Review

A Review of Deep Learning Techniques for Leukemia Cancer Classification Based on Blood Smear Images

Rakhmonalieva Farangis Oybek Kizi, Tagne Poupi Theodore Armand [ID] and Hee-Cheol Kim *[ID]

Institute of Digital Anti-Aging Healthcare, Inje University, Gimhae 50834, Republic of Korea; farangisrahmonaliyeva@gmail.com (R.F.O.K.); poupiarmand2@gmail.com (T.P.T.A.)
* Correspondence: heeki@inje.ac.kr; Tel.: +82-55-320-3720

Abstract: This research reviews deep learning methodologies for detecting leukemia, a critical cancer diagnosis and treatment aspect. Using a systematic mapping study (SMS) and systematic literature review (SLR), thirty articles published between 2019 and 2023 were analyzed to explore the advancements in deep learning techniques for leukemia diagnosis using blood smear images. The analysis reveals that state-of-the-art models, such as Convolutional Neural Networks (CNNs), transfer learning, Vision Transformers (ViTs), ensemble methods, and hybrid models, achieved excellent classification accuracies. Pre-processing methods, including normalization, edge enhancement, and data augmentation, significantly improved model performance. Despite these advancements, challenges such as dataset limitations, the lack of model interpretability, and ethical concerns regarding data privacy and bias remain critical barriers to widespread adoption. The review highlights the need for diverse, well-annotated datasets and the development of explainable AI models to enhance clinical trust and usability. Additionally, addressing regulatory and integration challenges is essential for the safe deployment of these technologies in healthcare. This review aims to guide researchers in overcoming these challenges and advancing deep learning applications to improve leukemia diagnostics and patient outcomes.

Keywords: leukemia; deep learning; CNN; ViT; classification; detection; blood smear

1. Introduction

Leukemia is a type of cancer that affects the blood and bone marrow, leading to an abnormal proliferation of immature white blood cells (also known as leukocytes). These abnormal cells, called leukemia cells, disrupt the normal function of healthy blood cells, impairing the body's ability to fight infections and control bleeding.

A total of 474,519 new cases of leukemia were reported in 2020. The global age-standardized incidence rate was 5.4 per 100,000, with an almost five-fold variation worldwide [1]. The American Cancer Society estimates that in the United States, in 2023, there will be over 6500 new cases of Acute Lymphoblastic Leukemia (ALL) and almost 1400 deaths will have occurred. Additionally, projections for the United States in 2024 estimate approximately 62,770 new leukemia cases and about 23,670 related deaths. Sixty percent of all ALL cases occur in children, with a peak incidence at age 2 to 5 years; a second peak occurs after age 50. ALL is the most common cancer in children and represents about 75% of leukemias among children < 15 years of age. The risk declines slowly until the mid-20s and then begins to rise again slowly after age 50. ALL accounts for about 20% of adult acute leukemias. The average lifetime risk of ALL in both sexes is about 0.1% (1 in 1000 Americans) [2].

A leukemia diagnosis is usually made by analyzing a patient's blood sample through a complete blood count (CBC) or microscopic evaluation of the blood or flow cytometry [3]. Leukemia can be broadly classified into four main types based on the speed of disease progression and the types of blood cells affected, as illustrated in Figure 1:

- Acute Lymphoblastic Leukemia (ALL): This type of leukemia progresses rapidly, affecting lymphoid cells, a type of white blood cell that produces antibodies to fight infections.
- Chronic Lymphocytic Leukemia (CLL): CLL progresses slowly and affects mature lymphocytes, a type of white blood cell that helps the body fight infection.
- Acute Myeloid Leukemia (AML): AML progresses rapidly, affecting myeloid cells, a type of white blood cell that gives rise to red blood cells, platelets, and other white blood cells.
- Chronic Myeloid Leukemia (CML): CML progresses slowly and affects myeloid cells. It is characterized by an abnormal chromosome called the Philadelphia chromosome.
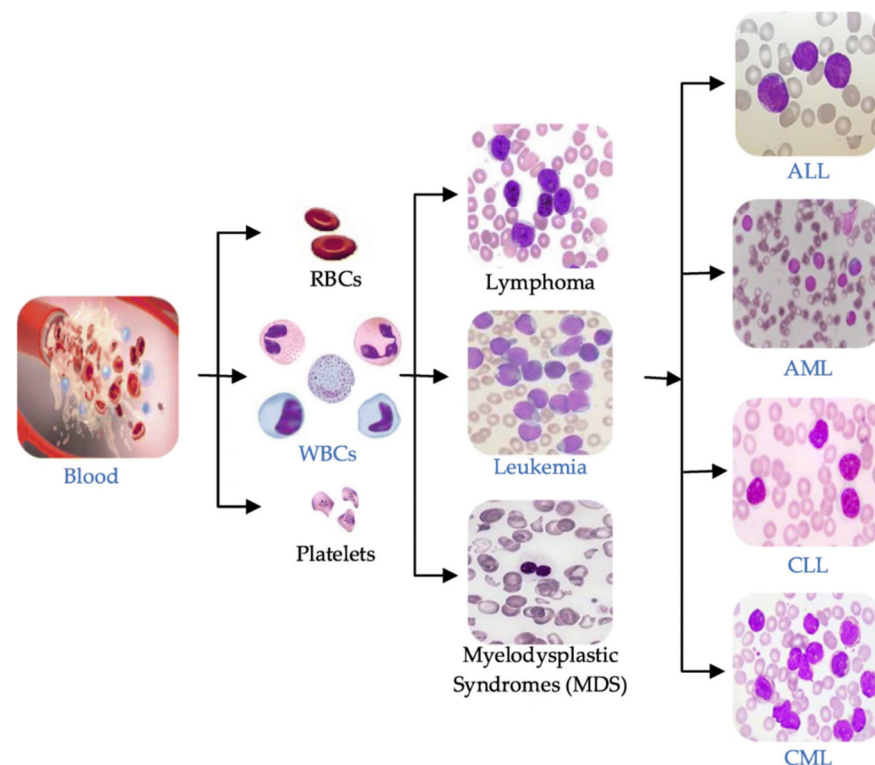

Figure 1. Blood components and main types of leukemia.

The leukemia cells divide rapidly, and the disease progresses quickly. If you have acute leukemia, you will feel sick within weeks of the leukemia cells forming. Acute leukemia is life-threatening and requires immediate initiation of therapy. The early detection of leukemia is crucial in improving patient outcomes, reducing disease-related complications, and increasing the likelihood of successful treatment. Deep learning techniques have revolutionized medical image analysis in recent years by enabling automated and precise classification of pathological conditions. Leveraging Convolutional Neural Networks (CNNs) and Vision Transformers, deep learning models can extract intricate features from blood smear images and discern subtle differences between leukemia subtypes. This capability holds immense potential for enhancing diagnostic accuracy and streamlining the workflow of hematologists. However, despite the growing interest in deep learning-based leukemia classification, there remains a need for a systematic review

to comprehensively assess the existing literature and identify gaps and opportunities for future research. Synthesizing findings from studies published between 2019 and 2023, this paper aims to provide insights into the effectiveness of different deep learning approaches, evaluate their performance metrics, and highlight areas for further investigation.

This review aims to consolidate evidence from diverse sources and offer valuable insights into the ultramodern methodologies for leukemia classification.

The suggested taxonomy includes various models that result in varying levels of accuracy and sensitivity for diagnosing leukemia. In this study, we divided the offered publications into five sections. The diagnosis of leukemia using deep learning consists of five steps. Figure 2 depicts the following input categories: preprocessing methods, DL techniques, ensemble methods, and hybrid methods.
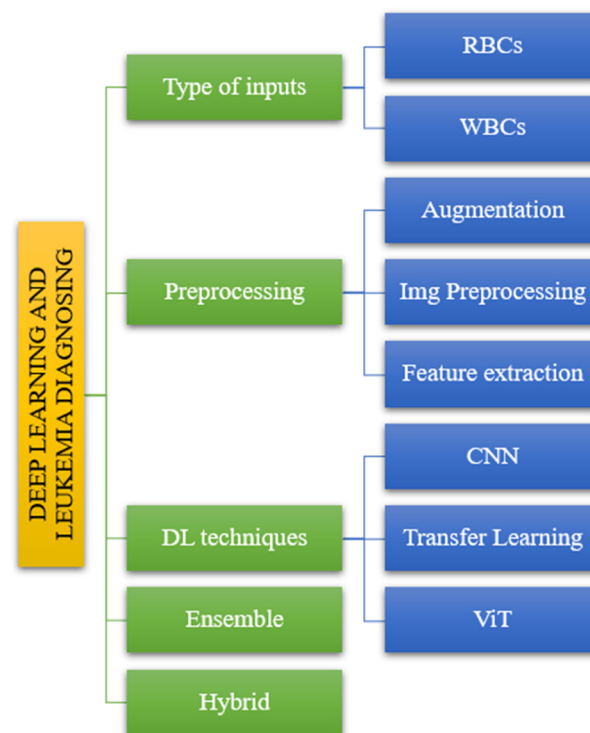

Figure 2. The taxonomy of leukemia diagnosis based on reviewed papers.

