Journal of Associated Medical Sciences 2026; 59(2): 187-195

JAMS
Journal of Associated Medical Sciences

Scopus Indexed & Thai-Journal Citation Index Centre (TCI)

Journal of Associated Medical Sciences

Journal homepage: https://www.tci-thaijo.org/index.php/bulletinAMS/index

Towards precision hematology: Machine learning-driven comparative feature importance analysis of hematologic parameters in newly diagnosed chronic myeloid leukemia vs healthy controls

Hossein Ayatollahi[1,2], Arefeh Mazhari[1,2], Seyed Mohammad Tabatabaei[3,4], Samaneh Talebi[5], Mahnaz Dehghan[2], Maryam Sheikhi[2], Yasamin Ayatollahi[2], Zahra Darvish Khalilabadi[1,2*]

[1]Department of Hematology and Blood Banking, Mashhad University of Medical Sciences, Mashhad, Iran.
[2]Cancer Molecular Pathology Research Center, Mashhad University of Medical Sciences, Mashhad, Iran.
[3]Department of Medical Informatics, Mashhad University of Medical Sciences, Mashhad, Iran.
[4]Applied Biomedical Research Center, Basic Sciences Research Institute, Mashhad University of Medical Sciences, Mashhad, Iran.
[5]Center for Healthcare Data Modeling, Department of Biostatistics and Epidemiology, School of Public Health, Shahid Sadoughi University of Medical Sciences, Yazd, Iran.

ARTICLE INFO

Article history:
Received 5 October 2025
Accepted as revised 2 February 2026
Available online 16 February 2026

Keywords:
Chronic myeloid leukemia, CML, machine learning, Random Forest, simulated annealing, NLR, BLR, dNLR.

ABSTRACT

Background: Chronic Myeloid Leukemia (CML) is a myeloproliferative neoplasm characterized by uncontrolled granulocytic proliferation and is often initially suspected based on peripheral blood smear findings. Utilizing machine learning to analyze routinely available Complete Blood Count (CBC) parameters and derived inflammatory indices may facilitate early identification of CML at the time of initial diagnosis.

Objectives: The objective of this study is to evaluate the diagnostic value of routinely available Complete Blood Count (CBC) parameters and derived inflammatory indices for the early identification of chronic myeloid leukemia at the time of initial diagnosis, using machine learning–based models.

Materials and methods: This study was conducted on 295 newly diagnosed cases of CML and 340 normal control samples. A total of 49 factors were subjected to study, including the following: variables included in the CBC of patients, inflammatory indices, and demographic data. Logistic regression analysis was performed to identify relevant variables, resulting in the selection of 22 features. Subsequently, multiple machine learning algorithms, including Random Forest (RF), Recursive Feature Elimination (RFE), Simulated Annealing (SA), Decision Tree (DT), K-Nearest Neighbor (KNN), and Xg-Boost (XGB), were applied to evaluate the diagnostic performance pf the selected features.

Results: The findings of this study indicate that the factors most pertinent in initial diagnosis in comparison with normal control include the WBC count, the relative percentage of cells including neutrophils, monocytes, and lymphocytes, and a series of indicators related to RBC such as RBC count and RDW-CV, as well as the index of inflammatory NLR, and BLR and PDW.

Conclusion: This study demonstrates that machine learning models based solely on routinely available CBC parameters and derived inflammatory indices can support the early identification of CML at the time of initial diagnosis. In addition to leukocyte-related variables, RBC-related and inflammatory indices provided complementary diagnostic information, highlighting their potential value in early-stage CML detection. The application of machine learning techniques could prioritize the development of more user-friendly dashboards to facilitate the diagnosis of CML at initial diagnosis.

* **Corresponding contributor.**
Author's Address: *Cancer Molecular Pathology Research Center, Mashhad University of Medical Sciences, Mashhad, Iran.*
E-mail address: zdarvish99@gmail.com
doi: 10.12982/JAMS.2026.050
E-ISSN: 2539-6056

Introduction

Chronic myeloid leukemia (CML) is a myeloproliferative neoplasm, defined as a disorder characterized by uncontrolled granulocyte proliferation. It is estimated that CML accounts for approximately 15% of all adult leukemias.[1] The disease predominantly affects older age groups; however, it can manifest in any age demographic. Approximately 50% of cases of this disease are identified incidentally during a routine medical examination. In instances where symptoms are present, the most observed manifestations are fatigue, weight loss, and fever.[2] The initial stage of diagnosis is the examination of a peripheral blood smear. Subsequent molecular testing is then performed to confirm the BCR-ABL fusion.

