Original Article

Check for updates

Available online at
www.heca-analitika.com/ijds

Infolitika Journal of Data Science

Vol. 4, No. 1, 2026

Assessing LightGBM Performance in Automated Leukemia Cell Classification

Rara Syifa Qaisa [1], Hayatun Maghfirah [1], Suryadi Suryadi [1], Noviana Husdayanti [2], and Rivansyah Suhendra [1,*]

[1] Department of Information Technology, Faculty of Engineering, Universitas Teuku Umar, Aceh Barat 23681, Indonesia: rarasyifaqaisa@gmail.com (R.S.Q); hayatunmaghfirah@utu.ac.id (H.M.); suryadi@utu.ac.id (S.S.); rivansyahsuhendra@utu.ac.id (R.S.);
[2] Teuku Umar Hospital, Aceh Jaya 23654, Indonesia: novianahdy@acehjayakab.go.id (N.H.)

* Correspondence: rivansyahsuhendra@utu.ac.id

Article History

Received 12 February 2026
Revised 27 April 2026
Accepted XX 8 May 2026
Available Online 16 May 2026

Keywords:
LightGBM
Leukemia
Image classification
Machine learning
Blood cell subtypes

Abstract

Leukemia is a type of blood cancer that requires fast and accurate diagnosis for effective treatment. Manual identification of leukemia blood cell subtypes is often challenging, time-consuming, and prone to observer variability, making automated image-based classification essential. This study evaluates the performance of the Light Gradient-Boosting Machine (LightGBM) as a computationally efficient and interpretable alternative to deep learning models for classifying leukemia subtypes. The dataset includes 3,000 microscopic images representing five classes: acute lymphocytic, acute myelogenous, chronic lymphocytic, chronic myelogenous, and healthy blood cells. Images were preprocessed using bilinear interpolation to balance quality and efficiency, and 90 statistical features were extracted across 13 distinct color spaces. The model was trained on an 80% subset and validated on a 20% hold-out set after hyperparameter optimization. LightGBM achieved robust performance with an accuracy of 93.3%, precision of 99.1%, recall of 94.9%, and an F-measure of 96.8%. Feature importance analysis revealed that texture variance in the YIQ color space (STD_YIQ_I) was the most critical predictor, highlighting the biological relevance of chromatin texture in classification. These results indicate that LightGBM is an effective, lightweight, and reliable approach for leukemia subtype classification, holding strong potential for implementation in resource-constrained automated diagnostic systems.

1. Introduction

Leukemia is a malignant disorder of the blood-forming tissues, characterized by the uncontrolled production of abnormal white blood cells [1]. It represents a significant global health burden, with approximately 474,519 new cases and 311,594 deaths reported annually [2]. Effective treatment relies heavily on the rapid and accurate identification of leukemia subtypes, as treatment protocols differ significantly between acute and chronic forms. However, the current gold standard for diagnosis, manual microscopic examination of peripheral blood smears, is labor-intensive, time-consuming, and prone to inter-observer variability due to the subjective nature of human interpretation [3]. These challenges have driven a surge in demand for automated, computer-aided diagnostic (CAD) systems to assist pathologists in making precise decisions [4, 5].

In recent years, various computational approaches have been proposed to automate leukemia classification. Early studies focused on traditional machine learning