This paper presents a comprehensive systematic review of deep learning techniques for leukemia classification using blood smear images, addressing the critical need for an in-depth assessment of existing literature. Synthesizing findings from recent studies provides valuable insights into advancements, challenges, and future directions in leveraging deep learning for accurate and efficient leukemia diagnosis. This work is distinct from prior reviews and offers a detailed comparison of state-of-the-art models, evaluating their performance, limitations, and real-world applicability. Moreover, it uniquely emphasizes ethical considerations, the integration of AI into clinical workflows, and underexplored areas such as the development of explainable AI and the creation of diverse, high-quality datasets. This review serves as a vital resource for researchers, clinicians, and stakeholders, guiding the future development and adoption of AI-driven solutions in leukemia diagnostics.

The rest of the article is organized as follows: Section 2 describes the proposed research method and questions. Section 3 consists of the main studies, and Section 4 consists of analyses based on the literature review. In addition, Section 5 answers research questions. Finally, Section 6 presents the conclusion and future work.

2. Methodology

This study employs a systematic approach like previous works, such as the Cruz-Benito et al. research and others [4–6]. The methodologies include systematic mapping study (SMS) and systematic literature review (SLR). The methods used during this study will be detailed further.

2.1. Research Questions

Systematic review is a fundamental term within academic discourse, denoting a meticulously structured process essential for synthesizing and evaluating existing literature on a particular subject. It encompasses a rigorous method aimed at comprehensively finding, selecting, and critically appraising relevant studies pertinent to a defined research question or topic of inquiry. This is the standard for quality and legal academic review articles. This type of article helps readers develop creative ideas [6]. This systematic review aims to address the following research questions:

RQ1: From 2019 to 2023, how many articles were published on leukemia classification using deep learning techniques? How many of these are SLR and SMS?

By addressing RQ1, researchers can gain insights into the growth trajectory of research in leukemia classification using deep learning techniques over the specified period and the prevalence of systematic review methodologies within this research domain.

RQ2: Which dataset(s) are commonly used for leukemia classification using deep learning? Which dataset is considered the most suitable or effective for this purpose?

Understanding prevalent datasets provides insights into the standardization and availability of data for training and evaluation purposes. Finding the most effective dataset can guide future research directions and benchmarking efforts.

RQ3: How do researchers preprocess blood images before applying deep learning techniques for leukemia classification? Which preprocessing methods have been identified as effective in improving classification accuracy?

This question explores the various preprocessing methods employed in leukemia classification research and their effectiveness. By evaluating the impact of preprocessing methods on classification accuracy, this question aims to find the most beneficial techniques for perfecting model performance.

RQ4: Which deep learning models are effective for leukemia classification? Are there specific architectures or approaches that are particularly successful?

This question investigates which models have shown effectiveness in accurately classifying leukemia subtypes.

RQ5: What are the main challenges encountered in the reviewed studies, and what future research directions are suggested to overcome these challenges and further improve the accuracy and interpretability of deep learning models in leukemia diagnosis?

This question is a comprehensive inquiry into the current research on deep learning models for leukemia diagnosis. It highlights the key obstacles and provides valuable insights into how these challenges can be addressed to advance the field and improve patient care.

This structured approach makes each research question distinct and allows for a clear exploration of several aspects of leukemia classification using deep learning techniques.

2.2. Inclusion and Exclusion Criteria

Using certain criteria, relevant articles exploring deep learning algorithms for leukemia cancer classification based on blood pictures were found for this systematic review. The

requirements were set to guarantee the selection of thorough and relevant literature while keeping the study's goals front and center(Table 1).

Table 1. Some of the inclusion and exclusion criteria.

Inclusion Criteria	Exclusion Criteria
Studies on different methods of leukemia classification and detection using DL	Studies that are not related to research questions
Comprehensive articles with clear evidence	Incomplete or insufficiently detailed articles
Studies were published from 2019 to 2023	Articles published before 2019 or after 2023
Only the articles that have been published in English are considered	Articles published in languages other than English
Articles that includes keywords "Deep learning", "Vision transformer", "CNN", "neural networks", "Transfer learning", "leukemia", "acute myeloid leukemia", "acute lymphoblastic leukemia", "Blood smears" are included.	Articles that include the keywords "DNA", "RNA", and "MICROARRAY" are excluded.

2.3. Search Process

Database Selection: We thoroughly searched multiple electronic databases for this systematic review to ensure comprehensive coverage of relevant literature. The selection of databases was crucial to accessing a wide array of scholarly articles from various disciplines. IEEE Xplore, PubMed, Google Scholar, Springer, and Science Direct databases were used for the search process.

Search Strategy: A comprehensive search strategy was developed using relevant keywords, Boolean operators, and controlled vocabulary (e.g., MeSH terms). Variations and synonyms for key concepts such as "deep learning", "leukemia", "blood smear", and "classification" were included. Here is the search string used to query our databases: ((deep learning OR neural networks OR convolutional neural networks OR vision transformer OR transfer learning) AND (leukemia OR blood cancer OR ALL) AND (blood smear OR white blood cell images)) AND ("1 January 2019" [Date—Publication]: "31 December 2023" [Date—Publication]) AND English [Language]

Execution of Search: The search was conducted across selected databases using the developed search strategy. The number of results retrieved from each database was recorded.

Screening of Search Results: The titles and abstracts of the retrieved articles were reviewed to determine their relevance to the research questions and inclusion/exclusion criteria.

2.4. Review Phases

Screening of titles and abstracts: We first screened the titles and abstracts of the retrieved articles against our inclusion and exclusion criteria. Articles that did not meet the requirements, such as those unrelated to leukemia classification or not involving deep learning techniques, were excluded from further consideration. This initial screening process helped narrow down the pool of articles to those most relevant to our systematic review.

Full-text assessment: After completing the initial title and abstract screening, the next phase involved a comprehensive evaluation of the full texts of articles that emerged as potentially relevant during the earlier screening process. This critical step was pivotal in determining the final eligibility of each article for inclusion in the systematic review. We meticulously assessed each article against the predetermined inclusion and exclusion criteria established for the review. This thorough examination involved a detailed scrutiny

of the content, methodology, and results presented in the articles. Rigorously applying the criteria allowed definitive determinations regarding the final eligibility of each article for inclusion in the systematic review. Clear documentation of the reasons for excluding articles at this stage ensured transparency and facilitated understanding of the selection process. This phase played a crucial role in upholding the integrity and reliability of the literature in the systematic review, ultimately contributing to its credibility and usefulness in advancing knowledge in the field.

Data extraction: We systematically gathered pertinent information from the selected studies using a standardized data extraction form during the data extraction phase. This form meticulously cataloged key details for synthesis and analysis, encompassing study characteristics such as authors, publication year, preprocessing techniques, feature extraction, deep learning methodologies, model architecture and training methods, and evaluation metrics. Additionally, we recorded details on datasets utilized, key findings, advantages, and disadvantages derived from each study. To ensure accuracy and reliability, data extraction was independently conducted by two reviewers, with any discrepancies resolved through discussion or consultation with a third reviewer when necessary. This meticulous approach to data extraction facilitated a comprehensive and systematic synthesis of information from the included studies, bolstering the validity and robustness of the systematic review findings.

Quality assessment: We used specific quality criteria to guide our interpretation of findings and assess the strength of inferences in the included studies. These criteria helped us evaluate the credibility of each study when synthesizing results [6]:

(1) The study's rigorous data analysis relied on evidence or theoretical reasoning rather than non-justified or ad hoc statements.
(2) The study described the research context.
(3) The design and execution of the research supported the study's aims.
(4) The study described the research method used for data collection.

All included studies met the four criteria, ensuring confidence in their credibility and relevance for data synthesis in the review.