Machine learning is a subfield of artificial intelligence that employs sophisticated algorithms to analyze intricate data sets, such as those pertaining to disease prevalence or disease outcomes. The application of machine learning enables the development of models capable of accurate estimation based on the analysis of raw data.[1]

The application of machine learning techniques has been limited within the context of CML.[3] In spite of the use of Tyrosine kinase inhibitors (TKIs), some patients require alternative treatments in accordance with their specific conditions. These treatments may be identified through the use of machine learning algorithms, enabling the selection of the optimal treatment.[1] In a study on CML, machine learning models have been employed to distinguish CML-related mature neutrophils from normal mature neutrophils, with a sensitivity of ≤95.30% and ≤95.80% specificity.[4] In addition, models have been constructed to predict treatment responses within a period of 6 to 18 months, utilizing data obtained from cyTOF.[5] Attempts have been made to apply machine learning in the field of diagnosis, particularly in the context of CML and other leukemias. These efforts have largely focused on the digitization of microscopic smear images. The training group employed blood smears with a definitive diagnosis of CML. Subsequently, the models created were subjected to a verification process with the test group.[6-8] A study was conducted on cell blood counts from CML patients just five years before the definitive diagnosis. The results indicated that the BCR-ABL test result can be predicted with a high degree of accuracy using blood cell count results.[9]

This study aims to examine the factors that can be identified at the initial stage of diagnosis through the utilization of a straightforward and readily accessible CBC (complete blood count) test, rather than the recommendation of a digital blood smear, which has been the focus of investigation in previous studies. The aforementioned factors have been identified as essential in the diagnostic process, particularly in the initial diagnosis of CML patients in comparison to normal controls. A diagnosis of CML can be reached with greater efficiency by investigating these factors, which can then be followed by the appropriate CML confirmation tests, if necessary.

Material and method

Patients and data collection

This retrospective study was conducted between 2010 and 2023 on 325 newly diagnosed cases of CML and 341 normal control samples at the Cancer Molecular Pathology Research Center, Mashhad University of Medical Sciences, Mashhad, Iran. The demographic data of the patients was extracted from the electronic medical records. The information encompasses the complete set of variables included in the CBC of patients and demographic data such as age and gender. The data obtained from CBC were utilized to calculate the inflammatory ratios, including neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio (PLR), monocyte-to-lymphocyte ratio (MLR), platelet-to-hemoglobin ratio (PHR), basophil-to-lymphocyte ratio (BLR), eosinophil-to-lymphocyte ratio (ELR), derived neutrophil-to-lymphocyte ratio (DNLR), neutrophil-to-lymphocyte*platelet ratio (NLPR), systemic inflammation index (SII), and systemic inflammation response Index (SIR-I). Furthermore, a comprehensive examination of the peripheral blood smear was conducted to ascertain the morphology of the blood cells, particularly in view of the presence of immature myeloid cells, and to evaluate the cells' differentiation. All the results are confirmed by a pathologist and hematologists. A total of 49 factors were subjected to study, including the following: age, gender, RBC count, hemoglobin, hematocrit, MCV, MCH, MCHC, RDW-CV, platelet count, PDW, MPV, WBC count, the absolute and relative blast count, absolute and relative promyelocyte count, absolute and relative myelocyte count, absolute and relative Metamyelocyte count, absolute and relative band cell count, absolute and relative neutrophil count, absolute and relative eosinophil count, absolute and relative basophil count, absolute and relative monocyte count, absolute and relative lymphocyte count, absolute and relative nRBC count, NLR, PLR, MLR, PHR, BLR, ELR, DNLR, NLPR, SIRI, and SII. Methods were carried out in accordance with relevant guidelines and regulations, including the Declaration of Helsinki.

Statistical analysis

In the initial phase of the study, 325 patients and 341 controls were considered, with those exhibiting deficiencies in information being excluded. Ultimately, 295 patients and 340 controls were included in the study. A total of 49 factors were investigated in the patient group in comparison to the controls. A logistic regression analysis was conducted between the control group and the patient utilizing R 4.4.2 and RStudio 2024.09.1, resulting in the incorporation of a total of 22 factors into the model. Furthermore, patients were divided into two groups—the training (80%) and the testing (20%) groups—for the purpose of developing

machine learning models. The following section of the analysis employs a variety of machine learning algorithms, including Random Forest (RF), Recursive Feature Elimination (RFE), Simulated Annealing (SA), Decision Tree (DT), K-Nearest Neighbor (KNN), and Xg-Boost (XGB). Subsequently, the importance of the factors in new cases was evaluated in comparison to the control group. Subsequently, to evaluate the predictive performance of the models, a range of metrics were employed, including sensitivity, specificity, accuracy, and the area under the ROC curve (AUC). To conclude, an overall assessment was conducted in order to ascertain the relative importance of each factor, with assistance from the derived importance of each factor from each model. Figure 1 depicts the sequence of steps undertaken during the study. Statistical analysis was performed using the R 4.4.2 and RStudio 2024.09.1 and packages including "Random Forest", "caret", "e1071", "XGBoost", "rminer" and "rpart" was employed for feature importance analysis with a significance level of 0.05.

