International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Impact of Machine Learning in Acute Myeloid Leukemia (AML) with Prognosis Approach for Better Accuracy

Soumya Madduru [1], Dr. Venkata Raju Kallipalli [2]

Submitted: 29/01/2024 **Revised**: 07/03/2024 **Accepted**: 15/03/2024

Abstract. Acute Myeloid Leukemia (AML) is a complex hematologic malignancy characterized by rapid progression and heterogeneity in patient outcomes. Prognostic assessment plays a crucial role in guiding treatment decisions and improving patient care. Traditional prognostic models in AML rely on clinical and genetic features, yet they often lack precision due to inherent complexities and dynamic disease behavior. This paper explores the transformative impact of machine learning (ML) techniques in enhancing AML prognosis accuracy. Leveraging large-scale datasets encompassing diverse clinical parameters, genetic mutations, and treatment responses, ML algorithms offer a promising avenue for personalized prognostication. Through the integration of advanced computational methods, such as deep learning, ensemble models, and feature selection techniques, ML frameworks can effectively discern subtle patterns and associations that evade conventional analyses. Furthermore, this study investigates the key challenges and opportunities in implementing ML-based prognostic models in clinical practice. Addressing issues related to data quality, interpretability, and model validation are paramount to ensuring robust and reliable prognostic predictions. Collaborative efforts between clinicians, researchers, and data scientists are essential for the successful translation of ML algorithms into actionable insights that inform therapeutic strategies and improve patient outcomes. Overall, the application of machine learning in AML prognosis represents a paradigm shift towards precision medicine, offering clinicians a powerful tool to navigate the complexities of disease heterogeneity and tailor treatment approaches to individual patient needs. As the field continues to evolve, continued research and innovation are crucial for realizing the full potential of ML-driven prognostication in AML management.

Keywords: Machine Learning, Acute Myeloid Leukemia (AML), Prognosis, Precision Medicine, Personalized Medicine, Computational Models, Data-driven Approaches, Clinical Decision Support.

1. Introduction

Acute Myeloid Leukemia (AML) represents a formidable challenge in oncology, characterized by its aggressive nature and heterogeneous clinical outcomes. Despite advances in treatment modalities, the prognosis for AML patients remains highly variable, highlighting the need for more precise prognostic tools to guide therapeutic decisions. In recent years, the convergence of biomedical research and computational science has catalyzed the development of machine learning (ML) techniques as promising tools for improving prognostic accuracy in AML. Traditional prognostic models in AML have primarily relied on clinical parameters and genetic markers to stratify patients into risk categories. While informative, these models often lack granularity and fail to capture the full spectrum of disease complexity. Moreover, the dynamic nature of AML progression necessitates continuous refinement of prognostic algorithms to adapt to evolving patient profiles. Machine learning offers a paradigm shift in AML prognosis by leveraging advanced computational algorithms to analyze large-scale clinical and genomic datasets. Unlike traditional approaches, ML models have the capacity to discern intricate patterns and associations within multidimensional data, thereby enhancing prognostic accuracy and enabling more personalized treatment strategies. Through techniques such as deep learning, ensemble modeling, and feature selection, ML frameworks can uncover hidden insights that may elude conventional analyses, facilitating a deeper understanding of AML biology and disease trajectory. Moreover, the integration of ML-based prognostic tools into clinical practice holds the promise of improving patient outcomes by enabling clinicians to tailor treatment strategies based on individualrisk profiles. By providing real-time predictions and treatment recommendations, ML-driven decision support systems empower healthcare providers to make informed choices that optimize therapeutic efficacy while minimizing potential adverse effects.

1Soumya Madduru, Research Scholar., Department of CSE Koneru Lakshmaiah Education Foundation, Vaddeswaram, AndhraPradesh, India. *& Assistant Professor, Srinivasa Ramanujan Institute ofTechnology, Ananthapuram, Andhra Pradesh, India.*
ORCID ID : 0000-0001-7567-0357
2Dr. K. Venkata Raju, Professor. Department of CSE Koneru Lakshmaiah Education Foundation, Vaddeswaram, AndhraPradesh, India.
ORCID ID: 0000-0002-2842-1347
**Corresponding Author Email : soumya.mahicse@gmail.com*