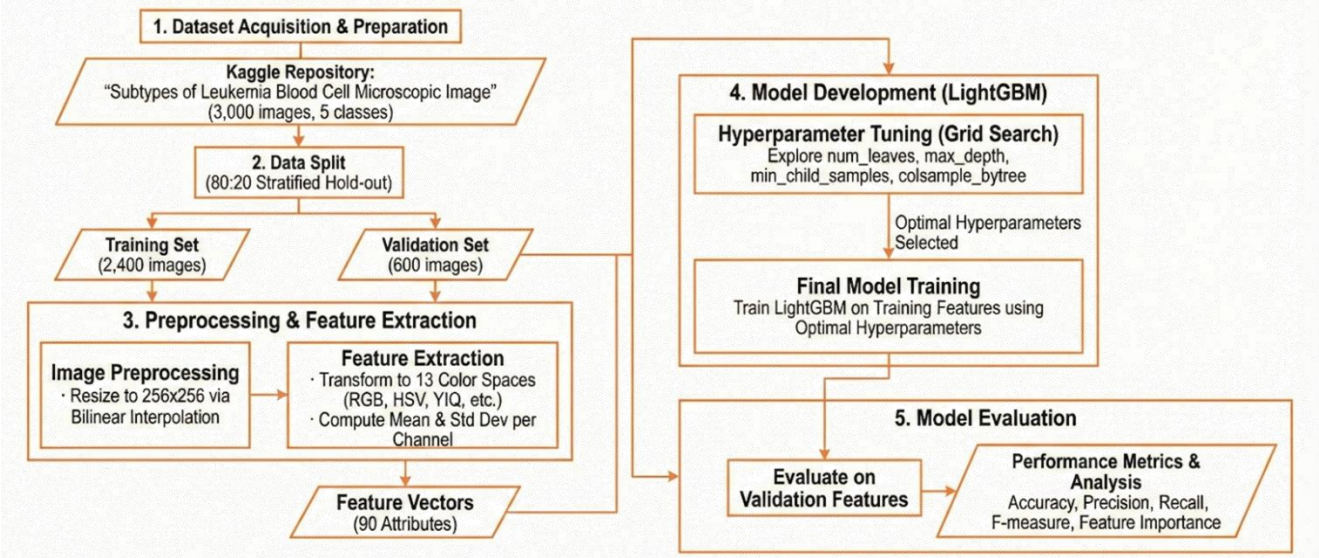

Figure 1. The overall research workflow.

algorithms such as Support Vector Machines (SVM) and K-Nearest Neighbors (k-NN) [6–9]. For instance, a study by Mahdi demonstrated the utility of SVM classifiers for leukemia detection, while other studies explored the use of Random Forest for morphological classification [1]. While these methods established the feasibility of automated diagnosis, they often struggle with the high-dimensional feature spaces inherent in high-resolution medical images. They may yield suboptimal accuracy when class boundaries are complex [10]. Conversely, Deep Learning (DL) models, particularly Convolutional Neural Networks (CNNs), have recently set new benchmarks in medical image analysis [11]. However, DL models are computationally expensive, require massive annotated datasets to prevent overfitting, and operate as "black boxes," offering little insight into which visual features drive the classification, a critical limitation in clinical diagnostics where interpretability is paramount [12].

This creates a specific research gap: the need for a diagnostic approach that bridges the efficiency and interpretability of traditional machine learning with the high performance required for clinical use. Light Gradient-Boosting Machine (LightGBM) has emerged as a powerful candidate to fill this gap. Unlike traditional boosting algorithms, LightGBM uses a leaf-wise tree growth strategy and histogram-based methods, enabling it to handle large-scale data and high-dimensional feature sets with superior training speed and lower memory usage [13]. Although LightGBM has shown success in other medical domains, such as diabetes prediction and drug discovery [14–18], its application specifically for multiclass leukemia subtype classification using color-space feature engineering has not been extensively investigated.

This study addresses this gap by evaluating the performance of LightGBM for classifying five distinct blood cell classes (four leukemia subtypes and healthy cells). Unlike "black box" approaches, this study utilizes explicitly extracted statistical features across multiple color spaces to maintain interpretability. The specific objectives of this research are to develop a robust feature extraction pipeline that captures color and texture variations across 13 distinct color spaces, to optimize the LightGBM algorithm through hyperparameter tuning to maximize classification accuracy for leukemia subtypes, and to comprehensively evaluate the model using standard metrics (accuracy, precision, recall, F-Measure) and analyze Feature Importance to identify the most critical visual attributes for differentiation. By achieving these objectives, this research aims to demonstrate that LightGBM can serve as a highly accurate, computationally efficient, and interpretable tool for automated leukemia diagnosis.