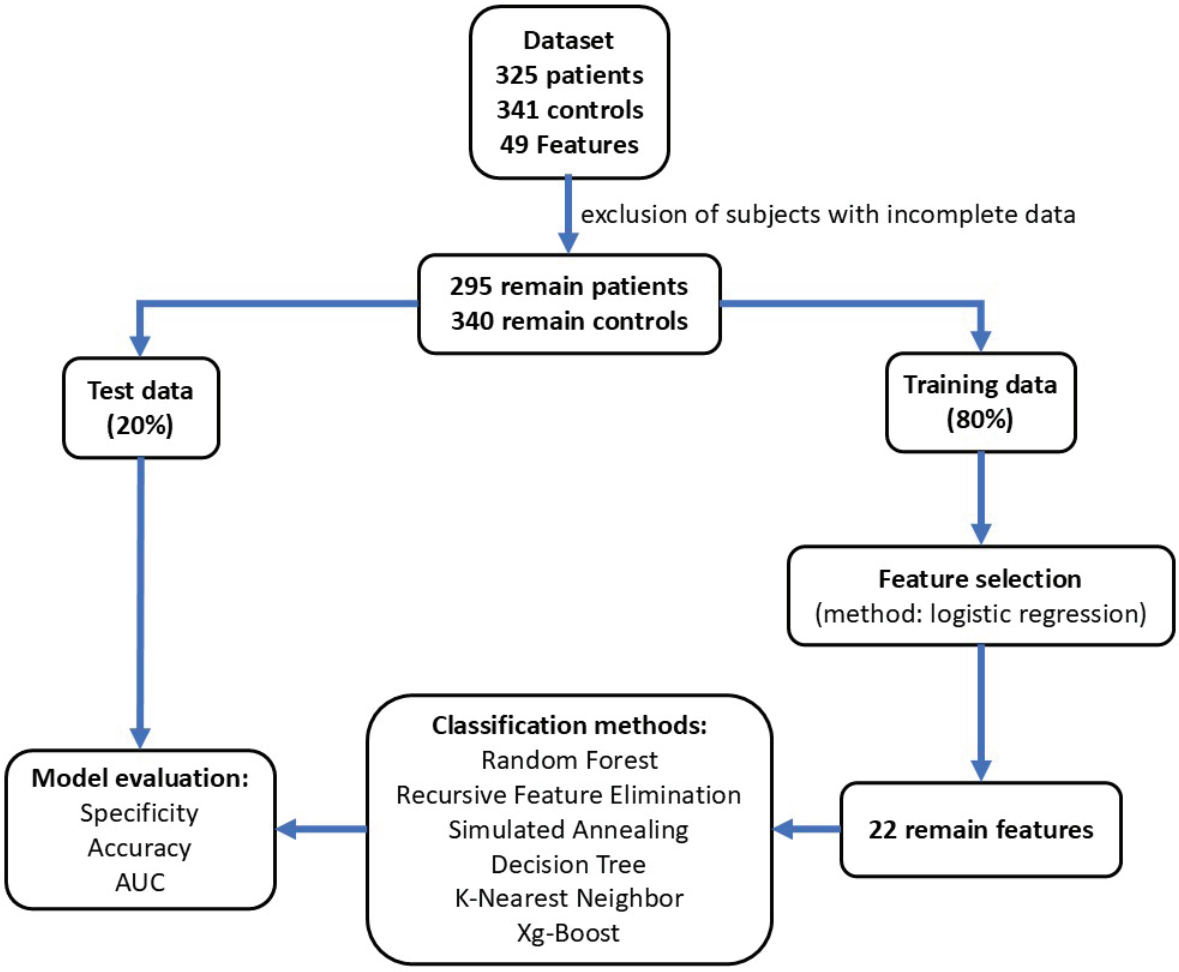

Figure 1. *The framework of the study design.*

Result

In total, 295 patients were included in the study, with 53% men (N=155) and 47% women (N=140) representing the sample population. An investigation was carried out into the 49 factors derived from the initial diagnoses of the subjects.

At the outset of the investigation, a logistic regression analysis was conducted between the two study groups, resulting in the identification of 22 statistically significant factors. The aforementioned factors encompass age, gender, RBC count, hematocrit, MCV, RDW-CV, PLT, PDW, WBC, neutrophil relative count, eosinophil relative count, basophil relative count, lymphocyte relative count, monocyte relative count, NLR, PLR, PHR, ELR, BLR, dNLR, and NLPR.

