows the validity bias to be estimated more objectively by repeating the model at each external cross-validation stage (Varma and Simon, 2006). If the classification success of the two methods used in the study on the same independent external data set is to be compared using ROC-AUC values, the DeLong statistical test can examine the relevant ROC curves (DeLong et al., 1988). To analyze the performance success of models based on the Confusion Matrix in binary classification results emerging at a fixed threshold value and analyze the method error difference, the McNemar test can be applied (McNemar, 1947).

6. Conclusion

Distinguishing PMF from the ET-PV disease group plays a critical role in patient treatment and survival; this study utilized two distinct gene expression datasets obtained from the GEO open archive. GSE26049 was selected as the training dataset, and GSE57793–untreated as the independent external validation dataset. The data were aligned at the gene level using GPL 570 annotation, and multiple probe data corresponding to the same gene were converted to a single value using the median method. The data preparation, preprocessing, and modelling stages were structured within a pipeline to minimize the risk of data leakage in the training and testing steps. The hyperparameters of the Elastic Net Logistic Regression and Linear SVM approaches were optimized using the cross-validation method in the training part. Model performance was tested using the independent external dataset.

When examining the validation results on the training dataset, both methods showed similar levels of BA and AP values, and no superiority was detected between the two approaches. In studies conducted on an independent external validation dataset, it was found that the BA and AP metrics of the Linear SVM method were higher compared to the outputs of the Elastic Net Logistic Regression method. The AP values obtained at the end of both methodologies were observed to be above the baseline level, indicating the positive class proportion. However, based on the ROC-AUC metrics, the Elastic Net Logistic Regression method demonstrated more successful performance. When examined based on the PMF sample, the Elastic Net Logistic Regression method provided a more advantageous precision value (0.667), while the Linear SVM method yielded a higher recall value (0.750). As the aim is to reduce the risk of missing PMF samples, the Linear SVM method can be considered a more effective option.

The results obtained demonstrate that the pipeline strategy, constructed using GEO-sourced data, yields robust classification results during the independent external validation phase. Furthermore, considering that the data distribution is not homogeneous, it is recommended that repeated analyses with new external validation datasets from different sources and platforms be conducted to reinforce the findings with threshold sensitivity and Bootstrap-based confidence interval analyses. Additionally, testing the difference in success between different methods through statistical comparisons can strengthen the reliability of the results obtained.

References

Arber, D. A., Orazi, A., Hasserjian, R. P., Borowitz, M. J., Calvo, K. R., Kvasnicka, H. M., ... and Tefferi, A., International Consensus Classification of Myeloid Neoplasms and Acute Leukemias: integrating morphologic, clinical, and genomic data. *Blood, The Journal of the American Society of Hematology*, vol. 140, no. 11, pp. 1200-1228, 2022.

Bassan, V. L., de Freitas Martins Felicio, R., Ribeiro Malmegrim, K. C. and Attié de Castro, F., Myeloproliferative Neoplasms Transcriptome Reveals Pro-Inflammatory Signature and Enrichment in Peripheral Blood Monocyte-Related Genes. *Cancer Investigation*, vol. 42, no. 7, pp. 605-618, 2024.

Boser, B. E., Guyon, I. M. and Vapnik, V. N., A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pp. 144-152, Pittsburgh, USA, July 27-29, 1992.

Brodersen, K. H., Ong, C. S., Stephan, K. E. and Buhmann, J. M., The balanced accuracy and its posterior distribution. In *2010 20th international conference on pattern recognition*, pp. 3121-3124, IEEE, İstanbul, Turkey, 23-26 August, 2010.

Brown, M. P., Grundy, W. N., Lin, D., Cristianini, N., Sugnet, C. W., Furey, T. S., ... and Haussler, D., Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proceedings of the National Academy of Sciences*, vol. 97, no. 1, pp. 262-267, 2000.

Bühlmann, P. and Van De Geer, S., *Statistics for high-dimensional data: methods, theory and applications*. 1st Edition, Springer Science & Business Media, 2011.

Cervantes, F., Dupriez, B., Pereira, A., Passamonti, F., Reilly, J. T., Morra, E., ... and Tefferi, A., New prognostic scoring system for primary myelofibrosis based on a study of the International Working Group for Myelofibrosis Research and Treatment. *Blood, The Journal of the American Society of Hematology*, vol. 113, no. 13, pp. 2895-2901, 2009.

Clough, E. and Barrett, T., The gene expression omnibus database. In *Statistical Genomics: Methods and Protocols*, New York, NY: Springer New York, vol. 1418, pp. 93-110, 2016.