2. Materials and Methods

This study evaluates the performance of the LightGBM in classifying leukemia blood cell subtypes. The experimental framework follows a systematic pipeline consisting of six main stages: dataset acquisition, image preprocessing, feature extraction, model training, hyperparameter tuning, and performance evaluation. To provide a clear overview of the research process and the flow of data between stages, the comprehensive methodology is visually illustrated in Figure 1. Each step was rigorously designed to ensure the reliability and reproducibility of the model in differentiating leukemia subtypes from microscopic images.

Table 1. Hyperparameter search space and candidate values evaluated during model optimization.

Parameter	Description	Tested Range
num_leaves	Controls the complexity of the tree model	[20, 31, 40, 50, 60]
max_depth	Limits the maximum depth of the tree	[-1 (unlimited), 10, 20, 30]
min_child_samples	Minimum number of data needed in a leaf	[10, 15, 17, 20, 30, 50]
colsample_bytree	Fraction of features (columns) used per tree	[0.7, 0.8, 0.9, 1.0]

2.1. Dataset

The dataset used in this study was obtained from the publicly available Kaggle repository entitled *"Subtypes of Leukemia Blood Cell Microscopic Image."* It consists of 3,000 microscopic images representing five different classes: acute lymphocytic, acute myelogenous, chronic lymphocytic, chronic myelogenous, and healthy blood cells. Each class contains 600 images, resulting in a balanced dataset. For model evaluation, the dataset was partitioned using an 80:20 stratified hold-out scheme, with 2,400 images allocated for training and 600 for validation (testing). This stratification ensures a balanced class distribution across both data subsets.

2.2. Preprocessing

Image preprocessing was performed to standardize the input data and improve the quality of features extracted later. The process involved resizing all images to 256 × 256 pixels using bilinear interpolation. This technique computes pixel intensities as a weighted average of neighboring pixels. Bilinear interpolation was selected as it offers an optimal balance between computational efficiency and output image quality. This method produces significantly smoother images compared to nearest-neighbor interpolation, which tends to introduce jagged artifacts, while remaining computationally faster than more complex techniques such as bicubic interpolation. This balance is crucial for maintaining the integrity of the extracted statistical features without imposing excessive computational overhead. The resizing step was crucial to ensure uniformity across all samples, reduce memory usage, and accelerate computation during training.

2.3. Feature Extraction

Feature extraction was performed to obtain representative numerical values capturing the visual characteristics of leukemia blood cell images. Each image was analyzed not only in its original RGB format but also transformed into 12 additional color spaces to capture diverse chromatic and intensity information. The 13 color spaces utilized in this study are: RGB, HSV, HSL, L*a*b*, L*u*v*, XYZ, YCbCr, YIQ, YPbPr, YUV, HED, CIE Lch, and Opponent color space. From these transformations, statistical features (mean and standard deviation) were computed for each channel, yielding a robust feature vector with 90 attributes. For instance, the YIQ and Opponent spaces separate luminance from chrominance, which is critical for distinguishing the nuclear texture of blast cells from the cytoplasm.

2.4. Model Training

The extracted features were used to train the LightGBM, an ensemble-based learning algorithm known for its efficiency and high predictive performance [19, 20]. The dataset was divided into training and validation subsets, with 80% of the data used for training and 20% for validation. The LightGBM model learns patterns from the extracted features by building decision trees sequentially, where each new tree attempts to correct the errors of the previous ones. This approach allows LightGBM to achieve high accuracy with relatively low computational cost, making it suitable for handling large image-based datasets.

2.5. Hyperparameter Tuning

Hyperparameter tuning was conducted to maximize the model's predictive accuracy while preventing overfitting or underfitting [21]. This process utilized a Grid Search strategy to systematically explore various combinations of key parameters. The adjusted hyperparameters included min_child_samples, which determines the minimum number of samples required to create a leaf node; colsample_bytree, which defines the fraction of features used per decision tree; max_depth, which limits the maximum depth of each tree; and num_leaves, which controls model complexity by defining the maximum number of leaves per tree [22]. The specific search ranges tested during the optimization process and the resulting optimal values selected for the final model are summarized in Table 1.

