SOHO ANNUAL MEETING 2024

“ LEVERAGING MACHINE LEARNING FOR RAPID AND ACCURATE DIAGNOSIS OF ACUTE LEUKEMIA”

BEULAH PRISCILLA MADDIRALA[1], GURLEEN OBEROI[1], ANAND KAKARLA[2], BEENA CHANDRASEKHAR[1], AJAY GUPTA[1], REENA NAKRA[1], VANDANA LAL[1]

1 Department of Hematopathology, National Reference Laboratory, Dr Lal Pathlabs, New Delhi, INDIA

2 School of Technology, Woxsen University, Telangana, INDIA

Dr Lal PathLabs

INTRODUCTION

- Early detection of acute leukemia (AL) is crucial for timely intervention in the disease course [1,2].
- Integrating machine learning (ML) predictive modeling holds the potential to revolutionize early screening and rapid detection of AL, thereby reducing the delay for higher-center referral [3,4,5].

AIM

1. To assess the utility of leukocyte cell population data (CPD) through ML models in detecting AL.
2. To subclassify AL into acute myeloid leukemia (AML) and acute lymphoid leukemia (ALL) based on CPD morphometry at a premicroscopic level.

METHOD

- A total of 1211 cases were analyzed, of which 810 were confirmed cases of AL by morphology, immunophenotype, and/or molecular methods.
- The remaining 401 cases were benign. Standard leukocyte parameters and CPD from a Sysmex XN1000 analyzer (WDF Channel) were analyzed individually across each disease category (AML, ALL, Benign).
- Statistical and data visualization tools, including heat-map and UMAP 2-dimensional projection, were used.
- Different machine learning models (LightGBM, CatBoost, TabNet, XGBoost), were trained for disease categorization, and the best model was selected based on the accuracy on the test set using k--fold cross validation (k=5).
- The Shapley tool was used to assess the contribution by each feature for better AL prediction.

RESULTS

Fig .1 (a) Feature-wise median disease category and the corresponding (b) heat map obtained with min-max standardization

	AML	ALL	Benign
WBC(10^3/uL)	35.4	37.96	8.16
NEUT#(10^3/uL)	4.935	1.47	4.41
NEUT#&(10^3/uL)	3.81	1.23	4.36
NE-SSC(ch)	135.5	135.75	141.5
NE-SFL(ch)	58.55	53.3	54.5
NE-FSC(ch)	72.5	75	81.1
NE-WX	546.5	548.5	333
NE-WY	1707.5	1538	618
NE-WZ	1363	1373	641
LYMPH#(10^3/uL)	9.31	26.115	2.52
LYMP#&(10^3/uL)	7.845	25.68	2.51
LY-X(ch)	90.1	82.55	67.8
LY-Y(ch)	67.85	71.6	76.5
LY-Z(ch)	55.9	57.55	51
LY-WX	631.5	671	638
LY-WY	1390	1384.5	925
LY-WZ	1016.5	934.5	540
MONO#(10^3/uL)	11.725	3.715	0.61
MO-X(ch)	113.7	107.6	108.1
MO-Y(ch)	117.9	112.1	126
MO-Z(ch)	65.7	66.3	58.8
MO-WX	398.5	356	278
MO-WY	1061	1053	671
MO-WZ	987	1167	613
IG#(10^3/uL)	0.45	0.21	0.03
HFLC#(10^3/uL)	0.24	0.38	0.01

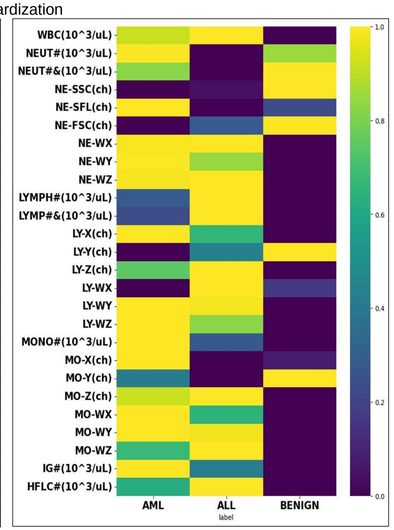


Table. 1 Training and test performance of various ML models for predicting disease categories

	Train					Test				
	TB	RF	CB	LGB	XGB	TB	RF	CB	LGB	XGB
Comp Time	4541	3307	3227	**3165**	3343	0.08	0.015	0.012	0.023	**0.016**
Accuracy (%)	86.21	82.82	95.79	98.35	98.6	84.48	81.67	86.04	87.7	**88.03**
Precision (%)	86.02	82.6	95.67	98.34	98.61	84.29	81.48	85.97	87.56	**87.91**
F1_score (%)	85.9	82.66	95.74	98.32	98.57	84.21	81.51	85.98	87.52	**87.86**
Sensitivity (%)	85.84	83.21	95.83	98.3	98.53	84.16	82.06	86.08	87.49	**87.81**
Specificity (%)	93.01	91.51	97.91	99.16	99.29	92.16	90.94	93.03	93.8	**93.97**
ROC_AUC	0.89	0.87	0.97	0.99	0.99	0.88	0.87	0.9	0.91	**0.91**

Among the compared models, XGBoost showed the best performance with an overall test sensitivity and specificity of 88% and 94% respectively and for each of the classes AML, ALL and Benign, the ROC-AUCs were 0.88, 0.87, 0.99 respectively.