Clough, E., Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., ... and Soboleva, A., NCBI GEO: archive for gene expression and epigenomics data sets: 23-year update. *Nucleic acids research*, vol. 52, no. D1, pp.138-144, 2024.

Collins, G. S., Moons, K. G., Dhiman, P., Riley, R. D., Beam, A. L., Van Calster, B., ... and Logullo, P., TRIPOD+ AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *bmj*, no. 385, e078378, 2024.

Davis, J. and Goadrich, M., The relationship between Precision-Recall and ROC curves. In *Proceedings of the 23rd international conference on Machine learning*, pp. 233-240, Pittsburgh, USA, June 25-29, 2006.

DeLong, E. R., DeLong, D. M. and Clarke-Pearson, D. L., Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, vol. 44, no. 3, pp. 837-845, 1988.

Duminuco, A., Mosquera-Orgueira, A., Nardo, A., Di Raimondo, F. and Palumbo, G. A., AIPSS-MF machine learning prognostic score validation in a cohort of myelofibrosis patients treated with ruxolitinib. *Cancer Reports*, vol. 6, no. 10, e1881, 2023.

Efthimiou, O., Seo, M., Chalkou, K., Debray, T., Egger, M., and Salanti, G., Developing clinical prediction models: a step-by-step guide. *BMJ: British Medical Journal (Online)*, vol. 386, e078276, 2024.

Elsayed, B., Elshoeibi, A. M., Elhadary, M., Ferih, K., Elsabagh, A. A., Rahhal, A., ... and Yassin, M., Applications of artificial intelligence in philadelphia-negative myeloproliferative neoplasms. *Diagnostics*, vol. 13, no. 6, 1123, 2023.

Fan, R. E., Chang, K. W., Hsieh, C. J., Wang, X. R. and Lin, C. J., LIBLINEAR: A library for large linear classification. *the Journal of machine Learning research*, vol. 9, pp. 1871-1874, 2008.

Friedman, J. H., Hastie, T. and Tibshirani, R., Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, vol. 33, no. 1, pp. 1-22, 2010.

Gangat, N., Caramazza, D., Vaidya, R., George, G., Begna, K., Schwager, S., ... and Tefferi, A., DIPSS plus: a refined Dynamic International Prognostic Scoring System for primary myelofibrosis that incorporates prognostic information from karyotype, platelet count, and transfusion status. *Journal of clinical oncology*, vol. 29, no. 4, pp. 392-397, 2011.

Gianelli, U., Thiele, J., Orazi, A., Gangat, N., Vannucchi, A. M., Tefferi, A. and Kvasnicka, H. M., International Consensus Classification of myeloid and lymphoid neoplasms: myeloproliferative neoplasms. *Virchows Archiv*, vol. 482, no. 1, pp. 53-68, 2023.

Griesshammer, M., Al-Ali, H. K., Eckardt, J. N., Fiegl, M., Göthert, J., Jentsch-Ullrich, K., ... and Heidel, F. H., How I diagnose and treat patients in the pre-fibrotic phase of primary myelofibrosis (pre-PMF)-practical approaches of a German expert panel discussion in 2024. *Annals of Hematology*, vol. 104, no. 1, pp. 295-306, 2025.

Guglielmelli, P., Lasho, T. L., Rotunno, G., Mudireddy, M., Mannarelli, C., Nicolosi, M., ... and Tefferi, A., MIPSS70: mutation-enhanced international prognostic score system for transplantation-age patients with primary myelofibrosis. *Journal of Clinical Oncology*, vol. 36, no. 4, pp. 310-318, 2018.

Grinfeld, J., Nangalia, J., Baxter, E. J., Wedge, D. C., Angelopoulos, N., Cantrill, R., ... and Campbell, P. J. Classification and personalized prognosis in myeloproliferative neoplasms. *New England Journal of Medicine*, vol. 379, no. 15, pp. 1416-1430, 2018.

Hasselbalch, H. C., Skov, V., Stauffer Larsen, T., Thomassen, M., Hasselbalch Riley, C., Jensen, M. K., ... and Kruse, T. A., Transcriptional profiling of whole blood identifies a unique 5-gene signature for myelofibrosis and imminent myelofibrosis transformation. *PLoS One*, vol. 9, no. 1, e85567, 2014.

Jung, S. H., Lee, S. E., Yun, S., Min, D. E., Shin, Y., Chung, Y. J. and Lee, S. H., Different inflammatory, fibrotic, and immunological signatures between pre-fibrotic and overt primary myelofibrosis. *Haematologica*, vol. 110, no. 4, pp. 938 - 951, 2025.