Data synthesis and analysis: This systematic review's data synthesis and analysis phase involved meticulously analyzing the findings extracted from the 30 selected papers identified from an initial pool of 212 papers using a structured search strategy and inclusion/exclusion criteria. Collating and synthesizing the data aimed to discern overarching themes and insights regarding the effectiveness of deep learning techniques for leukemia classification based on blood images. Through careful analysis, we identified prevalent approaches and methodologies employed by researchers in this field and key findings and outcomes reported in the literature. Furthermore, the synthesis phase involved interpreting the synthesized data in the context of the research questions and objectives of the systematic review. This process enabled us to provide valuable insights into the current state of deep learning-based leukemia classification research and identify areas for further investigation and improvement. The analysis was facilitated using spreadsheet software (Excel), which allowed for efficient organization, management, and analysis of the extracted data. This systematic approach to data synthesis and analysis ensures the robustness and reliability of the findings presented in this review paper, contributing to advancing knowledge in leukemia classification using deep learning techniques. To keep information consistent, the data extraction for the 30 studies was driven by a form shown in Table 2.

Table 2. Data extraction for each study.

Extracted Data	Description
Identity of study	Unique identity for the study
Bibliographic references	Author, year of publication, title, and source of publication
Type of study	Book, journal paper, conference paper, etc.
The focus of the study	Main topic area, concepts, motivation, and objective of the study
Used database of the study	Database name, size, classes, image types, file formats, and label type
Preprocessing techniques	Data augmentation, image preprocessing, feature extraction
Deep learning models	Description of the model used in the study.
Evaluation metrics	Description of the evaluation metrics used in the study
Performance of the study	The results obtained, such as accuracy F1-score
Key challenges and limitations	key challenges and constraints mentioned in the study

This flowchart (Figure 3) represents the sequential steps in the paper selection process. The process begins with searching the selected databases and retrieving relevant articles based on the search terms. Subsequently, screening is performed based on the title and abstract of the retrieved articles. A full-text review is conducted on the screened articles to assess their relevance further. The inclusion and exclusion criteria are applied to filter the articles and obtain selected papers that meet the eligibility criteria for the systematic review. Finally, the process concludes with the selected papers being included for analysis and synthesis in the systematic review.
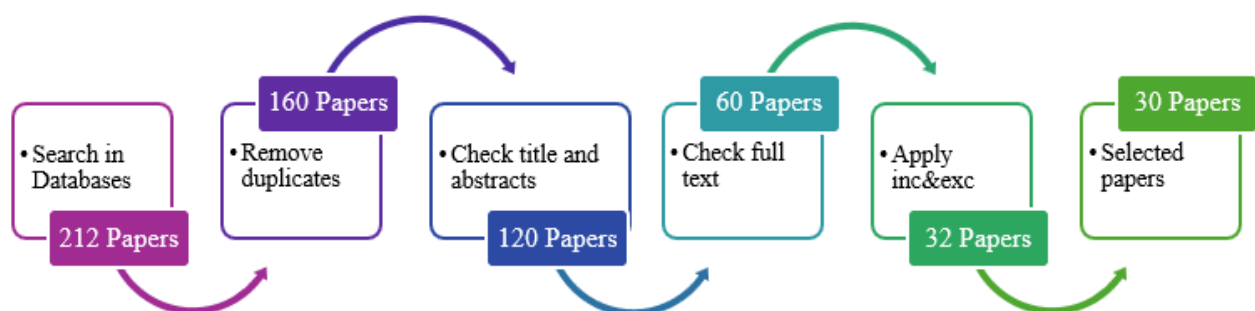
Figure 3. Paper selection process.

3. Literature Review

The articles were reviewed and analyzed from the oldest to the most recent publications within the specified period (2019–2023). This approach allows for the systematic exploration of how research in the field has evolved. Each article published within the defined period was systematically examined to understand its contributions, methodologies, findings, and relevance to the research questions. This method ensures thorough coverage of the existing literature. Following a chronological order enables researchers to track the development of ideas, methodologies, and trends in leukemia classification using deep learning techniques. This allows them to identify seminal works, key advancements, and gaps in knowledge as the research progresses.

In 2019, Nizar Ahmed et al. [7] proposed an innovative approach for diagnosing all leukemia subtypes from microscopic blood cell images using Convolutional Neural Networks (CNN). It explores the use of deep learning techniques to classify leukemia subtypes accurately. The proposed method uses CNN architecture for leukemia diagnosis, focusing on all four subtypes. Data augmentation techniques were applied to increase the dataset size, including rotation, height shift, width shift, zoom, horizontal flip, vertical flip, and shearing. Two publicly available leukemia datasets were used: ALL_IDB [8] and ASH [9]. The CNN model achieved 88.25% accuracy in binary classification (leukemia versus healthy) and 81.74% accuracy in multi-class classification of all leukemia subtypes. Comparative analyses were conducted with other machine learning algorithms, highlighting the effectiveness of the CNN model.

Rohit Agrawal et al. [10] proposed a novel method involving preprocessing, segmentation, feature extraction, and classification using a Convolutional Neural Network (CNN) to achieve a correct diagnosis. The proposed system aims to assist medical professionals in diagnosing several types and sub-types of white blood cell cancer diseases, including Acute Myeloid Leukemia (AML), Acute Lymphoblastic Leukemia (ALL), and Myeloma. The dataset comprises 100 microscopic blood sample images [11], with 62 images for training and 38 for testing. Images of blood samples are converted from RGB to YCbCr color space during input to improve picture quality and make segmentation easier. Utilizing Otsu Adaptive Thresholding in conjunction with Gaussian Distribution to segment images. The K-Means clustering approach is used to identify the segments of the cytoplasm and nucleus and to segment cells. Cells are segmented using K-Means clustering. Texture feature extraction was performed using the Gray Level Co-occurrence Matrix (GLCM). The exact categorization of white blood cell cancer subtypes is achieved using Convolutional Neural Networks (CNNs). The model achieved 97.3% accuracy.

Sara Hosseinzadeh Kassani et al. [12] proposed a hybrid method enriched with different data augmentation techniques to extract high-level features from input images. Features from intermediate layers of two CNN architectures, VGG16 and MobileNet, are fused to improve classification accuracy. Two methods for normalization are mean RGB subtraction divided by standard deviation and ImageNet mean subtraction. Images are resized from 450×450 pixels to 380×380 pixels using bicubic interpolation. For data augmentation, various techniques are applied, including contrast adjustments, brightness correction, horizontal and vertical flips, and intensity adjustments. Hybrid CNN Architecture combines features from VGG16 and MobileNet architectures to enhance classification accuracy. Features from specific abstraction layers are fused to improve discriminative capability. It utilizes low-level features from intermediate layers to generate high-level discriminative feature maps. The dataset used for experimentation is based on the classification of normal versus malignant cells in B-ALL white blood cancer microscopic images provided by SBILab. The dataset has 76 individual subjects, 47 ALL subjects, and 29 normal subjects, including 7272 ALL cell images and 3389 normal cell images [13]. The proposed method achieves an overall accuracy of 96.17%, sensitivity of 95.17%, and specificity of 98.58%. Comparative analysis shows that the proposed approach outperforms individual CNN architectures (VGG16 and MobileNet) and earlier studies regarding accuracy.

In 2020, Mohamed Loey et al. [13] proposed two automated classification models based on blood microscopic images to detect leukemia. It employs transfer learning using AlexNet pre-trained deep CNNs. Image preprocessing operations include converting images to the RGB color model, resizing them to fixed sizes (227×227), and data augmentation (translation, reflection, and rotation) to overcome the lack of large datasets. The first model involves preprocessing, feature extraction using AlexNet, and classification using various classifiers (SVM, LD, DT, K-NN). The second model used the same preprocessing

steps but employed fine-tuning of AlexNet for feature extraction and classification. The dataset consists of 564 (282 healthy and 282 unhealthy) blood microscopic images, with augmentation expanding to 2820 images. The images were collected from studies [9,14] and captured with an optical laboratory microscope and a camera. The first model achieved high accuracy using various classifiers, with SVM-Cubic performing the best with 99.79% accuracy. The second model, employing fine-tuned AlexNet, outperformed the first model in all metrics with 100% accuracy. Classification accuracy was evaluated through precision, recall, accuracy, and specificity, with the second model showing superior performance.

