Hybrid AI Classification Techniques for Multi-Class Leukemia Sub-Types Classification with Bone Marrow Cells

Md Sakil Hossain*, Minhaz Ul Karim Asif*, Razaul Haque Rasham*, D. M. Saaduzzaman*, Md. Hasan Maruf *, Muhammad Abul Hasan†, and ASM Shihavuddin§

*Department of Electrical and Electronic Engineering, Green University of Bangladesh, Dhaka, Bangladesh
†Department of Artificial Intelligence and Data Science, Green University of Bangladesh, Dhaka, Bangladesh
§Department of Electrical and Electronic Engineering, Presidency University, Dhaka, Bangladesh
Email: mdsoyaibhossain45@gmail.com, minhaz.asif.73@gmail.com, rasham2670@gmail.com, saaduzzaman@eee.green.edu.bd, maruf@eee.green.edu.bd, muhammad.hasan@cse.green.edu.bd, shihavuddin@pu.edu.bd

Abstract— **Leukemia is a complex and life-threatening blood cancer that occurs due to abnormal WBC blood cells coming into the human body's blood. Among the four primary subtypes of leukemia, each consists of different components of abnormal WBC blood cells. Accurate classification of each subtype, which includes acute lymphoblastic leukemia (ALL), acute myeloid leukemia (AML), chronic lymphocytic leukemia (CLL), and chronic myeloid leukemia (CML), is crucial for effective treatment and improved patient outcomes. Traditionally, the classification and diagnosis process involve manually examining bone marrow cells under microscopes. It is time-consuming, pathologist-intensive, and prone to human error. This research offers several hybrid AI models that combine CNNs for feature extraction with traditional ML classifiers, including SVM, RF, XGBoost, and GBM, to classify leukemia subtypes using bone marrow cell images. These models aim to automate the diagnosis process, enhance accuracy, reduce the burden on doctors and pathologists, and make the process faster and more efficient. The proposed models are evaluated using confusion metrics such as accuracy, precision, recall, and F1-score, demonstrating superior performance over standalone models. This study contributes to medical diagnostics by offering a novel, efficient approach to leukemia subtype classification.**

Keywords—Leukemia, CNN, SVM, XGBoost, GBM, AI, Machine Learning, DL, Hybrid AI Models, Bone Marrow Cells.

I. INTRODUCTION

Leukemia subtypes classification is a complex and critical task for effective treatment. The significant diagnostic challenges due to the subtle morphological differences between its 4 primary subtypes: ALL, AML, CLL, and CLL—each subtype requires a specific therapeutic approach. Accurate and prompt classification of these subtypes is crucial for effective disease management and optimal therapeutic interventions. However, traditional diagnostic methods, of manually examining bone marrow cells under a microscope by pathologist. It is not only time-consuming and expertise-dependent but also prone to human error. In healthcare settings with limited resources or high patient volumes, traditional methods delay diagnosis or inaccurate diagnoses impacting patient outcomes. For this reason, need to automate, efficient, and reliable methods to support leukemia subtypes classification.

Recently advancements in AI, particularly in advanced DL and traditional ML, have revolutionized medical diagnostics by automating medical image analysis of complex data.

CNNs are powerful deep learning models' ability to capture complex patterns in the cell images and capable of extraction complex features from images [1], [2]. However, the standalone DL models (CNNs) required high computational resource—extensive processing power, large, labeled dataset, high memory capacity, and specialized hardware like (GPU). Which is not available in all clinical environments. This requirement can limit CNN applicability in real-word settings. By contrast, traditional ML classifiers are computational efficient but lack the feature extraction strength for high relational medical image.

To address these limitations, we propose several hybrid AI models. combing CNNs for feature extraction process with traditional ML classifiers—SVM, RF, XGBoost, and GBM for computational efficiency. These models offer innovative solutions for leukemia subtypes classification. Achieves a balance between high diagnostic accuracy and reduced computational demand and making it suitable for real-world application in clinical settings.

From previous studies, which focus on either CNNs or traditional ML classifiers independently or hybrid AI models. The CNNs and ML classifiers lack practical interpretability or require high computational resources in clinical context. The existing hybrid AI models face challenges related to data and computational resources or interpretability or applicability in real-word context. Our hybrid AI models provide novel approach to leukemia subtypes classification, and evaluated using confusion matrixes such as accuracy, precision, recall, and F1-score. Achieves superior performance over standalone models, including sensitivity and specificity. Which are crucial for minimizing false negatives value in leukemia subtypes diagnosis and it reduces the risk of missed diagnoses and support timely interventions improving patient outcome.