2.6. Model Evaluation

The performance of the LightGBM model was evaluated using several standard metrics: accuracy, precision, recall, F-measure, and feature importance [23, 24]. Accuracy represents the overall proportion of correctly classified instances. Precision measures how many of the predicted positive samples are actually correct, while recall evaluates how well the model identifies all true positive samples [25]. The F-measure, as the harmonic

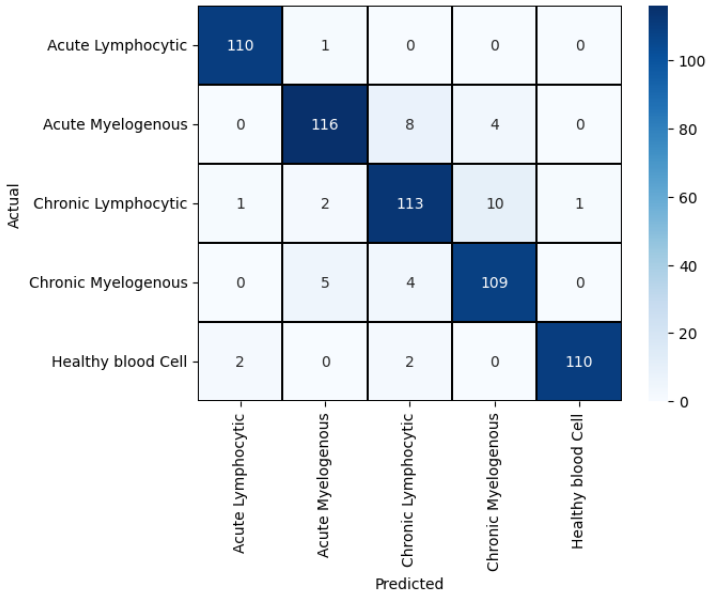

Figure 2. Confusion matrix of the LightGBM model.

mean of precision and recall, provides a balanced assessment of both metrics. Additionally, the feature importance analysis was conducted to identify which extracted features had the greatest influence on the classification outcome, offering interpretability and insight into the model's decision-making process [26].

2.7. Implementation Environment

The experiment was implemented using Python 3.10. Feature extraction was performed using the scikit-image library (v0.19), and the classification model was built using the LightGBM library (v3.3.5). The training process was conducted on a personal computer equipped with a Ryzen 7 4800 CPU and 16 GB of RAM. To ensure reproducibility, a random seed of 42 was set for all stochastic processes, including data splitting and model initialization.

3. Results and Discussion

This section presents the experimental results from classifying leukemia blood cell subtypes using the LightGBM algorithm. The analysis includes the outcomes of preprocessing, feature extraction, hyperparameter tuning, model evaluation, and feature importance interpretation.

The dataset used in this study consisted of 3,000 microscopic images representing five categories of blood cells. As the original images varied in dimension, a resizing operation was performed to standardize all images to 256 × 256 pixels using bilinear interpolation. This preprocessing step improved image uniformity and computational efficiency while preserving sufficient detail for subsequent analysis. The resized images demonstrated adequate clarity and consistency for reliable feature extraction and model training.

Following the preprocessing stage, feature extraction was performed to obtain numerical representations of each image. Ninety statistical features were extracted across 13 color spaces from the RGB channels, including the mean and standard deviation of pixel intensities. These features were found to capture meaningful variations in texture and color, which are essential for differentiating between the five leukemia subtypes and healthy blood cells. The comprehensive set of features contributed to improved model generalization and robustness during training.

To optimize model performance, several combinations of hyperparameters were tested. The parameter tuning process focused on four main hyperparameters—*min_child_samples*, *colsample_bytree*, *max_depth*, and *num_leaves*. The optimal configuration was obtained with *min_child_samples = 17*, *colsample_bytree = 0.9*, *max_depth = -1*, and *num_leaves = 31*. This configuration yielded the best trade-off between accuracy and generalization, achieving a training accuracy of 91.8%. The analysis showed that lower values of *min_child_samples* improved model flexibility, but excessive reductions led to overfitting, while moderate colsample_bytree values helped reduce feature redundancy and enhanced generalization.