As shown in Table 1, there was a statistically significant difference between CML patients and healthy controls in all hematologic parameters. Notably, the median WBC count was significantly higher in the CML group (p<0.001).

Table 1. Comparison of laboratory parameters between normal control and CML patients.

Parameters	Control normal (N=340)	Patients (N=295)	*p* value
Age (year)	51 (39, 64)	44 (32, 56)	<0.001
RBC (x10^6/µL)	4.7 (4.4, 4.9)	3.7 (3.1, 4.3)	<0.001
HCT (%)	40 (37.9, 42)	33.6 (28.9, 38.7)	<0.001
MCV (fL)	85.8 (83.5, 88.4)	89.4 (85.2, 93.5)	<0.001
RDW-CV (%)	13.2 (12.7, 13.8)	17.3 (16,18.7)	<0.001
PLT (x10^3/µL)	246 (207, 287)	322 (224, 516)	<0.001
PDW (fL)	15.9 (12,16.9)	12.5 (11.2,14.4)	<0.001
MPV (fL)	9.5 (8.4,10.4)	9.8 (9,10.5)	<0.001
WBC (x10^3/µL)	6.7 (5.7, 7.7)	92.2 (33.1,156.7)	<0.001
Neutrophils (%)	56.4 (50.9, 62.8)	43 (32, 56)	<0.001
Lymphocytes (%)	32.9 (28, 38.4)	6 (3,11)	<0.001
Monocytes (%)	6.2 (4.7, 8.3)	0 (0, 2)	<0.001
NLR	1.7 (1.3, 2.2)	7.3 (4.1,15)	<0.001
PLR	114.5 (91,142.3)	99.2 (41.8, 215.3)	0.023
PHR	1.8 (1.4, 2.1)	2.9 (2, 5)	<0.001
BLR	0.004 (0.002, 0.008)	0.00 (0.00, 0.00)	<0.001
ELR	0.07 (0.05, 0.11)	0.14 (0.00, 0.97)	0.008
DNLR	1.29 (1.04,1.69)	0.75 (0.47, 1.27)	<0.001
NLPR	0.67 (0.52, 0.96)	2.35 (0.99, 4.8)	<0.001

In the subsequent phase, the aforementioned factors were employed within the RF, RFE, SA, DT, KNN, and XGB models to ascertain the feature importance within the context of each model. The results of this step are presented in Figure 2.