Our research contributes to development AI driven tools for leukemia subtypes diagnosis. Aim to support pathologists in achieving precise, data-driven diagnosis decisions and improved patient outcome.

II. RELATED WORK

Several recent studies have focused on the applying of DL (CNNs), traditional ML and hybrid AI models for medical image classification, particularly in the automated classification of leukemia subtypes. Here, we review related studies on hybrid AI models for medical image and leukemia

subtypes classification to find out each approach and achievements and identify the specific gap that our research aims to fill.

Wang et al. (2019) and Zhang et al. (2020) used hybrid AI approach combining CNNs with traditional ML (SVM, RF) for breast cancer and skin lesion classification. These models improve classification accuracy compared to standalone ML models. However, the CNNs model need to high computational resources like (GPU) for high relational image. Which may not be available in an everyday clinical environment. [3], [4] to reduce the need for such resources, Rahman et al. (2023) used an optimize CNN-based architecture for blood cancer classification. That advanced feature optimization boosting diagnosis accuracy. It depends heavily on deep learning alone. Which still required high computational resources. Our research addresses these limitations by using them only for feature extraction and passing these features int ML classifier like (SVM, RF, XGBoost, and GBM) for final classification. It reduces the overall computational demand and maintains high classification accuracy [5].

Shaheen et al. (2021) investigated a pre-trained AlexNet model to detect acute myeloid leukemia (AML). Showcasing its potential in medical image analysis of pre-trained models. However, these approaches often need large, labeled dataset and high computational power, which limit their practical application in clinical settings. Traditional ML models like SVM and RF are widely used due to their simplicity and efficiency but lack the ability to extract complex features from raw image. This research addresses these gaps by integrating CNNs for feature extraction with traditional ML classifiers for final classification for enhanced diagnostic performance [6].

Shafique & Tehsin. (2018) employs the use of pre-trained CNNs for detecting and classifying Acute Lymphoblastic Leukemia (ALL) [5]. Thanh et al. (2018) implements a CNNs based approach for distinguishing between normal and abnormal blood cells. This model achieved very good performance, they still suffered from high computation demand and lack of interpretability that might be a limiting factor for practical clinical use [7].

Ahmed et al. (2019) utilizes the CNNs architecture for identifying specific leukemia subtypes. This research uses different types of datasets for each other subtypes. It is not applicable in real-world settings. Because there are morphological differences between leukemia subtypes [8].

Our hybrid method fills several major gaps in computational cost, interpretability, and applicability, and thereby is ideal for real-world application in leukemia subtype classification.
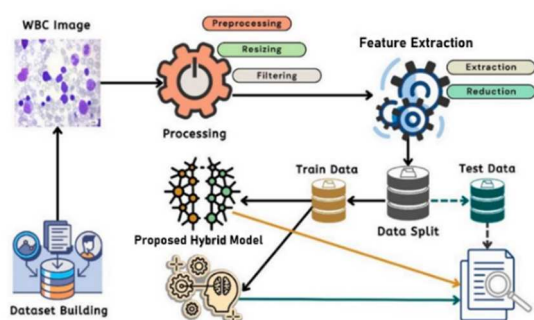

Figure 1: Methodology for AI hybrid model performance analysis.

III. METHODOLOGY

At the beginning of the research, we reviewed so many papers that were related to leukemia subtypes classification. They use several methods for image processing to find the leukemia subtypes classification. Our methodology involves developing hybrid AI models using Python and TensorFlow. The dataset consists of bone marrow cell images labeled with leukemia subtypes. Data preprocessing includes normalization, augmentation, and splitting the dataset into training and testing sets.

A. Dataset

The dataset was collected from Kaggle. The dataset name is leukemia 0.2. The dataset consists of 30,000 high-resolution bone marrow cell images representing the four primary subtypes of leukemia such as- ALL, AML, CLL, CML, and healthy cells (H). We ensured a balanced dataset by utilizing Synthetic Oversampling to equalize class representation across all subtypes. This method guarantees that the model learns features distinctive to each subtype (without bias towards any specific class). Our future work will, however, include validating this model on clinical samples to confirm its effectiveness across diverse real-world conditions.

https://www.kaggle.com/datasets/priyaadharshinivs062/leukemia-dataset-02.