The Confusion Matrix presented in Figure 2 provides deeper insight into the misclassifications. The model demonstrated excellent performance in identifying *Acute Lymphocytic* cells (110 correct, only 1 miss) and *Healthy* cells (110 correct). However, greater confusion was observed between the chronic subtypes. Specifically,

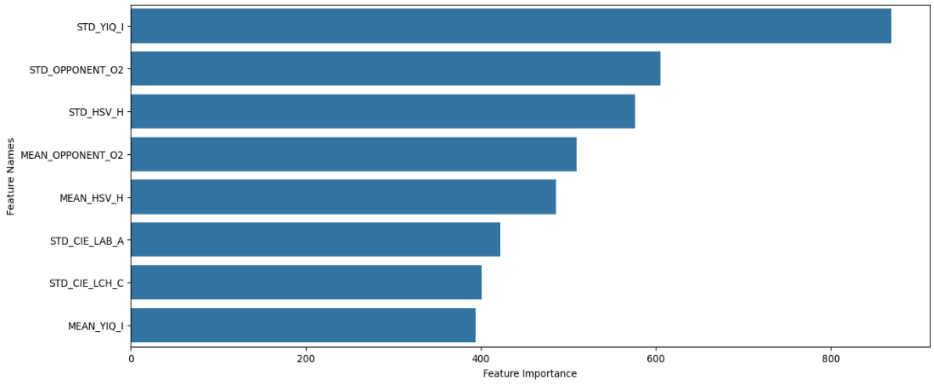

Figure 3. Feature importance of the LightGBM model.

Chronic Lymphocytic cells were occasionally misclassified as *Chronic Myelogenous* (10 instances) and *Acute Myelogenous* (8 instances). This pattern suggests that while the model excels at distinguishing cell types, it faces challenges with subtle morphological similarities, such as cytoplasm-to-nucleus ratios shared among chronic leukemia subtypes.

The trained LightGBM model was evaluated using five standard metrics: accuracy, precision, recall, F-measure, and feature importance. The results demonstrated that LightGBM achieved excellent performance in classifying leukemia blood cell subtypes, with an accuracy of 93.3%, precision of 99.1%, recall of 94.9%, and F-measure of 96.8%. The confusion matrix analysis indicated that the model effectively distinguished among the five classes, with minimal misclassification between morphologically similar subtypes such as acute and chronic leukemia cells (Fig. 2). These findings suggest that LightGBM can effectively capture complex patterns in feature distributions, making it suitable for multi-class medical image classification tasks.

To interpret the model's decision-making process, we analyzed the feature importance using the "split" criterion, which measures how frequently a feature is used to split data across all decision trees. Figure 3 shows the ranking of the 8 most influential features identified in this analysis. The results revealed that the std_yiq_i feature had the highest importance score, indicating its substantial contribution to distinguishing between leukemia subtypes. This feature represents the standard deviation in the intensity component of the YIQ color space, which correlates with variations in cell brightness and internal texture patterns. The prominence of this feature highlights the importance of color intensity distribution in differentiating microscopic images of leukemia cells.

The high performance achieved by LightGBM in this study is consistent with findings in recent literature, where gradient boosting algorithms are frequently reported to outperform traditional machine learning methods such as MLP, KNN, SVM, and Random Forest, particularly when handling high-dimensional feature sets [27–29]. This advantage is primarily attributed to LightGBM's ability to perform internal feature selection and handle complex nonlinear relationships between features more effectively. Its ability to handle large feature sets and perform internal feature selection contributes to its strong performance in this multi-class classification task.

The findings are consistent with previous studies that applied LightGBM in other medical imaging contexts, such as diabetes diagnosis and class imbalance handling, where the model also achieved high predictive accuracy. In this research, the high precision and recall values indicate that LightGBM not only minimizes false classifications but also effectively captures a wide range of positive cases. It is important to note that the use of a single public dataset (Kaggle) introduces potential bias, as variability in staining techniques, microscope lighting conditions, and equipment artifacts from different laboratories may not be fully represented. Consequently, a major challenge for the clinical translation of this model is the need for extensive validation using multi-centric datasets (from various hospitals) to test model generalizability. Furthermore, integrating this model into digital pathology workflows requires further consideration regarding medical device regulatory standards and the development of pathologist-friendly user interfaces.