Puneet Mathur et al. [15] proposed a novel architecture for detecting early-stage Acute Lymphocytic Leukemia (ALL) cells using a Mixup Multi-Attention Multi-Task Learning Model, which introduces Pointwise Attention Convolution Layers (PAC) and Local Spatial Attention (LSA) blocks to capture global and local features simultaneously. It incorporates Rademacher Paired Sampling Mixup to prevent the memorization of training data in cases of limited categorical shift. The MMA-MTL architecture combines Global Attention Channel (GAC) and Local Attention Channel (LAC) for joint attention on feature spaces. Pointwise Attention Convolutions (PACs) reinforce self-attention mechanisms in convolutional layers. Local Spatial Attention (LSA) layers are inspired by multi-gaze attention cells, enabling unsupervised localization. Rademacher Paired Sampling Mixup is applied as a regularization technique. The ISBI 2019 CNMC dataset [11] was used for experimentation, consisting of 10,661 cell images, with 7272 images belonging to the malignant class (ALL) and 3389 to the benign class (HEM). The dataset undergoes stain normalization and resizing to $224 \times 224 \times 3$ for input. The proposed MMA-MTL model shows competitive performance compared to baseline methods such as EfficientNet, WideResNet-50, WSDAN, and ensemble models. The model achieved an F1-score of 0.9189, precision of 0.8211, recall of 0.9385, and ROC-AUC of 0.8931.

Syadia Nabilah Mohd Safuan et al. [16] proposed a computer-aided system (CAS) using CNN for automated white blood cell (WBC) detection and classification, particularly for finding ALL. A CNN is chosen for its capability in visual recognition tasks and its ability to minimize overfitting, providing better robustness and efficiency. The proposed method involves training pre-trained CNN models (AlexNet, GoogleNet, and VGG-16) on two datasets: IDB-2 [8] has 260 images with two classes (lymphoblast and non-lymphoblast). At the same time, Leukocyte Image Segmentation and Classification LISC [17] consists of 242 images with five classes (basophil, eosinophil, neutrophil, lymphocyte, and monocyte). These models are evaluated for their ability to classify WBC types, particularly distinguishing between lymphoblast and non-lymphoblast cells in the IDB-2 dataset and classifying five WBC types in the LISC dataset. For the IDB-2 dataset, AlexNet achieved the highest training accuracy of 96.15% and testing accuracy of 97.74% for lymphoblast classification. VGG-16 achieved the highest accuracy for non-lymphoblast classification. AlexNet exhibited the shortest elapsed time compared to GoogleNet and VGG-16. For the LISC dataset, AlexNet attained the highest training accuracy of 80.82% and achieved the highest testing accuracy for all classes except monocyte. VGG-16 achieved the highest testing accuracy for monocytes. The paper concludes that AlexNet consistently outperforms other pre-trained models in WBC classification for both datasets.

Shamama Anwar et al. [18] proposed an automated diagnostic system to improve the efficiency and accuracy of diagnosis to detect ALL using a CNN. This study uses data from ALL_IDB [11] for training and testing purposes. ALL_IDB1 has 108 images, while ALL_IDB2 has 260 images, resulting in a total of 368 images initially. The dataset size is doubled to 736 images through data augmentation techniques, enhancing the diversity and quantity of data available for training and testing the model. The proposed method employs a 10-layer CNN architecture that directly processes raw images. Through a series

of convolutional layers, activation functions, and pooling layers, the model automatically extracts features from the images. These learned features are then used to classify the images into blast or non-blast cell categories. Notably, the model does not require explicit feature extraction techniques, as it learns to extract features during training. The architecture encompasses convolutional layers and activation functions like LeakyReLU, pooling, flattening, fully connected, dropout, and SoftMax layers to facilitate classification. The model achieved impressive results: During training, it reached a 95.45% accuracy after 100 iterations, with each iteration consisting of 20 epochs. In testing, the model showed even higher performance, with an average accuracy exceeding 99% and reaching 100% accuracy in most trials.

Md. Alif Rahman Ridoy et al. [19] proposed an automated method using a lightweight CNN for WBC classification to enhance efficiency and accuracy compared to manual or semi-automated techniques. Preprocessing involves data augmentation techniques, such as balancing to address uneven dataset distribution and downsizing images to 120×160 pixels to reduce computational cost. Based on LeNet architecture with modifications, the proposed CNN model comprises three convolutional blocks followed by an output-dense layer. Training employs the Adam optimization algorithm, early stopping, and categorical cross-entropy loss function. The BCCD dataset [14], containing 352 images of WBCs, is used. The evaluation metrics include accuracy, F1-score, precision, specificity, recall, Cohen's kappa index, confusion matrix, and micro-average AUC score. Despite fewer parameters and a less robust configuration, the model outperforms others in key metrics like the overall F1-score of 0.97 and Kappa index scores.

Nighat Bibi et al. [20] proposed an IoMT-based framework for automated detection and classification of leukemia using deep learning techniques. It introduces a system that utilizes IoT-enabled microscopes to upload blood smear images to a cloud platform for diagnosis. Two deep learning models, ResNet-34 and DenseNet-121, are employed to find leukemia subtypes, proving superior performance compared to existing methods. This study uses two datasets, the ASH [9] Image Bank and ALL-IDB [8], for experiments. The prediction accuracy for ALL and healthy cases is 100%, while the accuracy for other subtypes ranges from 99.65% to 100%.

In 2021, Jens P. E. Schouten et al. [21] explored the efficacy of CNNs in distinguishing between malignant and non-malignant leukocytes, crucial for leukemia diagnosis. The research uses two datasets to investigate how varying training set sizes impact CNN performance. A small ALL dataset has 250 single-cell images of leukocytes from Acute Lymphoblastic Leukemia patients and healthy individuals, and a larger AML dataset includes over 18,000 single-cell images from Acute Myeloid Leukemia (AML) patients and controls. The proposed method uses a small dataset to train a Sequential CNN for the binary classification of leukocytes. Data augmentation techniques such as horizontal and vertical flipping and random rotational transformations augment the training set. Remarkably, even with a few images, the CNN achieves high accuracy, with a ROC-AUC of 0.97 ± 0.02 achieved using only 200 training images.

Pradeep Kumar Das et al. [22] proposed a novel transfer-learning-based method for automatically detecting ALL using the lightweight SqueezeNet architecture. It employs data augmentation and resizing of images for preprocessing to improve classification performance. The proposed method utilizes transfer learning with ShuffleNet, a lightweight and computationally efficient CNN model. Channel shuffling and pointwise-group convolution are introduced to boost performance and speed. The experimental results demonstrate that the ShuffleNet-based ALL classification outperforms existing techniques regarding sensitivity, precision, and accuracy—96.97% IDB1 and 96.67% IDB2, specificity, AUC, and F1-score. Performance analysis is conducted on 70–30 and 50–50 dataset splits.

Zhencun Jiang et al. [23] proposed a novel approach for diagnosing ALL using computer vision technology. They address the challenge of distinguishing between cancer cells and normal cells by developing an ensemble model combining the Vision Transformer (ViT) and Convolutional Neural Network (CNN) models. They preprocess the ISBI 2019 dataset [11] using a difference enhancement-random sampling (DERS) method to balance the dataset. Various data augmentation techniques are applied to generate new images. The proposed method involves training separate models: a Vision Transformer model, which is based on the transformer structure and performs image classification without CNN layers, and an EfficientNet model, known for its efficiency and accuracy in image classification. These models are then integrated into a ViT-CNN ensemble model using a weighted sum method to combine their outputs. The ViT-CNN ensemble model achieves a high classification accuracy of 99.03% on the test set, with balanced recognition accuracy for cancer and normal cells. Comparative analysis with other models, including classic CNN models and a model proposed in the literature, shows the superiority of the proposed method.

Pradeep Kumar Das et al. [24] proposed a novel deep CNNs-based ALL classification. This paper presents a novel deep CNN-based framework for the efficient detection and classification of ALL. It introduces a probability-based weight factor to hybridize MobileNetV2 and ResNet18, addressing the challenge of correct ALL detection with smaller datasets. Data augmentation and resizing are employed for preprocessing, while depth-wise separable convolutions and skip connections enhance computational efficiency. The method achieves the best accuracy of 99.39% and 97.18% in ALLIDB1 and ALLIDB2 datasets, respectively, with a 70% training and 30% testing split, and outperforms recent transfer learning-based techniques.

Y. F. D. De Sant' Anna et al. [25] combined image processing and deep learning techniques to achieve high accuracy with lower computational cost than traditional Convolutional Neural Networks (CNNs). The proposed methodology encompasses feature extraction techniques including first-order statistical, textural, morphological, contour, and discrete cosine transform (DCT) features. These features are extracted from the C-NMC 2019 dataset [11] and augmented to balance the training set. The proposed method achieved an F1-score of 91.2%, focusing on model size and computational time efficiency.