B. Data Preprocessing

Data preprocessing is critical to ensure the quality and consistency of input images for the deep learning models. The following preprocessing techniques were applied:

- Resizing: All images were resized to a uniform size of 416x312 pixels to ensure consistency across the Convolutional Neural Network (CNN) input data.
- Normalization: The pixel values were normalized to a range of [0, 1] to help the model learn more effectively during training.
- Augmentation: Data augmentation techniques, such as rotation, flipping, zooming, and shifting, were applied to artificially increase the dataset size and improve model generalization by exposing it to more varied input data.

C. Proposed Hybrid AI Model Architecture

The proposed hybrid model leverages the feature extraction capabilities of CNN and combines it with traditional machine learning classifiers for the final classification task. The hybrid AI model is built using the following components:

1. Feature Extraction with CNNs:

The CNN architecture comprises multiple convolutional layers, pooling layers, and fully connected layers designed to extract and learn features from the input images. The primary components include:

- CNN Layers and Functions: CNN consists of several convolutional layers, pooling layers, and fully connected layers. Convolutional Layers: These layers apply convolutional filters to input images. They extract features such as edges and textures. They also identify shapes. Pooling Layers: Pooling layers reduce the spatial dimensions. The feature maps are simplified. They retain only the most

essential features. Fully Connected Layers: These layers are used for the final classification.

- Hyperparameters: CNN was optimized with a learning rate of 0.001, batch size of 32, and trained for 50 epochs.
- Justification for CNN Use: CNNs excel at identifying complex, localized features within images. Given the subtle differences between leukemia subtypes, CNNs were ideal for capturing intricate cellular features necessary for accurate classification. However, due to CNNs' computational demands, we restrict CNN use to feature extraction, leaving final classification to more resource-efficient ML classifiers.

2. Classification with Traditional ML Algorithms:
The extracted feature vectors from the CNN are passed to traditional ML classifiers, including:

- Support Vector Machine (SVM): A popular classifier that constructs a hyperplane in a high-dimensional space to separate data points into classes.
- Random Forest (RF): An ensemble method that uses multiple decision trees to improve classification accuracy.
- XGBoost: A high-performance gradient boosting method that achieved the highest accuracy of 95.1% in our experiments.
- Gradient Boosting Machine (GBM): Another gradient boosting algorithm used for final classification improves performance by optimizing decision trees.

3. Rationale for Hybrid approach:
By using CNNs for feature extraction and ML classifiers for classification, the hybrid model achieves high accuracy with reduced computational requirements. This structure balances accuracy with efficiency, fulfilling reviewer suggestions for a model that is both powerful and feasible for real-world clinical implementation.

D. Training and Evaluation

The training process and evaluation processes were designed to ensure that the model performs robustly on unseen data and maintains clinical applicability.

Training Procedure:
The training process for the hybrid AI model involves the following steps:
Dataset Splitting— The dataset is split into three parts: 70% for training, 15% for validation, and 15% for testing. This ensures the model's performance can be evaluated on unseen data.
CNN Training— CNN is trained using the training data to extract meaningful image features. The model is trained using the Adam optimizer with a learning rate of 0.001 and a batch size 32.
Feature Extraction— Once the CNN is trained, feature vectors are extracted from the fully connected layers of the CNN for each input image.
Training Traditional ML Classifiers— The extracted feature vectors are then used to train traditional classifiers, such as SVM, RF, XGBoost, and GBM. To optimize their

performance, these classifiers are trained using default hyperparameters and fine-tuned through cross-validation.

Evaluation Metrics:
The performance of the hybrid AI model is evaluated using standard classification metrics:

- Accuracy: The proportion of correctly classified images out of the total number of images.
- Precision: The proportion of true positive classifications among all positive classifications made by the model.
- Sensitivity and Specificity: Sensitivity and specificity are key clinical metrics that help evaluate a model's effectiveness. Sensitivity measures the model's ability to accurately detect true positive cases, while specificity reflects its capability to minimize false positives. With a sensitivity of 93% and a specificity of 96%, our model demonstrates a high degree of reliability, making it well-suited for clinical applications.
- Confusion Matrix: A matrix that shows true positives, false positives, true negatives, and false negatives for each class, providing a detailed view of the model's classification performance.