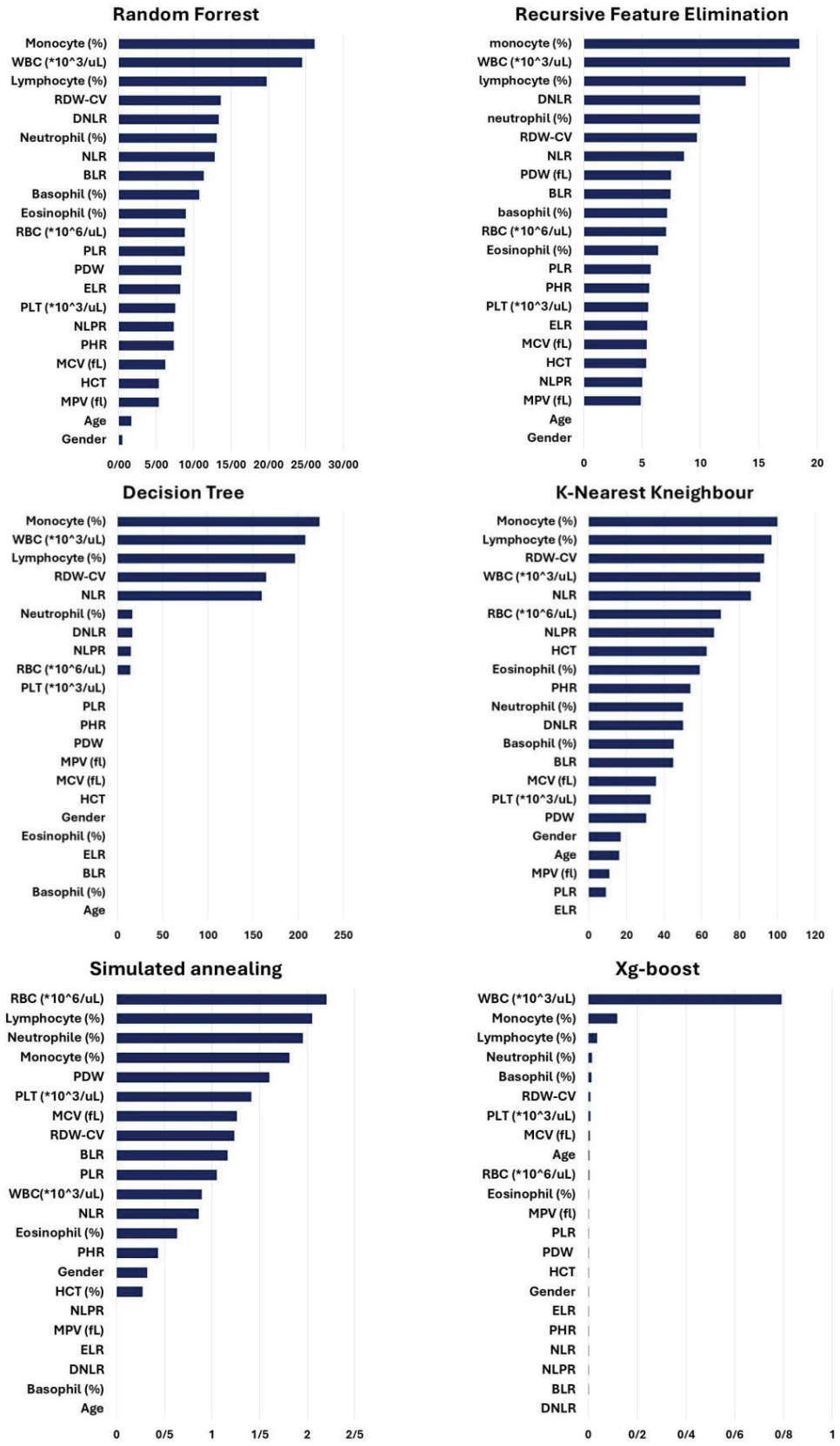

Figure 2. *Importance index of variables.*

The findings of this study indicate that the factors most pertinent in initial diagnosis in comparison with normal control include the WBC count, the relative percentage of cells including neutrophils, monocytes, and lymphocytes, and a series of indicators related to RBC such as RBC count and RDW-CV, as well as the index of inflammatory NLR, and BLR and PDW.In order to assess the robustness of the proposed models, a ROC curve has been constructed, the result of which is illustrated in Figure 3. The ROC curve analysis indicates that the RF and RFE models with an AUC level of 99 exhibit the greatest predictive efficacy, while other models with an AUC level of 98 demonstrate the second-highest level of performance.

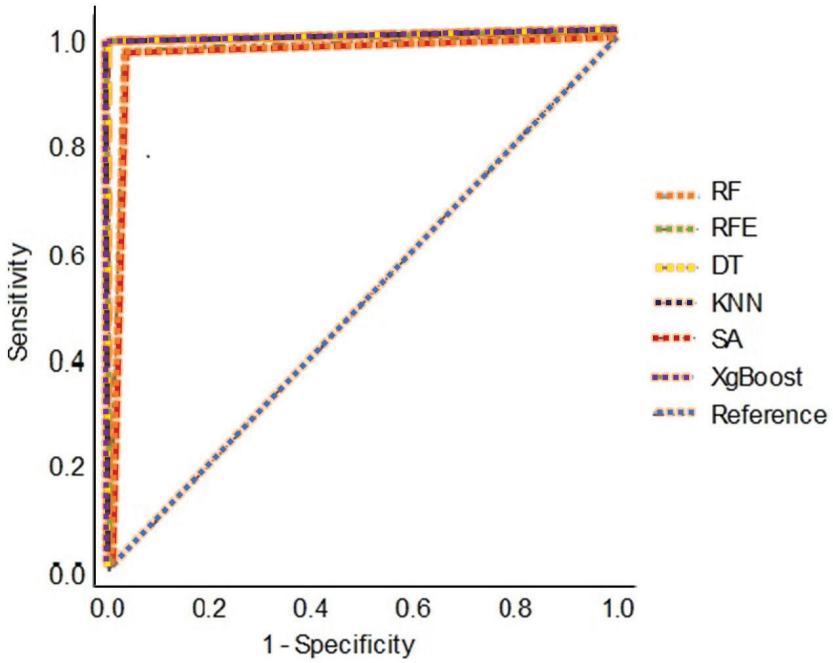

Figure 3. *Receiver operating characteristic (ROC) curve for different machine learning models.*

The subsequent phase entails the computation of sensitivity, specificity, and area under the curve (AUC) indices, which are utilized to assess the performance of these models, as illustrated in Table 2. Among the used models, the RF and RFE models exhibit the greatest sensitivity (equating to a value of 1) and the highest specificity (equating to a value of 0.98), respectively. These models indicate that WBC, the relative count of monocytes, neutrophils, and lymphocytes, as well as indices such as DNLR and NLR, as well as RDW, may be of diagnostic importance.