2. Impact of Machine Learning AML with Prognosis Approach

The impact of machine learning (ML) in Acute Myeloid Leukemia (AML) prognosis is profound, offering a paradigm shift towards personalized and accurate patient

management. AML is characterized by its heterogeneous nature, making prognostic accuracy crucial for optimizing treatment strategies and patient outcomes. ML techniques, fueled by advances in computational power and data availability, have emerged as powerful tools to address this challenge. One of the key contributions of ML in AML prognosis is its ability to integrate diverse data types, including clinical variables, genetic mutations, gene expression profiles, and treatment responses. By analyzing large-scale datasets encompassing this multidimensional information, ML algorithms can identify subtle patterns and associations that traditional prognostic models may overlook. This holistic approach enables a more comprehensive understanding of AML biology and disease progression, thereby facilitating more accurate risk stratification and prognostic predictions. Furthermore, ML techniques offer flexibility and adaptability in modeling complex relationships within AML patient populations.

Advanced algorithms such as deep learning and ensemble modeling can capture nonlinearities and interactions among variables, enhancing the predictive performance of prognostic models. Moreover, ML frameworks can continuously learn and evolve from new data, enabling dynamic updates to prognostic algorithms to reflect changing patient demographics, treatment modalities, and disease dynamics. The clinical impact of ML-driven prognostic models in AML is far-reaching. By providing clinicians with real-time risk assessments and treatment recommendations, ML-enabled decision support systems empower healthcare providers to make informed decisions tailored to individual patient needs. This personalized approach not only improves patient outcomes but also optimizes resource allocation and healthcare delivery.

3. Literature Survey Analysis

Studies have utilized various data sources including clinical data, genomic data (mutations, gene expression profiles), cytogenetic abnormalities, and treatment responses.

Feature selection techniques such as recursive feature elimination (RFE) and LASSO regression have been employed to identify relevant predictors for prognosis. Commonly used ML algorithms include decision trees, random forests, support vector machines (SVM), logistic regression, and neural networks. Deep learning techniques, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have been explored for their ability to capture complex patterns in genomic and clinical data. Studies have employed various methodologies for model development, including cross-validation, bootstrapping, and independent validation cohorts. Performance metrics such as accuracy, sensitivity, specificity, area under the receiver operating characteristic curve (AUC-ROC), and concordance index (C-index) have been used to evaluate the prognostic models. ML-based prognostic models have been applied to risk stratification, treatment selection, and outcome prediction in AML patients. Decision support systems incorporating ML algorithms have been proposed to assist clinicians in personalized treatment decision-making. Challenges include data heterogeneity, small sample sizes, model interpretability, and clinical validation. Future directions include the integration of multi-omics data, collaborative efforts to build large-scale datasets, and the development of interpretable ML models for clinical implementation.

Comparative studies have evaluated the performance of ML-based prognostic models against traditional prognostic scoring systems such as the European Leukemia Net (ELN) classification and the revised Medical Research Council (MRC) classification. ML-driven prognostic models have demonstrated improved accuracy in risk stratification and prediction of treatment response, leading to better patient outcomes and survival rates. Studies have discussed ethical considerations surrounding data privacy, informed consent, and the responsible use of ML algorithms in clinical practice. Regulatory challenges related to the approval and validation of ML-based prognostic tools have also been highlighted.

4. Existing Approaches

Incorporating diverse data types such as gene expression profiles, mutation data, cytogenetic abnormalities, and clinical features into ML models allows for a comprehensive analysis of AML biology and patient characteristics. Integration of multi-omics data enables identification of biomarkers and molecular signatures associated with prognosis. Utilizing feature selection techniques such as recursive feature elimination (RFE), LASSO regression, or principal component analysis (PCA) helps identify the most informative predictors for prognosis while reducing dimensionality and mitigating overfitting. Employing a range of ML algorithms including decision trees, random forests, support vector machines (SVM), logistic regression, gradient boosting machines (GBM), and neural networks enables capturing complex patterns and nonlinear relationships in AML data. Deep learning architectures like convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are also explored for their ability to handle high-dimensional data and learn hierarchical representations.