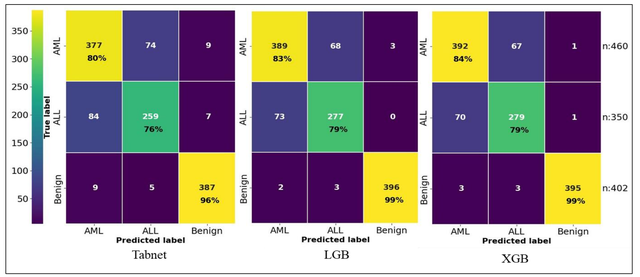


Fig .2 Confusion matrices of selected ML models along with sensitivity of each classes: AML-85%, ALL- 80%, Benign-99%.

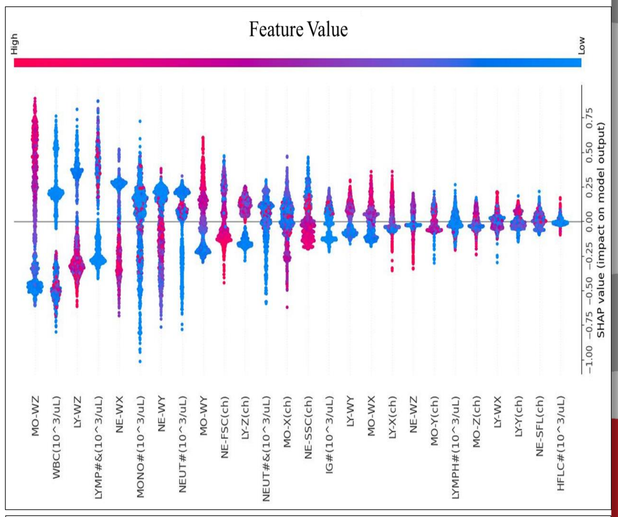


Fig 3. Feature analysis revealed that NE-WY, MO-WZ, LYMPH, NE-WZ, NEUT, and MONO# have significant impact on the disease classification.

CONCLUSIONS

- The leukocyte CPD-based ML model enhances early prediction of AL and its lineage differentiation at a pre-microscopic level.
- Integrating ML algorithms into a hematology analyzer offers a rapid, reliable, and cost-effective tool for both detecting and differentiating acute leukemia on one platform.
- This approach reduces subjectivity and is especially valuable in areas with few experts, shortening the time needed for further characterization and final diagnosis through immunophenotype and molecular studies.

REFERENCES

1. Eckardt JN, Bornhäuser M, Wendt K, Middeke JM. Application of machine learning in the management of acute myeloid leukemia: current practice and future prospects. Blood Advances. 2020 Dec 8;4(23):6077-85.
2. Salah HT, Muhsen IN, Salama ME, Owaidah T, Hashmi SK. Machine learning applications in the diagnosis of leukemia: Current trends and future directions. International journal of laboratory hematology. 2019 Dec;41(6):717-25.
3. Haider RZ, Ujjan IU, Khan NA, Urrechaga E, Shamsi TS. Beyond the in-practice CBC: the research CBC parameters-driven machine learning predictive modeling for early differentiation among leukemias. Diagnostics. 2022 Jan 7;12(1):138.
4. Virk H, Varma N, Naseem S, Bihana I, Sukhachev D. Utility of cell population data (VCS parameters) as a rapid screening tool for Acute Myeloid Leukemia (AML) in resource-constrained laboratories. Journal of Clinical Laboratory Analysis. 2019 Feb;33(2):e22679.
5. Mishra S, Chhabra G, Padhi S, Mohapatra S, Panigrahi A, Sable MN, Das PK. Usefulness of leucocyte cell population data by Sysmex XN1000 hematology analyzer in rapid identification of acute leukemia. Indian Journal of Hematology and Blood Transfusion. 2022 Jul 1:1-9.

ACKNOWLEDGEMENTS

I would like to thank Mr SUNNY KUMAR MAURYA, Mr KAILASH SINGH MEHRA and Mr PUSHKAR NEGI for technical assistance

CONTACT INFORMATION

DR BEULAH PRISCILLA MADDIRALA
HEMATO-ONCOPATHOLOGIST PDF

Email: beuspark@gmail.com
Orcid: https://orcid.org/0009-0007-8071-4343