Khoury, J. D., Solary, E., Abla, O., Akkari, Y., Alaggio, R., Apperley, J. F., ... and Hochhaus, A., The 5th edition of the World Health Organization classification of haematolymphoid tumours: myeloid and histiocytic/dendritic neoplasms. *Leukemia*, vol. 36, no. 7, pp. 1703-1719, 2022.

Kimura, K., Ai, T., Horiuchi, Y., Matsuzaki, A., Nishibe, K., Marutani, S., ... and Tabe, Y., Automated diagnostic support system with deep learning algorithms for distinction of Philadelphia chromosome-negative myeloproliferative neoplasms using peripheral blood specimen. *Scientific reports*, vol. 11, no. 1, 3367, 2021.

Li, R., Han, Q., Zeng, T., Wen, B., Zheng, Z., Bao, X. and Huang, J., Machine learning identifies inflammation-related diagnostic biomarkers for primary myelofibrosis with clinical validation. *Scientific Reports*, vol. 15, no. 1, 40986, 2025.

Ling, V. Y., Heidel, F. H. and Bywater, M. J., Pathogenesis and management of high molecular risk myeloproliferative neoplasms. *Haematologica*, vol. 110, no. 4, pp. 863 - 876, 2025.

Loghavi, S., Kanagal-Shamanna, R., Khoury, J. D., Medeiros, L. J., Naresh, K. N., Nejati, R. and Patnaik, M. M., Of the world health classification of tumors of the hematopoietic and lymphoid tissue: myeloid neoplasms. *Modern Pathology*, vol. 37, no. 2, 100397, 2024.

Martin, B., Bennett, T. D., DeWitt, P. E., Russell, S. and Sanchez-Pinto, L. N., Use of the Area Under the Precision-Recall Curve to Evaluate Prediction Models of Rare Critical Illness Events. *Pediatric Critical Care Medicine*, vol. 26, no. 6, e855-e859, 2025.

Medeiros, J. J., Zeng, A. G., Bansal, S., Kim, H., Yue, M. C. S., Woo, T., ... and Gupta, V., A Stem and Progenitor Cell-Derived Gene Expression Signature is Prognostic for Survival in Myelofibrosis. *Blood,* in press (Journal Pre-proof)*,* 2026.

Meggendorfer, M., Walter, W., Haferlach, C., Kern, W. and Haferlach, T., Deep learning algorithms support distinction of PV, PMF, and ET based on clinical and genetic markers. *Blood*, vol. 130, Suppl. 1, Abstract 4223, 2017, doi: 10.1182/blood.V130.Suppl_1.4223.4223.

McDermott, M., Zhang, H., Hansen, L., Angelotti, G. and Gallifant, J., A closer look at auroc and auprc under class imbalance. *Advances in Neural Information Processing Systems*, vol. 37, pp. 44102-44163, 2024.

McLornan, D. P., Psaila, B., Ewing, J., Innes, A., Arami, S., Brady, J., ... and Harrison, C. N., The management of myelofibrosis: A British society for haematology guideline. *British Journal of Haematology*, vol. 204, no. 1, pp. 136-150, 2024.

McNemar, Q., Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, vol. 12*,* no. 2, pp. 153-157, 1947.

Münch, M. M., Peeters, C. F., Van Der Vaart, A. W. and Van De Wiel, M. A., Adaptive group-regularized logistic elastic net regression. *Biostatistics*, vol. 22, no. 4, pp. 723-737, 2021.

NCBI GEO, Gene Expression Omnibus: GPL570, Available: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL570, Accessed on January 19, 2026a.

NCBI GEO, Gene Expression Omnibus: GSE26049, Available: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE26049, Accessed on January 19, 2026b.

NCBI GEO, Gene Expression Omnibus: GSE57793, Available: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE57793, Accessed on January 19, 2026c.

Ng, A. and Ma, T., CS229 Lecture Notes, Stanford University, Available: https://cs229.stanford.edu/main_notes.pdf, June 11, 2023. Accessed on January 22, 2026.

Passamonti, F., Cervantes, F., Vannucchi, A. M., Morra, E., Rumi, E., Pereira, A., ... and Tefferi, A., A dynamic prognostic model to predict survival in primary myelofibrosis: a study by the IWG-MRT (International Working Group for Myeloproliferative Neoplasms Research and Treatment). *Blood, The Journal of the American Society of Hematology*, vol. 115, no. (9), pp. 1703-1708, 2010.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... and Duchesnay, É., Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, vol. 12, pp. 2825-2830, 2011.