Azamossadat Hosseini et al. [26] developed a lightweight Convolutional Neural Network (CNN) model integrated into a mobile application for diagnosing B-cell Acute Lymphoblastic Leukemia (B-ALL) using blood microscopic images. Using a local dataset of 3242 images from 89 patients, this study addressed challenges distinguishing benign hematogones from malignant lymphoblasts through advanced preprocessing techniques, including LAB color space transformation and K-means clustering. Among three lightweight CNN architectures, including EfficientNetB0, MobileNetV2, and NASNet Mobile, MobileNetV2 achieved 100% accuracy in classifying B-ALL and its subclasses. The model was implemented in a mobile app for real-time diagnosis by combining original and segmented images for optimal feature extraction, reducing the need for invasive bone marrow biopsies. Tested in clinical settings, the app demonstrated high sensitivity and specificity, offering a scalable, efficient, and accessible tool to enhance B-ALL screening and diagnosis.

In 2022, Ibrahim Abunadi et al. [27] proposed a multi-method diagnosis of blood microscopic samples for early detection of Acute Lymphoblastic Leukemia based on deep learning and hybrid techniques. This paper aims to develop diagnostic systems for the early detection of leukemia using two Acute Lymphoblastic Leukemia Image Databases (ALL_IDB1 and ALL_IDB2). It proposes three systems: one based on artificial neural networks (ANNs), feed-forward neural networks (FFNNs), and support vector machine

(SVM); another based on Convolutional Neural Network (CNN) models using transfer learning; and a third hybrid system combining CNN and SVM technologies. Images were enhanced using color channel calculations and scaling, then noise reduction using average and Laplacian filters. The first system used Neural Networks and machine learning, employing an Adaptive Region-Growing Algorithm for segmentation, morphological methods for image improvement, and feature extraction using LBP, GLCM, and FCH algorithms. Features were fused, and classification was performed using an ANN, FFNN, and SVM. The second model focused on Convolutional Neural Networks (CNNs), employing three models (AlexNet, GoogleNet, ResNet-18) with various layers like convolutional, ReLU, pooling, fully connected, and SoftMax. Additionally, a third model involved deep learning–machine learning hybrid techniques, incorporating architectures like AlexNet + SVM, GoogleNet + SVM, and ResNet-18 + SVM. The ANN and FFNN achieved an accuracy of 100%, while the SVM achieved 98.11% accuracy. CNN models (AlexNet, GoogLeNet, and ResNet-18) achieved 100% accuracy for both datasets. Hybrid systems combining a CNN and SVM achieved promising results, with accuracies ranging from 98.1% to 100%.

Maryam Bukhari et al. [28] used squeeze-and-excitation learning to improve channel connections at all levels of feature representation, allowing for a clear representation of channel interdependence and better feature discrimination between leukemic and normal cells. The input images were preprocessed by augmentation and directly fed into the CNN model. Data augmentation techniques such as rotation and random shifts were applied to increase the dataset size. The proposed method employs a deep CNN architecture with squeeze-and-excitation learning for diagnosing leukemia from microscopic blood sample images. Squeeze-and-excitation learning is integrated into the model to improve feature learning discriminability by explicitly modeling channel interdependencies and recalibrating kernel outputs. The dataset used for evaluation is the Acute Lymphoblastic Leukemia Image Database (ALL-IDB), including ALL-IDB1 and ALL-IDB2 datasets. The proposed model achieves 100% accuracy, precision, recall, and F1-score in diagnosing leukemia from the ALL-IDB1 dataset. The model achieves high accuracy, precision, recall, and F1-score on the ALL-IDB2 dataset, with accuracies ranging from 96% to 99.98% in different runs. When both datasets are combined, the model exhibits high accuracy and performance. The proposed model outperforms the traditional deep learning model with an average accuracy improvement of 5.5%.

Zahra Boreiri et al. [29] propose an automated approach using a deep convolutional neuro-fuzzy network. Fuzzy reasoning rules integrated with connectionist networks create a neuro-fuzzy system for ALL detection. The preprocessing involves a two-stage fuzzy color segmentation technique to separate leukocytes from other blood components. Selective filtering and unsharp masking methods are applied for noise reduction. RGB images are converted to Lab* color space to help segmentation. Image segmentation uses color in the bounding box method to extract single nucleus images, which is employed to address the limited training samples, including vertical and horizontal flips and image rotation. The dataset is increased to 1024 images to minimize overfitting. The proposed method focuses on fuzzy inference and pooling operations to implement a convolutional neuro-fuzzy network for image classification. Fuzzy inference operation uses a TSK rule-based system. A fuzzy pooling operation is applied after the fuzzy inference layer for downsampling. A convolutional neuro-fuzzy network architecture is designed for ALL classifications. The ALL-Image Database (IDB) dataset, consisting of ALL-IDB1 and ALL-IDB2 datasets, is used for evaluation. The proposed model achieves an average accuracy of 97.31% for ALL detection.

Tanzilal Mustaqim et al. [30] proposed a modification to the YOLOv4 and YOLOv5 models for detecting Acute Lymphoblastic Leukemia (ALL) subtypes in multicellular

microscopic pictures. It introduces the GhostNet convolution module as a replacement for the regular convolution module in the YOLOv4 and YOLOv5 models. The change looks to lower computational complexity, parameter count, and GFLOPS values while keeping detection accuracy and accelerating calculation time. The preprocessing stage includes manually labeling the dataset using the YOLO format and resizing images to 416×416 pixels. The GhostNet convolution module replaces the standard convolution module in the YOLOv4 and YOLOv5 backbones. The Ghost Bottleneck, consisting of CBS, GhostNet convolution modules, and depth-wise convolutions, is used to reduce the number of parameters and GFLOPS values. The modification aims to perfect the convolution process by reducing redundant feature maps and computational costs. This study uses a multi-cell blood microscopic image dataset [31] containing information on L1, L2, and L3 subtypes. The modified models showed comparable performance with the original YOLO models, with slight differences in accuracy metrics. The GFLOPS value and number of parameters were reduced by approximately 40% and 35%, respectively.

Protiva Ahammed et al. [32] proposed a multistage transfer learning approach for ALL classification. They introduce a 3-stage transfer learning approach and stacks of Convolutional Neural Networks for automated leukemia identification and classification. The proposed method includes three main phases: data preprocessing, image segmentation, and multistage transfer learning. Data preprocessing involves extracting the region of interest (ROI) from microscopic images using RGB channels with a monochrome filter. Specific values for the RGB channels are applied to isolate lymphoblasts, followed by threshold operations to refine the ROI. Image augmentation techniques such as flipping, rotating, shifting, and scaling are used to create added synthetic data for training. A modified Residual U-Net architecture is employed for image segmentation to isolate lymphoblasts and cut other blood components. Multistage transfer learning involves transferring knowledge from pre-trained models, InceptionV3, Xception, and InceptionResNetV2 on natural images, to domain-specific medical datasets through three stages, improving the model's ability to classify leukemia. The dataset comprises 100 images for leukemia detection and 240 for segmentation, obtained from the ALL-IDB1 and ALL-IDB2 repositories. Utilizing InceptionResNetV2 architecture, the proposed method proves high accuracy (99.60% for normal vs. leukemia and 94.67% for normal to L3) and reduces error rates.

Chastine Fatichah et al. [33] propose the use of object detection and instance segmentation techniques, specifically comparing YOLO (You Only Look Once) and Mask R-CNN (Mask Region-based Convolutional Neural Network), for the detection of Acute Lymphoblastic Leukemia (ALL) subtypes. Data augmentation techniques such as rotation, zoom, and flip generate new image variations, augmenting the training dataset and reducing overfitting. This study compares the performance of YOLOv4, YOLOv5, and Mask R-CNN in detecting ALL subtypes. YOLOv4 and YOLOv5 are used for object detection, while Mask R-CNN is used for instance segmentation. The YOLO models are built from the training data, and the testing data are used to evaluate their performance. Similarly, the Mask R-CNN model is trained and evaluated. The dataset comprises white blood cell microscopy images from patients with ALL [31], totaling 301 images, including 128 L1-type images, 63 L2-type images, and 110 L3-type images. YOLOv4 outperforms YOLOv5 and Mask R-CNN in detecting ALL subtypes, achieving an F1 value of 89.5% and a mean Average Precision (mAP) value of 93.2%. YOLOv5 exhibits the best performance in detecting L1 subtypes. Mask R-CNN shows the lowest performance in detecting ALL subtypes.