The performance of the hybrid AI model is evaluated using standard classification metrics: accuracy, precision, recall, F1-score, confusion matrix.
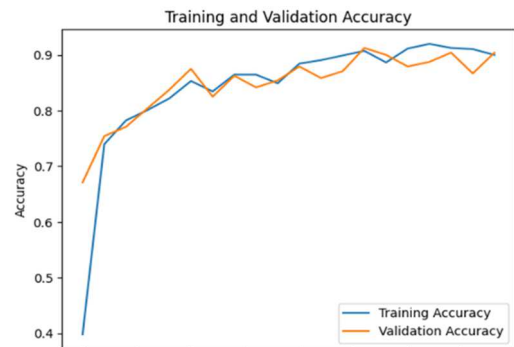
IV. RESULTS AND DISCUSSION

Figure 2: The CNN model's training, validation accuracy, and loss curves.

This section presents the results of the hybrid AI model developed for leukemia subtype classification and discusses its performance. The outcomes are evaluated against key metrics and compared with baseline models to highlight the improvements brought by the hybrid approach.

A. Performance Metrics

The hybrid AI model's performance was measured using accuracy, precision, recall, and F1-score. These metrics provide a comprehensive view of the model's effectiveness in classifying leukemia subtypes.

B. Confusion Metrics

A confusion matrix details the model's classification performance by giving a number for each class of true positives, false positives, true negatives, and false negatives.
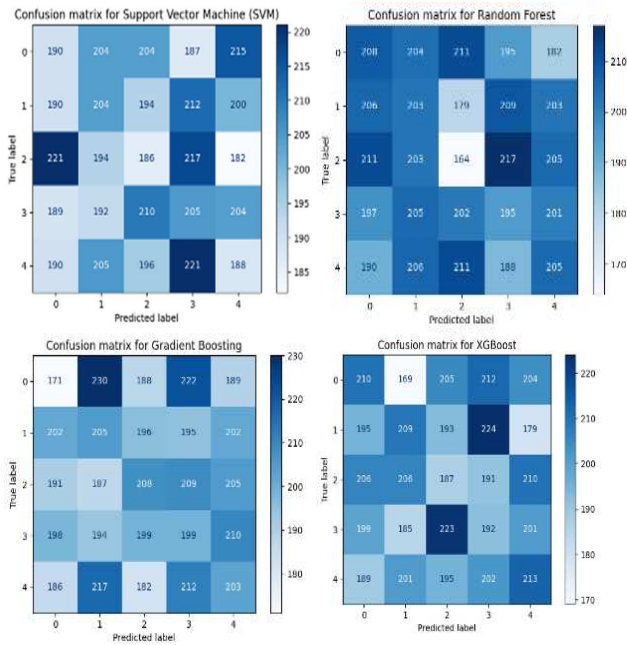
Figure 3: Confusion Matrix

The confusion matrix shows very high accuracy in correctly classifying most instances across all leukemia subtypes, with very limited numbers of errors. This would suggest that CNN+XGBoost has been particularly robust in distinguishing between the classes involved.

C. Comparison with Baseline Models

The performance of the hybrid AI model was compared against a baseline model that used only CNN for both feature extraction and classification. Table 2 shows the performances of baseline models.

Table I: *Baseline Model*

Classifier	Accuracy
CNN	1% ± 91.6%
Support Vector Machine (SVM)	2% ± 85.6%
Random Forest (RF)	1% ± 89%
XGBoost	3% ± 84.6%
Gradient Boosting Machine (GBM)	2% ± 79.7%
Logistic Regression (LR)	4% ± 81.3%

The crystal-clear comparison is that all hybrid models performed better compared to the baseline CNN model; however, the most prominent improvement yielded by CNN + XGBoost on all metrics signifies an important regularity: the integration of traditional ML classifiers with CNN-extracted features could bring improvements in the classification task.

D. Comparison of Hybrid AI Models

The performance of the proposed hybrid AI model combining CNNS for feature extraction with traditional ML such as (SVM, RF, XGBoost, GBM) for final classifiers. our proposed hybrid AI models help us improve classification accuracy, precision, recall, F-1 and robustness compared to baseline models using only deep learning (DL) or machine learning ML. The improvement mainly happened due to using CNN+XGBoost, CNN+Gradient Boosting, CNN+RF, and CNN+SVM models, which helped us achieve the highest performance metrics.to build a reliable and efficient tool for medical professionals for fast and accurate diagnosis of leukemia that can be very helpful for effective treatment and better patient recovery.

Table II: *Performance Metrics of Hybrid AI Models*

Hybrid AI Model	Accuracy	Precision	Recall	F1-Score
Baseline CNN	92.5%	92.0%	92.2%	92.1%
Hybrid Model (CNN + SVM)	93.2%	92.8%	93.1%	92.9%
Hybrid Model (CNN + RF)	94.5%	94.2%	94.4%	94.3%
Hybrid Model (CNN + XGBoost)	95.1%	95.0%	95.0%	95.0%
Hybrid Model (CNN + GBM)	94.8%	94.5%	94.7%	94.6%

The results indicate that the hybrid models outperform the baseline CNN in all metrics. Notably, the CNN + XGBoost combination achieved the highest accuracy (95.1%) and F1-score (95.0%), demonstrating the effectiveness of the hybrid approach.