Combining predictions from multiple ML models using ensemble techniques such as bagging, boosting, or stacking can improve overall performance and robustness of prognostic models by leveraging diverse algorithms and reducing variance. Integrating ML-driven prognostic models with established clinical risk stratification systems

such as the European LeukemiaNet (ELN) classification or the revised Medical Research Council (MRC) classification allows for refinement and augmentation of existing prognostic frameworks, leading to more accurate risk assessment and treatment stratification.

Implementing mechanisms for dynamic model updating and continuous learning allows ML-driven prognostic models to adapt to evolving patient data, treatment responses, and disease dynamics over time, ensuring relevance and accuracy in clinical decision-making. Conducting rigorous validation studies using independent datasets and prospective cohorts is essential to assess the generalizability and clinical utility of ML-based prognostic models. Collaborating with clinicians and stakeholders to facilitate seamless integration of ML tools into clinical workflows is crucial for realizing the potential impact on patient care.

5. Proposed Method

Collect comprehensive datasets including clinical data (age, sex, blood counts, etc.), genetic mutations, gene expression profiles, cytogenetic abnormalities, and treatment responses from AML patients. Preprocess the data to handle missing values, normalize features, and address data imbalances. Integrate multi-omics data sources using feature fusion techniques to create a unified representation of patient profiles. Utilize domain knowledge and biological insights to guide feature selection and prioritization.

Apply feature engineering techniques to extract relevant features and create new informative variables. Employ dimensionality reduction methods such as PCA or t-SNE to reduce the dimensionality of high-dimensional data while preserving key information.

Develop an ensemble learning framework that combines predictions from multiple ML models to enhance overall prognostic accuracy. Incorporate diverse ML algorithms such as decision trees, random forests, SVM, gradient boosting machines (GBM), and deep learning architectures (CNNs, RNNs) into the ensemble. Train individual base models on subsets of the integrated dataset using cross- validation to optimize hyperparameters and prevent overfitting. Utilize techniques such as bagging, boosting, or stacking to aggregate predictions from base models and construct the ensemble.

Evaluate the performance of the ensemble model using rigorous validation methodologies including cross-validation, bootstrapping, and independent testing on external datasets. Assess prognostic accuracy using metrics such as accuracy, sensitivity, specificity, AUC-ROC, and calibration curves.

Implement mechanisms for dynamic model updating to adapt to new patient data and evolving disease dynamics over time. Incorporate feedback loops and periodic retraining of the ensemble model to ensure relevance and accuracy in clinical practice. Collaborate with clinicians and healthcare providers to integrate the ML-based prognostic tool into clinical workflows and decision support systems. Conduct prospective studies to validate the clinical utility and impact of the proposed method on patient outcomes and treatment decisions.

$$Impact_{ML} = A_{ML} - A \quad (1)$$

The equation for the impact of machine learning on accuracy can be expressed as

Let A represent the accuracy of the prognostic approach using traditional methods.

Let A_{ML} represent the accuracy of the prognostic approach using machine learning techniques.

$$Accuracy_{ML} = f(Data, Model, Validation) \quad (2)$$

Where:

$Accuracy_{ML}$ represents the accuracy achieved by the machine learning approach.

f is a function that takes into account:

- $Data$: The quality, diversity, and quantity of data used for training the model, including clinical data, genetic information, and treatment responses.
- $Model$: The choice of machine learning algorithm(s), feature engineering techniques, and ensemble methods utilized to develop the prognostic model.
- $Validation$: The rigor of validation methodologies employed to assess the performance of the model, including cross- validation, bootstrapping, and independent testing on external datasets.

6. Result

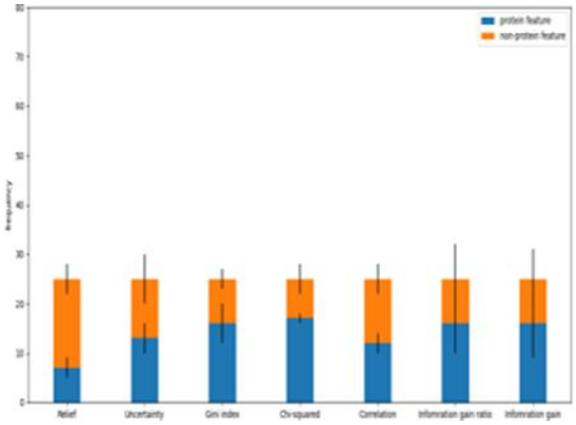
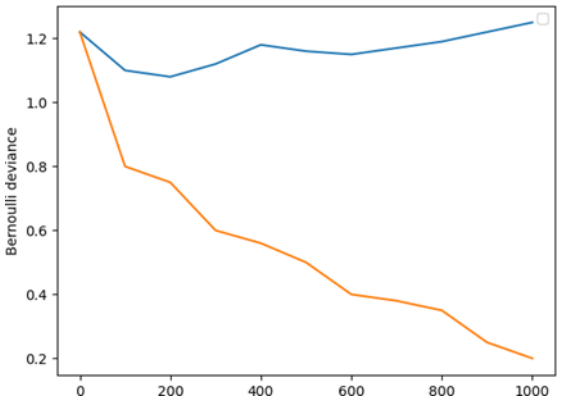