Table 2. Comparison of predictive power indices for models.

Model	**Sensitivity**	**Specificity**	**Accuracy** (0.95% CI)	**AUC** (0.95% CI)
Random Forest	1.00	0.98	0.99 (0.95, 0.99)	0.992 (0.97, 1)
Recursive Feature Alimination	0.97	0.96	0.97 (0.92, 0.99)	0.97 (0.93, 1)
Simulated Annealing	1.00	0.98	0.99 (0.95, 0.99)	0.992 (0.97, 1)
Decision Tree	0.97	0.96	0.97 (0.92, 0.99)	0.97 (0.93, 1)
K-Nearest Neighbor	0.97	0.96	0.97 (0.92, 0.99)	0.97 (0.93, 1)
Xg-Boost	0.97	0.96	0.97 (0.92, 0.99)	0.97 (0.93, 1)

To enhance the clinical utility of our findings, ROC curve analysis was conducted for the important variables based on the RF model. The optimal cutoff value WBC count was 10.35 (AUC: 0.97, Sensitivity: 90.8%, Specificity: 99.7%). Detailed thresholds for other parameters are summarized in Table 3.

Table 3. Cut-off points for the important variables based on the RF model.

Variable	Cut-off point	Sensitivity	Specificity	AUC (0.95% CI)
WBC (×10^3/uL)	>10.35	90.8	99.7	0.97 (0.95,0.98)
Lymphocyte (%)	<21.03	94.4	94.6	0.94 (0.92,0.96)
RDW-CV (%)	>15.15	87.5	93.1	0.93 (0.90,0.96)
Monocyte (%)	<3.05	97.1	91.2	0.98 (0.96,0.99)
DNLR	<0.82	85.9	85.7	0.91 (0.89,0.95)
Neutrophil (%)	<45.03	83.9	87.2	0.89 (0.87,0.91)

Discussion

As previously indicated, recent research findings have demonstrated that machine learning is a highly prevalent method in the field of medical diagnostics and prognostication. While this method has received less attention in the context of CML. The present study evaluated multiple machine learning models to identify key predictors of CML at the initial stage of diagnosis using CBC parameters, derived inflammatory indices, and demographic data.

The application of multiple machine learning algorithms in this study was intended to enhance the robustness and reliability of the identified predictors for early CML detection. Different machine learning methods capture distinct patterns within the data, and agreement across models increases confidence in the observed findings.[10]

Despite methodological differences among the applied algorithms, a high degree of consistency was observed in the ranking of key features. Variables such as WBC count, relative proportions of neutrophils, lymphocytes, and monocytes, as well as RDW-CV and inflammatory indices including NLR and DNLR, were repeatedly identified as important predictors across Random Forest, Recursive Feature Elimination, Simulated Annealing, and XGBoost models. Such convergence across independent modeling strategies supports the biological plausibility and clinical relevance of these CBC-derived parameters in the early identification of CML.[11,12]

Tree-based models, particularly Random Forest and Decision Tree algorithms, demonstrated strong diagnostic performance and relative interpretability, making them well-suited for clinical screening applications. Random Forest has been shown to perform robustly in biomedical datasets characterized by nonlinearity and complex feature interactions, while also providing stable feature importance measures.[13] Feature selection techniques such as Recursive Feature Elimination and Simulated Annealing further strengthened the analysis by reducing dimensionality and mitigating overfitting, thereby enhancing model generalizability.[14,15]

Other algorithms, including K-Nearest Neighbor and XGBoost, contributed complementary perspectives but present specific limitations. KNN performance is sensitive to feature scaling and class imbalance, while XGBoost requires careful hyperparameter tuning to avoid overfitting, particularly in single-center retrospective datasets.[16,17]

Machine learning has been increasingly applied to leukemia diagnosis and management, but most studies focus on prognosis or subtype classification rather than early clinical diagnosis of CML. Previous Machine learning studies using CBC or molecular markers reported moderate to high accuracy but did not systematically integrate inflammatory or RBC-related indices. The integration of low-cost, interpretable biomarkers into machine learning models provides a potentially useful tool for early screening, especially where molecular assays are unavailable. Logistic regression identified several significant predictors, including WBC count, neutrophil and lymphocyte percentages, NLR, BLR, RDW, and PDW. In addition, the Random Forest and Simulated Annealing models consistently ranked these variables among the most important features, contributing to the strong diagnostic performance (sensitivity =1.00, specificity =0.98) observed in our analysis.