Pescia, C., Lopez, G., Cattaneo, D., Bucelli, C., Gianelli, U. and Iurlo, A., The molecular landscape of myeloproliferative neoplasms associated with splanchnic vein thrombosis: Current perspective. *Leukemia Research*, vol. 136,107420, 2024.

Ren, M., Chen, Q. and Zhang, J., Repurposing MALDI-TOF MS for effective antibiotic resistance screening in Staphylococcus epidermidis using machine learning. *Scientific Reports*, vol. 14, no. 1, 24139, 2024.

Riley, R. D., Archer, L., Snell, K. I., Ensor, J., Dhiman, P., Martin, G. P., ... and Collins, G. S., Evaluation of clinical prediction models (part 2): how to undertake an external validation study. *Bmj*, *384,* 2024.

Rung, J. and Brazma, A., Reuse of public genome-wide gene expression data. *Nature Reviews Genetics*, vol. 14, no. 2, pp. 89-99, 2013.

Saelmans, A., Seinen, T., Pera, V., Markus, A. F., Fridgeirsson, E., John, L. H., ... and Williams, R., Implementation and Updating of Clinical Prediction Models: A Systematic Review. *Mayo Clinic Proceedings: Digital Health*, vol. 3, no. 3, 100228, 2025.

Saito, T. and Rehmsmeier, M., The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PloS one*, vol. 10, no. 3, e0118432, 2015.

Salmi, M., Atif, D., Oliva, D., Abraham, A. and Ventura, S., Handling imbalanced medical datasets: review of a decade of research. *Artificial intelligence review*, vol. 57, no. (10), 273, 2024.

Schischlik, F., Transcriptional configurations of myeloproliferative neoplasms. *International Review of Cell and Molecular Biology*, vol. 366, pp. 25-39, 2022.

Schmidt, A., Bernhardt, C., Bürkle, D., Fries, S., Hannig, C. V., Jentsch-Ullrich, K., ... and Heidel, F. H., Diagnosis and treatment of MPN in real life: exploratory and retrospective chart review including 960 MPN patients diagnosed with ET or MF in Germany. *Journal of Cancer Research and Clinical Oncology*, vol. 149, no.10, pp. 7197-7206, 2023.

Shao, R., Ryder, C., Wang, L., Zhang, H., Moscinski, L., Martin, M., ... and Song, J., A Review of the Pathological and Molecular Diagnosis of Primary Myelofibrosis. *Cancers*, vol. 18, no. 1, 50, 2025.

Singhal, S., Mishra, R., Arora, S., Sharafeldin, N., Desai, R., Jain, T. and Vachhani, P., Epidemiology and outcomes of patients with myeloproliferative neoplasms in the United States: A SEER database analysis, years 2000-2021. *Blood*, vol. 146, suppl. 1, pp. 2048, 2025.

Sirinukunwattana, K., Aberdeen, A., Theissen, H., Sousos, N., Psaila, B., Mead, A. J., ... and Royston, D., Artificial intelligence–based morphological fingerprinting of megakaryocytes: a new tool for assessing disease in MPN patients. *Blood advances*, vol. 4, no. 14, pp. 3284-3294, 2020.

Skov, V., Thomassen, M., Riley, C. H., Jensen, M. K., Bjerrum, O. W., Kruse, T. A., ... and Larsen, T. S., Gene expression profiling with principal component analysis depicts the biological continuum from essential thrombocythemia over polycythemia vera to myelofibrosis. *Experimental Hematology*, vol. 40, no. 9, pp. 771-780, 2012.

Smith, C. J., Thomas, J. W., Ruan, G., Hyun, M. C., Bansal, R., McLaughlin, N., ... and Tefferi, A., A population-based study of outcomes in polycythemia vera, essential thrombocythemia, and primary myelofibrosis in the United States from 2001 to 2015: comparison with data from a Mayo Clinic single institutional series. *American journal of hematology*, vol. 96, no. 12, E464-E468, 2021.

Srisuwananukorn, A., Krull, J. E., Ma, Q., Zhang, P., Pearson, A. T. and Hoffman, R., Applications of artificial intelligence to myeloproliferative neoplasms: a narrative review. *Expert review of hematology*, vol. 17, no. 10, pp. 669-677, 2024.

Tay, J. K., Narasimhan, B. and Hastie, T., Elastic net regularization paths for all generalized linear models. *Journal of statistical software*, vol. 106, pp. 1-31, 2023.