Protiva Ahammed et al. [34] proposed a transfer learning-based CNN model from microscopic images for Acute Lymphoblastic Leukemia (ALL) prognosis. It aims to address the challenge of accurately and rapidly diagnosing ALL, which is crucial for timely treatment and improving patient outcomes. The ISBI C-NMC 2019 dataset containing

microscopic images of white blood cells was used. Regarding data preprocessing, images were resized to 300 × 300 with three channels. Image augmentation techniques like horizontal flipping and resizing were applied to minimize class imbalance issues. The proposed model uses the Inception V3 architecture for feature extraction and classification. The model includes feature extraction using pre-trained Inception V3 layers and trainable dense layers for deep neural connection and classification. Dropout regularization was used to reduce overfitting. The model achieved an overall accuracy of 98.00%, precision of 98.00%, and F1-score of 98.00%.

In 2023, Sanam Ansari et al. [35] proposed a model that solves the problems of low detection speed and accuracy in previous studies by designing an adaptive CNN model based on the Tversky loss function. Additionally, it introduces a novel dataset augmented with a GAN to enhance scalability. Preprocessing involves resizing, grayscale conversion, normalization, and data augmentation. The proposed CNN architecture includes convolution, pooling, and dense layers with SoftMax activation. The dataset is collected from the Shahid Ghazi Tabatabai Oncology Center in Tabriz, containing images from 44 patients with 184 ALL images and 469 AML images. The experimental results show a 99% accuracy rate, outperforming prior methods.

Another study by Sanam Ansari et al. [36] proposed a novel method for diagnosing acute leukemia by distinguishing between Acute Lymphocytic Leukemia (ALL) and Acute Myeloid Leukemia (AML) through image analysis. It employs a type-II fuzzy deep network for feature extraction and classification. The dataset is collected from the Shahid Ghazi Tabatabai Oncology Center in Tabriz, containing images from 44 patients with 184 ALL images and 469 AML images. The proposed method achieves high accuracy (98.8%) and F1-score (98.9%) in distinguishing between ALL and AML. Additionally, it outperforms models using traditional activation functions like Relu and Leaky-Relu.

Qile Fan et al. [37] proposed a model called QCResNet for the classification of Acute Lymphoblastic Leukemia (ALL) using Convolutional Neural Networks (CNNs). The model achieved a high accuracy of 98.9% on a dataset of 15,135 images. The preprocessing includes resizing, balancing, gamma transforms, and shuffling of the dataset. QCResNet, a modification of ResNet-18, incorporates additional layers and activation functions, leading to faster convergence and improved adaptability to complex tasks. The dataset comprises 15,135 images from 118 patients, with two classes: normal cell and leukemia blast. The proposed method achieves high accuracy (98.9%).

Pradeep Kumar Das et al. [38] proposed a novel model for detecting and classifying Acute Lymphoblastic Leukemia (ALL) or Acute Myelogenous Leukemia (AML) using deep learning, specifically addressing challenges with small medical image datasets. The proposed model integrates ResNet18-based feature extraction with an Orthogonal SoftMax Layer (OSL) for improved classification. Extra dropout and ReLu layers are introduced for enhanced performance. The model is evaluated on standard datasets (ALLIDB1, ALLIDB2, ASH, and C_NMC_2019) and outperforms existing methods regarding the accuracy, precision, recall, specificity, AUC, and F1-score. The proposed model achieved 99.39% accuracy in ALLIDB1, 98.21% in ALLIDB2, and 97.50% in the ASH dataset.

Oguzhan Katar et al. [39] proposed an innovative method for automated detection and classification of white blood cells (WBCs) in blood film images, aiming to overcome the limitations of manual screening, which is time-consuming and subjective. It proposes an explainable Vision Transformer (ViT) model that uses a self-attention mechanism to extract features from input images, dividing them into 16 × 16 pixel patches for classification. The model achieves high accuracy (99.40%) in classifying five types of WBCs and offers precise localization using the Score-CAM algorithm for visualization.

Pranavesh Kumar Talupuri et al. [40] proposed a comprehensive comparative study of four prominent deep learning models (ResNet, InceptionNet, MobileNet, EfficientNet) to classify and detect ALL from microscopic blood cell images. The goal is to assess their performance and guide model selection for automated disease detection. A dataset of 15,135 images is meticulously constructed and subjected to preprocessing, including normalization, noise reduction, and augmentation. The proposed method involved constructing a diverse dataset, exploring four deep learning architectures, training and fine-tuning models on the leukemia dataset, and evaluating their performance using accuracy and F1-score metrics. The achieved accuracies are as follows: ResNet—95.75%, Inception—95.37%, MobileNet—94.81%, EfficientNet—96.44%.

Avinash Bhute et al. [41] offered "ALL Detection and Classification Using an Ensemble of Classifiers and Pre-Trained Convolutional Neural Networks" as a complete technique for enhancing leukemia detection accuracy via ensemble learning and pre-trained CNN models. By using blood microscopic images, this study addresses the challenge of early-stage leukemia detection, offering solutions such as image preprocessing techniques, data augmentation to address dataset imbalance, and the use of various pre-trained CNN models, including ResNet50, VGG16, and InceptionV3 for feature extraction and classification. The Raabin Leukemia dataset includes 1800 microscopic images across different leukemia classes. The dataset is split into training, testing, and validation sets. The proposed method involves ensemble learning, combining models to enhance predictive performance, achieving an impressive accuracy of 99.8%. While individual models show accuracy ranging from 90% to 92.5%, the approach outperforms them, albeit requiring a longer execution time.

Mustafa Ghaderzadeh et al. [42] proposed an automated model for classifying B-lymphoblast cells from normal B-lymphoid precursors in blood smear images using a majority voting ensemble technique. This study used the C-NMC 2019 dataset, including 12,528 lymphocyte nucleus images (8491 B-ALL and 4037 normal cases), to train and evaluate the model. The dataset was preprocessed using normalization, edge enhancement, and data augmentation techniques to improve feature extraction. Eight pre-trained CNN models were tested, and four (DenseNet121, Inception V3, Inception-ResNet-v2, and Xception) were selected to construct an ensemble classifier. By combining predictions from these models, the ensemble achieved a sensitivity of 99.4%, specificity of 96.7%, and accuracy of 98.5%. This model offers high precision in distinguishing cancerous cells from normal ones and can be implemented in digital laboratory equipment, aiding faster and more reliable leukemia diagnosis.

4. Analysis of the Reviewed Studies

Our analysis is structured in three sections and aims to evaluate leukemia diagnosis using deep learning techniques:

- Firstly, we covered the classification analysis based on the literature review (See Figure 4).
- The second part handles dataset analysis, preprocessing, DL model, and evaluations.
- The third analysis is about the advantages and disadvantages based on the reviewed literature.
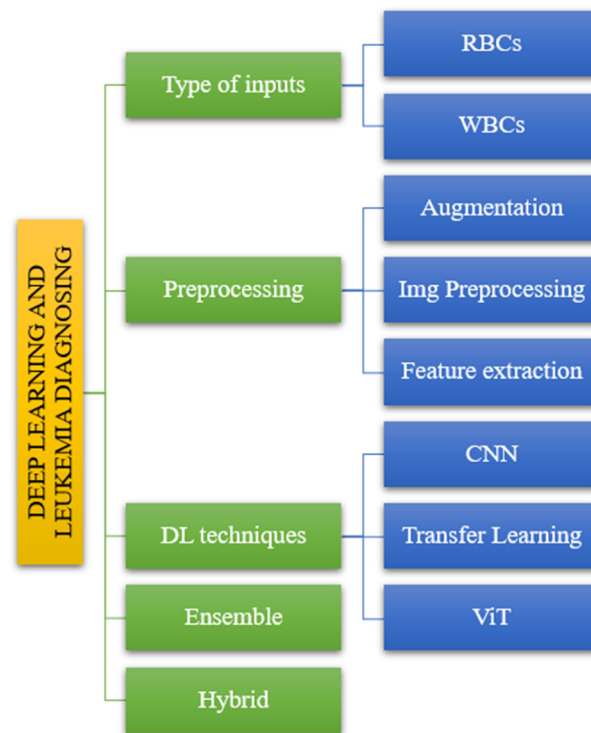
Figure 4. The taxonomy of leukemia classification based on reviewed papers.