The high feature importance of WBC aligns with established clinical guidelines, where leukocytosis is a hallmark of CML. In the context of CML, WBC count is a critical parameter, ranking amongst the most significant factors in the clinical management of the disease. A study of 56 CML patients revealed that individuals presenting with elevated WBC levels, abdominal discomfort, and splenomegaly should undergo evaluation for CML.[18] Furthermore, another research has indicated that a WBC count exceeding 20,000/μL in conjunction with basophilia may serve as a useful indicator for the identification of CML.[19]

Relative percentages of neutrophils, lymphocytes, and monocytes provide additional information beyond absolute counts, reflecting early disease-related shifts in white cell distributions. relative percentage of monocytes and lymphocytes was identified as a

significant factor in most models. Despite findings from some studies indicating that absolute cell count is not associated with molecular response, there has been minimal discussion in the literature regarding its significance during the initial diagnosis.[20] In some studies, an absolute lymphocyte count (ALC) above 4000/uL within 3-5 months after treatment has been reported to be associated with a decreased overall survival (OS) rate.[21] The relative percentage of neutrophil cells is another significant factor that must be considered. Neutrophilia is a common initial finding in CML patients.[1]

Inflammatory indices such as NLR and BLR also emerged as significant predictors. Although their role in CML diagnostics has been less explored, their accessibility and low cost make them attractive for early screening and prioritization.

Similarly, red blood cell distribution width (RDW) and platelet distribution width (PDW) were among the top predictors identified. While RDW has been recognized as a prognostic marker in other hematologic disorders, its application in early CML diagnosis has not been extensively evaluated.[22] PDW remains understudied in the context of CML, but its association with platelet variability could reflect underlying pathophysiological changes in hematopoiesis.[21] The identification of accessible, cost-effective predictors highlights the potential for developing decision support tools that flag high-risk individuals for confirmatory testing. Future studies should validate these models across multi-center cohorts and explore the integration of additional clinical variables, such as symptomatology and physical examination findings. Furthermore, future research could prioritize the development of more user-friendly dashboards to facilitate the diagnosis of CML.

Limitations

One potential limitation of this study is the relatively small sample size. This study is limited by its retrospective design and single-center data, which may affect generalizability. External validation and prospective studies would strengthen the evidence for clinical implementation.

Conclusion

CML is a myeloproliferative disorder, defined as a disorder characterized by uncontrolled granulocyte proliferation. While several models performed exceptionally well, we recommend the Random Forest model for high-precision diagnostic tasks and the Decision Tree model for its ease of visual interpretation in clinical settings. This study aims to examine the factors that can be identified at the initial stage of diagnosis through the utilization of a straightforward and readily accessible CBC test, rather than the recommendation of a digital blood smear, which has been the focus of investigation in previous studies. Key discriminative features included WBC count, relative leukocyte proportions, RBC-related indices (RBC count and RDW-CV), and inflammatory markers such as NLR, BLR, and PDW. These findings highlight the potential of integrating CBC-derived features with machine learning approaches to enable efficient early screening and support clinical decision-making. Furthermore, future research could prioritize the development of more user-friendly dashboards to facilitate the diagnosis of CML.

Funding

This article has not been supported or received any grant or fund.

Ethical approval

This study received approval from the Ethics Committee of MUMS, and all participants provided informed consent (Ethics code: IR.MUMS.MEDICAL.REC.1402.525).

Conflict of interest

Authors declared no conflict of interest.

Data availability

All the data have been included in the manuscript and will be made available upon publication of the manuscript.

CRediT authorship contribution statement

Hossein Ayatollahi: conceptualization, data curation, investigation, methodology, visualization, project administration, validation, writing: original draft, review and edit; **Arefeh Mazhari**: data curation, methodology, writing: original draft, review and edit; **Seyed Mohammad Tabatabaei**: formal analysis, methodology, writing: original draft, review and edit; **Samaneh Talebi**: formal analysis, methodology, writing : original draft, review and edit; **Mahnaz Dehghan**: data curation, writing: review and edit; **Maryam Sheikhi**: methodology, writing: review and edit; **Yasamin Ayatollahi**: data curation, writing: review and edit; **Zahra Darvish Khalilabadi**: conceptualization, data curation, investigation, methodology, visualization, writing: original draft, review and edit.

Acknowledgements

The authors wish to thank all colleagues at Mashhad University of Medical Sciences.

References

[1] Elhadary M, Elsabagh AA, Ferih K, Elsayed B, Elshoeibi AM, Kaddoura R, et al. Applications of machine learning in chronic myeloid leukemia. Diagnostics. 2023;13.

[2] Pandey MK, Pal S. Evaluation of chronic myelogenous leukemia (CML) as the chronic phase of disease using machine learning techniques. Int J Mech Eng Educ. 2022; 7: 1-12.