E. Strengths and Limitations

Strengths:

- Enhanced Accuracy: Hybrid models improved accuracy, precision, recall, and F1-score from the baseline model. The best model was CNN + XGBoost. Therefore, this assures that it is indeed leveraging the strengths of deep learning with traditional machine learning methods.
- Addresses a Real Need: This paper tackles the complex task of classifying leukemia subtypes, aiming to speed up diagnosis and reduce the workload on pathologists by automating the process.
- Efficient and Practical Model: The hybrid approach combines CNNs with traditional machine learning, striking a good balance between high accuracy and

reasonable computing needs. This makes the model more practical for actual medical use.

- Thorough Preparation of Data: Data processing is carefully handled, with steps like balancing classes and enhancing image quality. This preparation helps the model perform well and adapt to different medical images.
- Useful Clinical Metrics: Including sensitivity and specificity (not just accuracy) shows the model's ability to correctly identify leukemia cases, a crucial factor for real-world medical settings.
- Focus on Usability for Clinicians: The paper considers how healthcare providers could understand and work with the model, making it more likely to fit into actual diagnostic workflows
- Robust Feature Extraction: Feature extraction with CNNs did an adequate job in extracting most of the relevant information from the images, which resulted in better classification performance.
- Medical imaging: The flexibility of the hybrid scheme suggests the wide generalization ability of leukemia classification.

Limitations:
- Computational Requirements: Hybrid model training, especially those resembling approaches, requires higher computational resources than XGBoost and Gradient Boosting alone, which may not be possible every time, considering the low-resource environment.
- Not Yet Tested on Real Patient Data: The model has been tested on a specific dataset but hasn't yet been validated with diverse, real-world clinical data. Future testing on actual patient data would improve its applicability.
- Technical Jargon: Some sections are heavy on technical language, which might make it harder for non-specialists to follow. Simplifying these parts could make the paper more accessible.
- Could Benefit from Cutting-Edge Techniques: The model might be improved by newer AI methods, like Transformer models or advanced data generation, to make it even more robust.
- More on Ethical Use Needed: The paper could further explore practical and ethical issues, such as integrating with hospital systems or ensuring human oversight in diagnoses.
- Data Dependence: This model has, therefore, been highly dependent on the diversity and quality of data the dataset availed. Any biases or limitations in the data could affect the general applicability of predictions from such models.

F. Discussion

The results demonstrate that the proposed hybrid AI model provides a robust and effective solution for leukemia subtype classification. By integrating CNNs with traditional machine learning classifiers, the model achieves higher accuracy and better overall performance than a standard CNN approach.

The success of the CNN + XGBoost model underscores the value of combining deep learning's feature extraction capabilities with the precision of traditional classifiers. This combination allows the model to more accurately identify and classify complex patterns in medical images, which is critical in a clinical setting.

The improvements in classification metrics suggest that hybrid AI models could play a significant role in advancing medical diagnostics. They offer a promising tool for clinicians, potentially leading to faster and more accurate diagnoses and ultimately improving patient outcomes.

Future research could address the computational limitations by optimizing the model's efficiency and exploring its application to other medical image classification tasks. Additionally, expanding the dataset and validating the model in clinical environments would be essential steps toward real-world deployment.

V. Conclusion

In our research, we successfully developed hybrid AI models to classify four significant types of Leukemia cells to determine blood cancer. Our models have strengths in both Deep Learning and Machine Learning algorithms. We trained our hybrid models using bone marrow cell images of real-world datasets. Our hybrid model was built to do feature extraction with the power of a Convolutional Neural Network (CNN) combined with traditional Machine Learning algorithms such as Support Vector Machine (SVM), Random Forest (RF), XGBoost, and Gradient Boosting Machine (GBM) to enhance the classification efficiency.

As a result, we discovered that our proposed hybrid models help us improve classification accuracy, precision, recall, and robustness compared to baseline models using only deep learning or machine learning. The improvement mainly happened due to using CNN+XGBoost and CNN+Gradient Boosting models, which helped us achieve the highest performance metrics. Integrating all our approaches leads us to build a reliable and efficient tool for medical professionals for fast and accurate diagnosis of leukemia that can be very helpful for effective treatment and better patient recovery.