4.1. Summary of Taxonomy Based on Literature Review

Following the taxonomy delineated in the introductory section (Figure 2), we systematically evaluated the articles to elucidate their alignment with the classification depicted in Figure 4. The reviewed literature underwent classification based on distinct parameters, including input types, processing methodologies, DL techniques, and ensemble and hybrid approaches, as illustrated in Figure 4. This categorization facilitated a structured analysis of the articles, offering insights into the diverse methods and methodologies employed in the domain of interest.

4.2. Summary of Primary Studies Based on Literature Review

A comprehensive examination of seminal studies on leukemia classification leveraging deep learning methodologies has been meticulously conducted. The synthesized insights from this review are briefly presented in Appendix A, encompassing critical analyses of datasets, preprocessing techniques, model architecture, and evaluation. A distinctive feature of this research is identifying all datasets and model architecture that successfully met the predefined study objectives. We refer readers to the dedicated Results Section for an in-depth exploration of the datasets and algorithms, which provides a more comprehensive elucidation.

4.3. Summary of Advantages and Disadvantages Based on Literature Review

Following a thorough literature review, we succinctly encapsulated the objectives, advantages, and disadvantages in Appendix B.

5. Results

This section answers the research questions after thoroughly analyzing the selected papers.

5.1. Answer to Question RQ1

Searches across the databases specified in the search process section identified 212 relevant articles; three were systematic review papers, and no systematic mapping study (SMS)

papers were encountered. The content, methodology, and full text of the identified systematic literature review (SLR) papers were thoroughly examined, revealing no similarities with our manuscript. A synopsis of the identified systematic literature reviews (SLRs) can be found in the table below (Table 3).

Table 3. An overview of the identified systematic literature reviews (SLRs).

N°	Ref	Year	Title	Main Idea
1	[43]	2022	A Systematic Review on Acute Leukemia Detection Using Deep Learning Techniques	This systematic review aims to explore the use of deep learning techniques for detecting and classifying acute leukemia. By analyzing selected publications, this review addresses four research questions regarding the effectiveness of deep learning in leukemia classification.
2	[44]	2023	Deep Learning for the Detection of Acute Lymphoblastic Leukemia Subtypes on Microscopic Images: A Systematic Literature Review	This study reviews advancements in detecting and classifying Acute Lymphoblastic Leukemia (ALL) subtypes. The discussion primarily centers around using deep learning techniques for object detection and classification in this domain.
3	[45]	2023	Acute Lymphoblastic Leukemia Detection Challenges and Systematic Review	This paper provides a literature review on detecting Acute Lymphoblastic Leukemia using various development approaches. It highlights existing studies on leukemia detection using machine or deep learning, performance analysis, and associated challenges.

5.2. Answer to Question RQ2

To answer this question, we reviewed the databases used in the 30 articles of this study. As depicted in Appendix A, an exhaustive examination of the datasets presented in the articles was conducted. The most optimal dataset is visually represented in Figure 5, as Appendix A indicates.
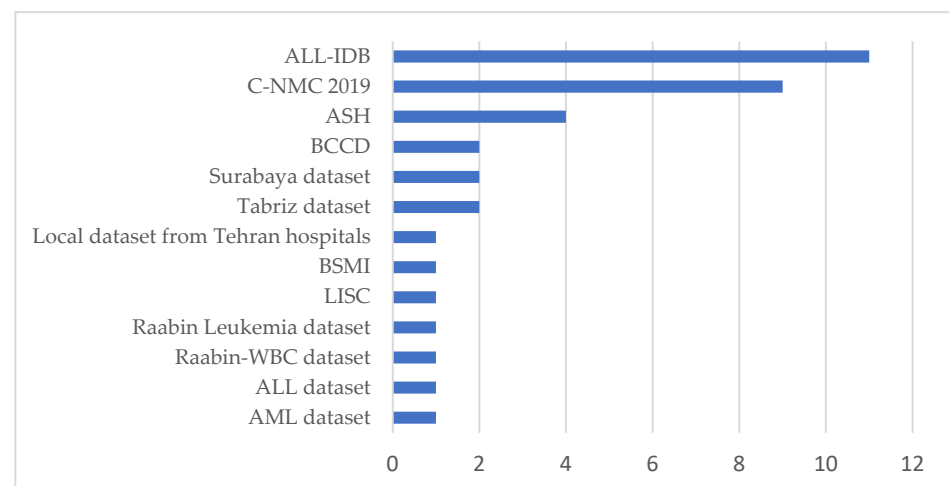

Figure 5. Dataset usage for selected studies.

From Figure 5, it can be observed that within the selected studies, the datasets commonly used in leukemia classification research are ALL-IDB, C-NMC 2019, and the ASH Image Bank. These datasets provide critical resources but also present significant limitations that challenge their application in deep learning. For instance, The ALL-IDB dataset [11] is specifically designed for Acute Lymphoblastic Leukemia (ALL) research and contains annotated microscopic blood cell images. It is widely used for segmentation and classification tasks due to its targeted focus and expert annotations. However, the dataset's scope

is narrow, as it only includes ALL cases and healthy samples, omitting other leukemia types such as Acute Myeloid Leukemia (AML). Its small sample size further limits its utility for deep learning, requiring large datasets to train complex architectures effectively. Moreover, the dataset lacks variability in image acquisition conditions, such as differences in staining, illumination, and noise, reducing its ability to generalize to real-world clinical settings. These shortcomings highlight the need for more comprehensive and diverse datasets representing a broader spectrum of leukemia cases and imaging conditions.

On the other hand, the ASH Image Bank [12] provides a publicly available repository of hematological images and is often used as a supplementary dataset in leukemia research. Its comprehensive nature, covering a wide array of hematological topics, makes it a valuable resource for training and validation. However, this dataset lacks the standardization required for specific tasks such as leukemia classification. Images vary in quality, resolution, and format, complicating their direct use in deep learning pipelines. Furthermore, inconsistent and incomplete annotations limit its utility for supervised learning, where precise labels are critical for model performance. These factors necessitate preprocessing steps to normalize the images and detailed re-annotation efforts to align the data with the requirements of deep learning techniques.

The C-NMC 2019 dataset [13] offers a larger collection of segmented microscopic images, comprising 15,135 images from 118 patients. It includes real-world noise, such as staining imperfections and illumination errors, making it more representative of clinical scenarios. Additionally, the dataset is annotated by expert oncologists, ensuring high accuracy in its ground truth labels. Despite these strengths, the C-NMC 2019 dataset faces significant challenges. The dataset exhibits class imbalance, primarily focusing on two categories: normal cells and leukemia blasts, leading to biases in model training. Furthermore, the morphological similarity between these two classes makes feature extraction and classification particularly difficult for deep learning models. Limited patient diversity, with images derived from a relatively small number of individuals, further restricts the generalization of models trained on this dataset. Addressing these issues requires datasets with more balanced class distributions and increased patient diversity to ensure broader applicability.

These datasets collectively highlight the pressing need for more robust data resources for deep learning applications. The key requirements include larger sample sizes to train deep learning architectures effectively, greater diversity in patient demographics and disease subtypes to improve generalization, and detailed and consistent annotations. Additionally, incorporating real-world variability in imaging conditions, such as noise, lighting, and staining differences, can enhance the robustness of models in clinical settings. Advanced techniques like data augmentation can artificially expand datasets by introducing variations such as rotations, flips, and brightness adjustments, thereby mitigating the issue of limited data. Synthetic data generation using methods like Generative Adversarial Networks (GANs) can also create realistic images to address dataset scarcity. Furthermore, integrating multiple datasets through cross-dataset training can enhance model performance by exposing it to diverse data sources. Researchers can significantly improve models' accuracy, robustness, and applicability in leukemia diagnosis by addressing the underlined limitation in combination with advanced deep learning techniques. The most suitable or effective dataset for leukemia classification using deep learning may vary depending on data quality, diversity, and relevance to clinical scenarios. However, the ALL-IDB dataset is often cited in the literature due to its specific focus on leukemia and its established use in research studies.