However, the model's performance is influenced by the quality and diversity of the dataset. Since this study relied on a single dataset from Kaggle, additional validation with different or larger datasets would strengthen the generalizability of the results. Additionally, incorporating texture-based or deep learning–derived features could potentially enhance classification accuracy in future studies.

4. Conclusions

This study successfully evaluated the LightGBM for classifying leukemia blood cell subtypes, achieving a robust accuracy of 93.3% and precision of 99.1% on the validation set. Beyond numerical metrics, the research demonstrates that a lightweight, feature-based ensemble model can rival complex deep learning architectures in performance while offering superior computational efficiency and interpretability, particularly by identifying chromatin texture (STD_YIQ_I) and color contrast as critical predictors. However, these findings are constrained by the use of a single public dataset, which limits generalizability across clinical settings, and by the observation that testing accuracy slightly exceeds training accuracy due to aggressive regularization, which warrants caution regarding potential dataset-specific artifacts. Consequently, while the current results establish a promising baseline, the system is not yet ready for standalone deployment; future work must focus on validation with multi-centric datasets and the integration of morphological features to resolve subtle misclassifications among chronic subtypes, thereby moving toward a reliable human-in-the-loop diagnostic aid.

Author Contributions: Conceptualization, R.S. and N.H.; methodology, R.S. and R.S.Q.; software, R.S.Q.; validation, N.H., and R.S.; formal analysis, R.S.; investigation, R.S.Q; resources, N.H. and H.M.; data curation, R.S.Q.; writing—original draft preparation, R.S., N.H. and R.S.Q.; writing—review and editing, R.S., R.S.Q., S.S., and H.M.; visualization, S.S., and R.S.Q. All authors have read and agreed to the published version of the manuscript.

Funding: This study does not receive external funding.

Ethical Clearance: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset used in this study is publicly available online and can be accessed from the Kaggle repository at https://www.kaggle.com/datasets/sheikhlubna/subtypes-of-leukemia-blood-cell-microscopic-image.

Acknowledgments: The authors would like to express their sincere gratitude to all individuals and institutions who contributed to this research. Special thanks are extended to Universitas Teuku Umar and the Faculty of Engineering for their continuous support, as well as to all collaborators who provided valuable insights and assistance throughout the study.

Conflicts of Interest: All the authors declare no conflicts of interest.