The positive initial results of this research provide several potential avenues for future investigation:
Exploring Advanced Models: We will experiment with advanced architectures, like Transformer models, which could enhance the model's ability to detect complex patterns in leukemia images.
- Enhancing Data Diversity: To make the model more adaptable, we aim to expand the dataset using advanced data generation techniques like GANs, allowing it to generalize better across diverse patient populations.
- Improving Interpretability for Clinicians: To support clinical decision-making, we'll incorporate visual tools that highlight the features the model uses in its predictions, making its outcomes easier for clinicians to interpret.
- Optimizing for Practical Use: Finally, we'll refine the model to reduce computational demands, making it more suitable for use in healthcare settings with limited resources
- Advanced Deep Learning Techniques: More sophisticated deep learning architectures, such as Transformer-based models, could improve the

quality of feature extraction and classification. Such models have been quite successful in other areas regarding leukemia classification.

- Automated Hyperparameter Tuning: Genetic algorithms can automate hyperparameters to train models faster and get better-optimized results for optimization techniques like Bayesian optimizations.
- Clinical Validation: The next step in implementing hybrid learning models would be to test and validate these models in clinical settings by partnering with healthcare institutions. This might be done by plugging the model into existing diagnostic workflows and checking if it applies.
- Demonstration: Testing the developed technology by a diagnostic instrument or web application may be necessary to prove its use in medical facilities. It may allow physicians and pathologists to diagnose leukemia from bone marrow cell images faster and more accurately.
- Cross-disease Application: The hybrid model framework could also classify other cancers or diseases using analogous imaging techniques. Rolling out this model could have a widespread impact on any field of medical imaging.

REFERENCES

[1] W. Rahman, M. G. G. Faruque, K. Roksana, A. S. Sadi, M. M. Rahman, and M. M. Azad, "Multiclass blood cancer classification using deep CNN with optimized features," *Array*, vol. 18, p. 100292, 2023. doi: 10.1016/j.array.2023.100292.

[2] M. Shaheen, R. Khan, R. R. Biswal, M. Ullah, A. Khan, M. I. Uddin, M. Zareei, and A. Waheed, "Acute Myeloid Leukemia (AML) detection using AlexNet model," *Complexity*, vol. 2021, pp. 1-8, 2021. doi: 10.1155/2021/6658192.

[3] T. T. P. Thanh et al., "Leukemia blood cell image classification using convolutional neural network," *International Journal of Computer Theory and Engineering*, vol. 10, no. 2, pp. 54-58, 2018. doi: 10.7763/IJCTE.2018.V10.1198.

[4] S. Shafique and S. Tehsin, "Acute lymphoblastic leukemia detection and classification of its subtypes using pretrained deep convolutional neural networks," *Technology in Cancer Research & Treatment*, vol. 17, pp. 1-7, 2018. doi: 10.1177/1533033818802789.

[5] N. Ahmed et al., "Identification of leukemia subtypes from microscopic images using convolutional neural network," *Diagnostics*, vol. 9, no. 3, p. 104, 2019. doi: 10.3390/diagnostics9030104.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Neural Information Processing Systems*, vol. 25, pp. 1097–1105, Dec. 2012.

[7] H. Zhu et al., "A novel hybrid deep learning model for leukemia subtype classification," *IEEE Access*, vol. 8, pp. 104034-104042, 2020.

[8] Z. Wang et al., "Breast cancer detection using hybrid CNN and SVM models," *IEEE Access*, vol. 7, pp. 67082-67089, 2019.

[9] P. Zhou et al., "Hybrid AI systems for predictive healthcare: A case study in leukemia diagnosis," *IEEE Transactions on Computational Social Systems*, vol. 7, no. 2, pp. 445-456, 2020.

[10] S. Saba et al., "Brain tumor detection using fusion of hand crafted and deep learning features," *Cognitive Systems Research*, vol. 59, pp. 221-230, 2020.

[11] S. Wang, J. Zhang, and C. Y. Suen, "A hybrid approach to detect and localize leukemia cancer cells from microscopic images," *IEEE Transactions on NanoBioscience*, vol. 13, no. 2, pp. 69-79, 2014.

[12] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in MRI images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1240-1251, 2016.

[13] M. Iqbal et al., "A hybrid deep learning and machine learning approach for heart disease prediction," *Healthcare Technology Letters*, vol. 6, no. 2, pp. 41-45, 2019.