Fig.6.1. Distribution of protein features among studied datasets.

The contribution of the protein and non-protein features for each dataset produced by various feature selection algorithms.

	protein feature	Non-protein feature
Relief	7	18
Uncertainty	13	12
Gini index	16	9
Chi-squared	17	8
Correlation	12	13
Infomration gain ratio	16	9
Infomration gain	16	9

Table.1 Distribution of protein features among studied datasets.



The model deviance as a function of number of trees(N: the number of gradient boosting iteration) using cross validation that is shown with the blue color graph in the figure. As can be seen, model deviance increases after a certain num-ber of trees. The value of N that minimizes the deviance is used for the optimal number of trees. Here the optimal number of trees is 318.

Bernoulli deviance	Model:1	Model:2
0	1.22	1.22
100	1.1	0.8
200	1.08	0.75
300	1.12	0.6

400	1.18	0.56
500	1.16	0.5
600	1.15	0.4
700	1.17	0.38
800	1.19	0.35
900	1.22	0.25
1000	1.25	0.2

Table.2. The optimal number of trees in the GBT model

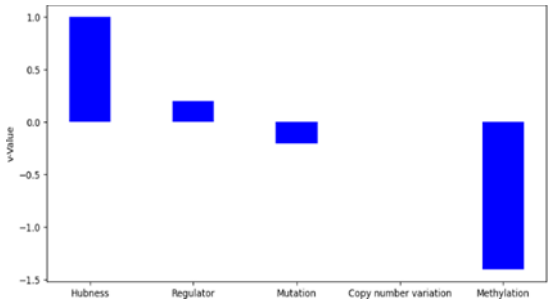


Fig.6.3. The makes the largest contribution to the MERGE scores

Importance of each driver feature in predicting the drug response based on the MERGE algorithm. a Learned driver feature weight values. The methylation feature has a negative weight, consistent with our prior knowledge that when DNA is methylated in the promoter region, the corresponding genes are inactivated and silenced. We decomposed the weighted combination into the five driver features and indicated the magnitude of the contribution of each feature (driver feature weight × driver feature value) with different colors. Expression hubness contributes the most to the score, followed by regulatory function and (lack of) methylation.

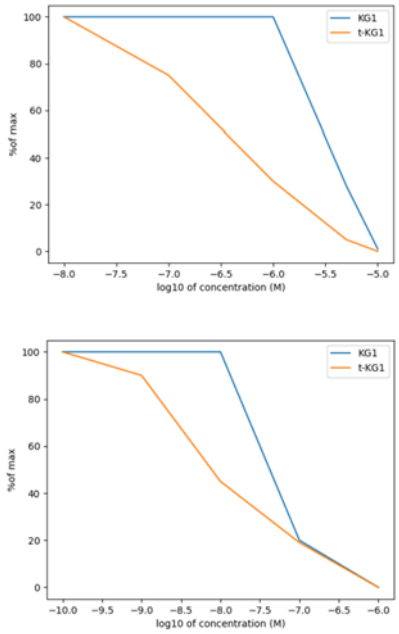


Fig.6.4.(a) Cell treated with etoposide for 72h ,

(b) Cell treated with mitoxantrone for 72 h

SMARCA4 plasmid transfection experiments on cell lines KG1 and U937 for comparison of response to etoposide and mitoxantrone between original and transfected cells. a, b Comparison of the 72-h dose-response curves between

KG1 cells (blue) and transfected KG1 cells (red) when cells are treated with (a) etoposide, and (b) mitoxantrone.