References

1. Mahdi, G. J. M. (2020). A Modified Support Vector Machine Classifiers Using Stochastic Gradient Descent with Application to Leukemia Cancer Type Dataset, *Baghdad Science Journal*, Vol. 17, No. 4, 1255–1266. doi:10.21123/bsj.2020.17.4.1255.
2. Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries, *CA: A Cancer Journal for Clinicians*, Vol. 71, No. 3, 209–249. doi:10.3322/caac.21660.
3. Shah, A., Naqvi, S. S., Naveed, K., Salem, N., Khan, M. A. U., and Alimgeer, K. S. (2021). Automated Diagnosis of Leukemia: A Comprehensive Review, *IEEE Access*, IEEE, 132097–132124. doi:10.1109/ACCESS.2021.3114059.
4. Quinn, L., Tryposkiadis, K., Deeks, J., De Vet, H. C. W., Mallett, S., Mokkink, L. B., Takwoingi, Y., Taylor-Phillips, S., and Sitch, A. (2023). Interobserver Variability Studies in Diagnostic Imaging: A Methodological Systematic Review, *British Journal of Radiology*, The British Institute of Radiology., 20220972. doi:10.1259/bjr.20220972.
5. Khalifa, M., Albadawy, M., and Iqbal, U. (2024). Advancing Clinical Decision Support: The Role of Artificial Intelligence across Six Domains, *Computer Methods and Programs in Biomedicine Update*, Elsevier, 100142. doi:10.1016/j.cmpbup.2024.100142.
6. Laosai, J., and Chamnongthai, K. (2018). Classification of Acute Leukemia Using Medical-Knowledge-Based Morphology and CD Marker, *Biomedical Signal Processing and Control*, Vol. 44, 127–137. doi:10.1016/j.bspc.2018.01.020.
7. Ramya, B., and Rani, U. (2020). Acute Lymphoblastic Leukemia Detection and Classification of Its Subtypes Using Pretrained Deep Convolutional Neural Networks, *International Journal of Scientific and Technology Research*, Vol. 9, No. 2, 6177–6180.
8. Zolfaghari, M., and Sajedi, H. (2022). A Survey on Automated Detection and Classification of Acute Leukemia and WBCs in Microscopic Blood Cells, *Multimedia Tools and Applications*, Vol. 81, No. 5, 6723–6753. doi:10.1007/s11042-022-12108-7.
9. Vineela, M., Reddy, G. D. S., Karthik, G., Muthukumaran, N., and Al Deen, S. H. (2024). Classification of Leukemia White Blood Cell Cancer, *7th International Conference on Inventive Computation Technologies, ICICT 2024*, IEEE, 1233–1236. doi:10.1109/ICICT60155.2024.10544978.
10. Wani, A. A. (2025). Comprehensive Review of Dimensionality Reduction Algorithms: Challenges, Limitations, and Innovative Solutions, *PeerJ Computer Science*, Vol. 11, e3025. doi:10.7717/peerj-cs.3025.
11. Abdou, M. A. (2022). Literature Review: Efficient Deep Neural Networks Techniques for Medical Image Analysis, *Neural Computing and Applications*, Springer, 5791–5812. doi:10.1007/s00521-022-06960-9.
12. Rane, N., Choudhary, S. P., and Rane, J. (2024). Ensemble Deep Learning and Machine Learning: Applications, Opportunities, Challenges, and Future Directions, *Studies in Medical and Health Sciences*, Vol. 1, No. 2, 18–41. doi:10.48185/smhs.v1i2.1225.
13. Hajihosseinlou, M., Maghsoudi, A., and Ghezelbash, R. (2023). A Novel Scheme for Mapping of MVT-Type Pb–Zn Prospectivity: LightGBM, a Highly Efficient Gradient Boosting Decision Tree Machine Learning Algorithm, *Natural Resources Research*, Vol. 32, No. 6, 2417–2438. doi:10.1007/s11053-023-10249-6.
14. Goswami, B., Bhuyan, M. K., Alfarhood, S., and Safran, M. (2024). Classification of Oral Cancer Into Pre-Cancerous Stages From White Light Images Using LightGBM Algorithm, *IEEE Access*, Vol. 12, 31626–31639. doi:10.1109/ACCESS.2024.3370157.
15. Sai, M. J., Chettri, P., Panigrahi, R., Garg, A., Bhoi, A. K., and Barsocchi, P. (2023). An Ensemble of Light Gradient Boosting Machine and Adaptive Boosting for Prediction of Type-2 Diabetes, *International Journal of Computational Intelligence Systems*, Vol. 16, No. 1, 14. doi:10.1007/s44196-023-00184-y.
16. Kanber, B. M., Smadi, A. Al, Noaman, N. F., Liu, B., Gou, S., and Alsmadi, M. K. (2024). LightGBM: A Leading Force in Breast Cancer Diagnosis Through Machine Learning and Image

Processing, *IEEE Access*, Vol. 12, 39811–39832. doi:10.1109/ACCESS.2024.3375755.