[3] Shanbehzadeh M, Afrash MR, Mirani N, Kazemi-

Arpanahi H. Comparing machine learning algorithms to predict 5-year survival in patients with chronic myeloid leukemia. BMC Med Inform Decis Mak. 2022; 22(1): 236. doi:10.1186/s12911-022-01980-w.

[4] Ni W, Tong X, Qian W, Jin J, Zhao H. Discrimination of malignant neutrophils of chronic myelogenous leukemia from normal neutrophils by support vector machine. Comput Biol Med. 2013; 43(9): 1192-5. doi:10.1016/j.compbiomed.2013.06.004

[5] Bergenheim MB. Towards prediction of CML treatment outcomes with machine learning and CyTOF data: The University of Bergen; 2024.

[6] Cerrato TR. Use of artificial intelligence to improve access to initial leukemia diagnosis in low-and middle-income countries. American Society of Clinical Oncology; 2020.

[7] Dese K, Raj H, Ayana G, Yemane T, Adissu W, Krishnamoorthy J, et al. Accurate machine-learning-based classification of leukemia from blood smear images. Clin Lymphoma Myeloma Leuk. 2021; 21(11): e903-e14. doi: 10.1016/j.clml. 2021. 06.025.

[8] Salah HT, Muhsen IN, Salama ME, Owaidah T, Hashmi SK. Machine learning applications in the diagnosis of leukemia: Current trends and future directions. Int J Lab Hematol. 2019; 41(6): 717-25. doi:10.1111/ijlh.13089

[9] Hauser RG, Esserman D, Beste LA, Ong SY, Colomb DG, Bhargava A, et al. A aachine learning model to successfully predict future diagnosis of chronic myelogenous leukemia with retrospective electronic health records data. Am J Clin Pathol. 2021; 156(6): 1142-8. doi:10.1093/ajcp/aqab086

[10] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning. Springer series in statistics New-York; 2009.

[11] Guyon I, Elisseeff A. An introduction to variable and feature selection. J Mach Learn Res. 2003;(3): 1157-82.

[12] Saeys Y, Inza I, Larranaga P. A review of feature selection techniques in bioinformatics. Bioinformatics. 2007; 23(19): 2507-17. doi:10.1093/bioinformatics/btm344.

[13] Couronné R, Probst P, Boulesteix A-L. Random forest versus logistic regression: a large-scale benchmark experiment. BMC Bioinformatics. 2018; 19(1): 270. doi: 10.1186/s12859-018-2264-5.

[14] Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. Machine learning. 2002; 46(1): 389-422.

[15] Kirkpatrick S, Gelatt Jr CD, Vecchi MP. Optimization by simulated annealing. Science. 1983; 220(4598): 671-80.

[16] Chen T. XGBoost: A scalable tree boosting system. Cornell University. 2016.

[17] Kotsiantis SB, Zaharakis I, Pintelas P. Supervised machine learning: A review of classification techniques. Emerging artificial intelligence applications in computer engineering. 2007; 160(1): 3-24.

[18] Algahtani FH, Alqahtany FS. Evaluation and characterisation of Chronic myeloid leukemia and various treatments in Saudi Arabia: A retrospective study. J Infect Public Health. 2020; 13(2): 295-8. doi:10.1016/j.jiph.2019.12.006

[19] Fenu E, O'Neill SS, Insuasti-Beltran G. BCR-ABL1 p210 screening for chronic myeloid leukemia in patients with peripheral blood cytoses. Int J Lab Hematol. 2021; 43(6): 1458-64. doi:10.1111/ijlh.13635

[20] Pepedil-Tanrikulu F, Buyukkurt N, Korur A, Sariturk C, Aytan P, Boga C, et al. Significance of lymphocyte count, monocyte count, and lymphocyte-to-monocyte ratio in predicting molecular response in patients with chronic myeloid leukemia: a single-centre experience. Clin Lab. 2020; 66(3). doi:10.7754/Clin.Lab.2019.190628

[21] Mulas O, Mola B, Madeddu C, Caocci G, Macciò A, Nasa GL. Prognostic role of cell blood count in chronic myeloid neoplasm and acute myeloid leukemia and its possible implications in hematopoietic stem cell transplantation. Diagnostics. 2022; 12(10): 2493. doi.org/10.3390/diagnostics12102493

[22] Iriyama N, Hatta Y, Kobayashi S, Uchino Y, Miura K, Kurita D, et al. Higher red blood cell distribution width is an adverse prognostic factor in chronic-phase chronic myeloid leukemia patients treated with tyrosine kinase inhibitors. Anticancer Res. 2015; 35(10): 5473-8. PMID: 26408711