Tefferi, A., Primary myelofibrosis: 2023 update on diagnosis, risk-stratification, and management. *American Journal of Hematology*, vol. 98, no. 5, pp. 801-821, 2023.

Tefferi, A., Guglielmelli, P., Lasho, T. L., Gangat, N., Ketterling, R. P., Pardanani, A. and Vannucchi, A. M., MIPSS70+ version 2.0: mutation and karyotype-enhanced international prognostic scoring system for primary myelofibrosis. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*, vol. 36, no. 17, pp. 1769-1770, 2018a.

Tefferi, A., Guglielmelli, P., Nicolosi, M., Mannelli, F., Mudireddy, M., Bartalucci, N., ... and Vannucchi, A. M., GIPSS: genetically inspired prognostic scoring system for primary myelofibrosis. *Leukemia*, vol. 32, no. 7, pp. 1631-1642, 2018b.

Tefferi, A. and Vannucchi, A. M., Risk models in myelofibrosis—the past, present, and future. *American journal of hematology*, vol. 99, no. 4, pp. 519-522, 2024.

Thiele, J., Kvasnicka, H. M., Gianelli, U., Arber, D. A., Tefferi, A., Vannucchi, A. M., ... and Orazi, A., Evolution of WHO diagnostic criteria in "Classical Myeloproliferative Neoplasms" compared with the International Consensus Classification. *Blood cancer journal*, vol. 15, no. 1, 31, 2025.

Thiele, J., Kvasnicka, H. M., Orazi, A., Gianelli, U., Gangat, N., Vannucchi, A. M., ... and Tefferi, A., The international consensus classification of myeloid neoplasms and acute leukemias: myeloproliferative neoplasms. *American journal of hematology*, vol. 98, no. 1, pp. 166-179, 2023.

Torang, A., Gupta, P. and Klinke, D. J., An elastic-net logistic regression approach to generate classifiers and gene signatures for types of immune cells and T helper cell subsets. *BMC bioinformatics*, vol. 20, no. 1, 433, 2019.

Van Calster, B., Collins, G. S., Vickers, A. J., Wynants, L., Kerr, K. F., Barreñada, L., ... and Steyerberg, E. W., Evaluation of performance measures in predictive artificial intelligence models to support medical decisions: overview and guidance. *The Lancet Digital Health,* 2025.

Varma, S. and Simon, R., Bias in error estimation when using cross-validation for model selection. *BMC bioinformatics*, vol. 7, no. 1, 91, 2006.

Verstovsek, S., Yu, J., Scherber, R. M., Verma, S., Dieyi, C., Chen, C. C. and Parasuraman, S., Changes in the incidence and overall survival of patients with myeloproliferative neoplasms between 2002 and 2016 in the United States. *Leukemia & Lymphoma*, vol. 63, no. (3), pp. 694-702, 2022.

Yates, L. A., Aandahl, Z., Richards, S. A. and Brook, B. W., Cross validation for model selection: a review with examples from ecology. *Ecological Monographs*, vol. 93, no. 1, e1557, 2023.

Yu, D., Zhang, H., Song, Y., Tao, Y., Zhou, F., Wang, Z., ... and Zhang, L., Artificial intelligence-based quantitative bone marrow pathology analysis for myeloproliferative neoplasms. *Haematologica*, vol. 110, no. 11, 2691 – 2701, 2025.

Zou, H. and Hastie, T., Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 67, no. 2, pp. 301-320, 2005.

Biographies

Yeşim Kayabas Alkan completed her undergraduate and master's degrees in the Department of Industrial Engineering at Karabuk University. She is currently pursuing her doctoral studies in the Department of Industrial Engineering at Gazi University. In 2023, she served as a research assistant at the Department of Industrial Engineering at Mudanya University, and since 2024, she has been continuing her academic work as a research assistant at the Department of Industrial Engineering at Tokat Gaziosmanpasa University. Her research interests include multi-criteria decision-making methods, decision-making in fuzzy environments, and decision support systems.

Mehmet Kabak has been a faculty member at the Department of Industrial Engineering at Gazi University since 2015. Appointed to the professorship in 2018, the author teaches courses in Operations Research, Mathematical Modelling, Project Management, and Multi-Criteria Decision Making. His research areas include mathematical modelling, multi-criteria decision making, energy, and strategic management. Prof. Dr. Mehmet Kabak has published nearly 70 scientific articles, more than 25 of which are indexed in SCI, SCI-E, and SSCI journals, presented over 80 conference papers at national and international conferences, edited 2 international books, and authored more than 15 international book chapters.