17. Qiuqian, W., GaoMin, KeZhu, Z., and Chenchen. (2025). A Light Gradient Boosting Machine Learning-Based Approach for Predicting Clinical Data Breast Cancer, *Multiscale and Multidisciplinary Modeling, Experiments and Design*, Vol. 8, No. 1, 75. doi:10.1007/s41939-024-00662-6.
18. Ramalingam, K., Yadalam, P. K., Ramani, P., Krishna, M., Hafedh, S., Badnjević, A., Cervino, G., and Minervini, G. (2024). Light Gradient Boosting-Based Prediction of Quality of Life among Oral Cancer-Treated Patients, *BMC Oral Health*, Vol. 24, No. 1, 349. doi:10.1186/s12903-024-04050-x.
19. Noviandy, T. R., Maulana, A., Idroes, G. M., Maulydia, N. B., Patwekar, M., Suhendra, R., and Idroes, R. (2023). Integrating Genetic Algorithm and LightGBM for QSAR Modeling of Acetylcholinesterase Inhibitors in Alzheimer's Disease Drug Discovery, *Malacca Pharmaceutics*, Vol. 1, No. 2, 48–54. doi:10.60084/mp.v1i2.60.
20. Noviandy, T. R., Idroes, G. M., Maulana, A., Afidh, R. P. F., and Idroes, R. (2024). Optimizing Hepatitis C Virus Inhibitor Identification with LightGBM and Tree-Structured Parzen Estimator Sampling, *Engineering, Technology & Applied Science Research*, Vol. 14, No. 6, 18810–18817. doi:10.48084/etasr.8947.
21. Noviandy, T. R., Idroes, G. M., Patwekar, M., and Idroes, R. (2025). Fine-Tuning ChemBERTa for Predicting Activity of AXL Kinase Inhibitors in Oncogenic Target Modeling, *Grimsa Journal of Science Engineering and Technology*, Vol. 3, No. 2, 73–84. doi:10.61975/gjset.v3i2.98.
22. Suhendra, R., Husdayanti, N., Suryadi, S., Juliwardi, I., Sanusi, S., Ridho, A., Ardiansyah, M., Murhaban, M., and Ikhsan, I. (2023). Cardiovascular Disease Prediction Using Gradient Boosting Classifier, *Infolitika Journal of Data Science*, Vol. 1, No. 2, 56–62. doi:10.60084/ijds.v1i2.131.
23. Maulana, A., Noviandy, T. R., Suhendra, R., Earlia, N., Prakoeswa, C. R. S., Kairupan, T. S., Idroes, G. M., Subianto, M., and Idroes, R. (2024). Psoriasis Severity Assessment: Optimizing Diagnostic Models with Deep Learning, *Narra J*, Vol. 4, No. 3, e1512. doi:10.52225/narra.v4i3.1512.
24. Noviandy, T. R., and Idroes, R. (2025). Interpretable Machine Learning QSAR Models for Classification and Screening of VEGFR-2 Inhibitors in Anticancer Drug Discovery, *Malacca Pharmaceutics*, Vol. 3, No. 2, 58–66. doi:10.60084/mp.v3i2.339.
25. Sasmita, N. R., Ramadeska, S., Kesuma, Z. M., Noviandy, T. R., Maulana, A., Khairul, M., and Suhendra, R. (2024). Decision Tree versus K-NN: A Performance Comparison for Air Quality Classification in Indonesia, *Infolitika Journal of Data Science*, Vol. 2, No. 1, 9–16. doi:10.60084/ijds.v2i1.179.
26. Han, J., Kamber, M., and Pei, J. (2011). Data Mining Concepts and Techniques Third Edition, *The Morgan Kaufmann Series in Data Management Systems*, 83–124.
27. Bagasjvara, R. G. (2017). *Klasifikasi Jenis Sel Darah Putih Dan Sel Acute Lymphoblastic Leukemia Menggunakan Pengolahan Citra Digital Dengan Metode Multilayer Perceptron*, Universitas Gadjah Mada.
28. Wirawan, R. I. (2020). Klasifikasi Sel Darah Putih Limfoblas Dan Nonlimfoblas Pada Pasien Acute Lymphoblastic Leukemia Tipe L1 Dengan Menggunakan Metode Random Forest.
29. Alamsyah, F., Dzikrullah Suratin, M., and Hamid, M. (2022). Analisis Perbandingan Metode Support Vector Machine (SVM) Dan Decision Tree Pada Klasifikasi Penyakit Acute Lymphoblastic Leukimia (ALL), *Jurnal PRODUKTIF*, Vol. 6, No. 2, 581–587.