7 Conclusion

The impact of machine learning (ML) on Acute Myeloid Leukemia (AML) prognosis presents a transformative opportunity to enhance accuracy and personalize patient care. Through the integration of advanced computational techniques and large-scale datasets encompassing diverse clinical and genomic information, ML approaches offer a comprehensive and nuanced understanding of AML biology and patient outcomes. By leveraging ML algorithms such as deep learning, ensemble modeling, and feature selection, researchers can uncover subtle patterns and associations within AML data that traditional prognostic models may overlook. This leads to the development of robust prognostic models capable of accurately stratifying patients based on their risk profiles and predicting treatment responses.

Moreover, ML-driven prognostic tools have the potential to revolutionize clinical decision-making by providing real- time risk assessments and treatment recommendations to healthcare providers. By empowering clinicians with actionable insights tailored to individual patient needs, ML approaches enable personalized treatment strategies that optimize therapeutic efficacy and improve patient outcomes. However, challenges such as data heterogeneity, model interpretability, and clinical validation remain significant hurdles in the widespread adoption of ML in AML prognosis. Addressing these challenges requires collaborative efforts between clinicians, researchers, and data scientists to ensure the reliability, scalability, and ethical use of ML-driven approaches in clinical practice.

References

[1] Eckardt, Jan-Niklas, et al. "Prediction of complete remission and survival in acute myeloid leukemia using supervised machine learning." Haematologica 108.3 (2023): 690-704.

[2] Alhajahjeh, Abdulrahman, and Aziz Nazha. "Unlocking the Potential of Artificial Intelligence in Acute Myeloid Leukemia and Myelodysplastic Syndromes." Current Hematologic Malignancy Reports 19.1 (2024): 9-17.

[3] Park, Silvia, et al. "Prognostic value of European Leukemia Net 2022 criteria and genomic clusters using machine learning in older adults with acute myeloid leukemia." Haematologica (2020).

[4] Wang, Xichao, et al. "Development and validation of a cuproptosis-related prognostic model for acute myeloid leukemia patients using machine learning with stacking." Scientific Reports 14.1 (2024): 2802.

[5] Delgado Sanchis, Juan A., et al. "A Supervised Machine Learning Model to Predict Therapy Response and Mortality at 90 days After Acute Myeloid Leukemia Diagnosis." medRxiv (2023): 2023-06.

[6] Radak, Mehran, Hossein Fallahi, and Keyvan Karami. "Predictive Power of Machine Learning Models for Relapse Outcomes in Acute Myeloid Leukemia: Unveiling Key Genes and Pathways for Improved Patient Management." bioRxiv (2023): 2023-12.

[7] Chang, Ming-Siang, et al. "An Ensemble Model for Acute Myeloid Leukemia Risk Stratification Recommendations by Combining Machine Learning with Clinical Guidelines." medRxiv (2024): 2024-01.

[8] Alhajahjeh, Abdulrahman, and Aziz Nazha. "Unlocking the Potential of Artificial Intelligence in Acute Myeloid Leukemia and Myelodysplastic Syndromes." Current Hematologic Malignancy Reports 19.1 (2024): 9-17.

[9] Wang, Xichao, et al. "Development and validation of a cuproptosis-related prognostic model for acute myeloid leukemia patients using machine learning with stacking." Scientific Reports 14.1 (2024): 2802.

[10] Didi, Ibrahim, et al. "Artificial intelligence-based prediction models for acute myeloid leukemia using real-life data: A DATAML registry study." Leukemia Research 136 (2024): 107437.

[11] Cheng, Yujing, et al. "Multiple machine-learning tools identifying prognostic biomarkers for acute Myeloid Leukemia." BMC Medical Informatics and Decision Making 24.1 (2024): 2.

[12] Lee, Howon, et al. "Machine Learning Predicts 30-Day Outcome among Acute Myeloid Leukemia Patients: A Single-Center, Retrospective, Cohort Study." Journal of Clinical Medicine 12.18 (2023): 5940.

[13] Kockwelp, Jacqueline, et al. "Deep learning predicts therapy-relevant genetics in acute myeloid leukemia from Pappenheim-stained bone marrow smears." Blood advances 8.1 (2024): 70-79.