5.3. Answer to Question RQ3

Researchers preprocess blood images using various techniques before applying deep learning algorithms for leukemia classification. These preprocessing methods aim to enhance image quality, reduce noise, and improve feature extraction, ultimately improving classification accuracy. Some commonly employed preprocessing methods are as follows:

- Data Augmentation: Techniques such as rotation, translation, flipping, zooming, and shearing increase the diversity of the dataset, which helps prevent overfitting and improve model generalization.
- Color Space Conversion: Converting images from RGB to other color spaces like YCbCr or Lab* can enhance image quality and make segmentation easier.
- Image Resizing: Resizing images to a standardized size helps reduce computational complexity and ensure consistency across the dataset.
- Segmentation: Segmenting images to isolate specific components, such as cells or nuclei, can aid in feature extraction and improve classification accuracy.
- Thresholding: Techniques like Otsu Adaptive Thresholding separate objects from the background by finding an optimal threshold value.
- Filtering: Gaussian or Laplacian filters help reduce noise and enhance image clarity.
- Normalization: Normalizing pixel values helps standardize the input data and improve convergence during model training.
- Feature Extraction: Extracting features using methods like the Gray Level Co-occurrence Matrix (GLCM) helps capture texture information, which is useful for distinguishing between different cell types.

The effectiveness of these preprocessing methods in improving classification accuracy varies depending on the dataset and the specific characteristics of the images. Comparative analyses and the experimental results in the reviewed literature suggest that a combination of techniques, including data augmentation, color space conversion, feature extraction, and segmentation, often yield the best results in terms of classification accuracy. Additionally, employing domain-specific preprocessing methods tailored to the characteristics of blood cell images can further enhance model performance.

5.4. Answer to Question RQ4

The reviewed papers used several deep learning models for leukemia diagnosis, each with unique strengths and performance characteristics. The choice of the model often depends on the dataset, the amount of data available, and the specific requirements of the classification task. Some of the deep learning models and approaches for leukemia classification are as follows:

- Convolutional Neural Networks (CNNs): CNNs have been extensively used in medical image analysis, including leukemia classification. Their ability to automatically learn hierarchical features from images makes them well suited for this task. The types of CNNs used in reviewed papers include but are not limited to DCNN (DenseNet121), Sequential CNN, Neural Network, DCNN with squeeze-and-excitation learning, CNFN: fuzzy interface, Customized CNN, and Type II fuzzy CNN.
- Transfer Learning: Pre-trained models can be fine-tuned for leukemia classification tasks, especially those trained on large datasets like ImageNet. This approach leverages the knowledge gained from diverse datasets. The selected papers identified the following transfer learning techniques: VGG, MobileNet, AlexNet, GoogleNet, EfficientNet, ResNet, Inception, ShuffleNet, and QCResNet.
- Vision Transformer (ViT): ViT is a relatively newer architecture that has succeeded in various computer vision tasks. Its attention mechanism lets it capture long-range image dependencies, making it promising for leukemia classification.

- Ensemble Models: Combining predictions from multiple models, such as an ensemble of CNNs or a combination of different architectures, has enhanced overall classification performance.
- Hybrid Models: Hybrid models offer flexibility in leveraging the strengths of different deep learning architectures or approaches, potentially leading to improved classification accuracy and generalization capability. However, designing and training hybrid models requires careful consideration of model architecture, parameter tuning, and computational resources.

While CNNs remain the dominant choice for leukemia classification, recent research has explored various architectures and approaches, including transfer learning, transformers with attention mechanisms, hybrid models, and ensemble methods, to improve classification accuracy and robustness further, as illustrated in Figure 6 below. The choice of model architecture depends on factors such as the size and complexity of the dataset, computational resources, and the specific requirements of the classification task.
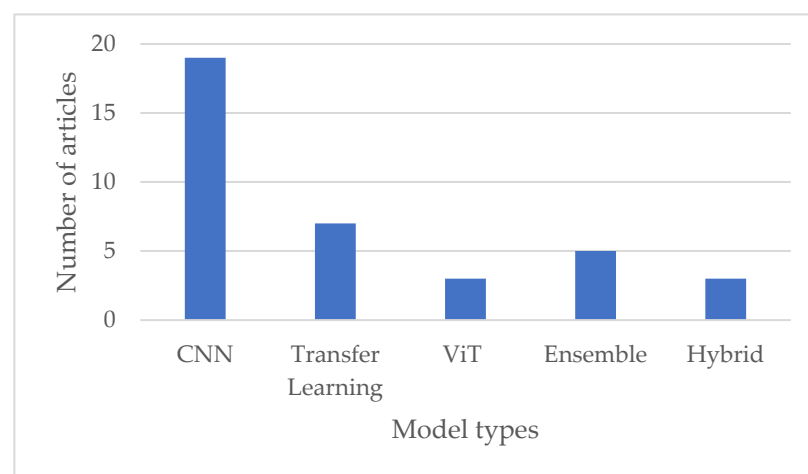

Figure 6. Comparison of various deep learning techniques used in leukemia diagnosis.

5.5. Answer to Question RQ5

The selected studies on deep learning models for leukemia diagnosis have identified several challenges researchers face in achieving high accuracy and interpretability. These challenges and proposed future directions are pivotal for enhancing the accuracy and interpretability of deep learning models in leukemia diagnosis. Here is an overview:

- Limited Datasets: Many studies face challenges due to the scarcity of annotated leukemia images, leading to small and imbalanced datasets. This hampers the ability of models to generalize across diverse cases.
- Class Imbalance: Class imbalance, where certain leukemia subtypes are underrepresented in the dataset, can bias the model towards the majority class and lead to suboptimal performance for minority classes.
- Interpretability and Explainability: Deep learning models, particularly complex architectures like CNNs and transformers, are often considered "black boxes." Interpreting and explaining the decision-making process of these models remains challenging, raising concerns in clinical applications where transparency is crucial.
- Generalization to Unseen Cases: Ensuring the robustness and generalization of deep learning models to handle unseen or rare leukemia subtypes is a persistent challenge. Models need to be capable of adapting to variations in image quality and patient demographics.

- Integration with Clinical Workflow: Deploying deep learning models into clinical practice requires seamless integration with clinical workflows and electronic health record systems. Ensuring the usability and practicality of the models is essential for their adoption by healthcare professionals.
- Ethical and Regulatory Considerations: As deep learning models progress towards clinical deployment, ethical and regulatory aspects become increasingly important. Critical challenges include ensuring patient privacy, model fairness, and compliance with medical regulations.

To address these challenges and further improve the accuracy and interpretability of deep learning models in leukemia diagnosis, several future research directions are suggested:

- Data Augmentation and Synthesis: Techniques such as data augmentation and synthesis can help address the limited data availability by generating synthetic images or augmenting existing data to increase the diversity and size of the dataset.
- Attention Mechanisms and Explainable AI: Integrating attention mechanisms and developing explainable AI techniques can enhance the interpretability of deep learning models, allowing clinicians to understand the model's decision-making process and trust its predictions.
- Transfer Learning and Domain Adaptation: Further exploration of transfer learning and domain adaptation methods can improve generalization across different datasets and unseen cases.
- Ensemble and Multi-Modal Approaches: Ensemble learning and multi-modal approaches, which combine information from multiple sources or models, can improve classification accuracy and robustness by leveraging complementary information.
- Hybrid Models and Integration with Clinical Data: Investigating hybrid models that combine imaging data with clinical information can provide a more comprehensive understanding of leukemia cases. Developing effective frameworks for integrating diverse data sources is essential. Incorporating AI into healthcare faces technical and organizational barriers. Key challenges are ensuring interoperability with hospital systems, addressing data variability, and overcoming infrastructure limitations. High computational demands and the need for skilled personnel hinder adoption, particularly in resource-limited settings. Organizational resistance and legal uncertainties regarding liability further complicate implementation. Addressing these issues requires standardized data protocols, robust quality assurance, and fostering collaboration between technologists and clinicians to align AI solutions with clinical needs effectively.
- Ethical AI and Regulatory Frameworks: Ethical AI in healthcare requires transparency, fairness, accountability, and explainability. Explainable AI (XAI) is critical for fostering clinician trust by enabling them to understand and validate AI predictions, especially in sensitive applications like leukemia diagnosis. Addressing biases in datasets is equally essential to prevent inequitable outcomes, ensuring that AI systems are fair and representative across diverse populations. Regulatory frameworks must evolve to include standards for model validation, periodic reassessment, and compliance with data protection laws like GDPR (General Data Protection Regulation). Collaboration between regulators, healthcare professionals, and developers is crucial to align innovation with ethical integrity, promote the adoption of XAI, and safeguard patient safety [46].
- Patient-Centric Approaches: Future research should prioritize patient-centric approaches, considering individual variations and tailoring models to specific patient demographics and characteristics.

Addressing these challenges and pursuing these research directions enables deep learning for leukemia diagnosis to advance, ultimately resulting in more accurate, interpretable, and clinically impactful models that improve patient care.

6. Conclusions

This paper provides insights from a systematic mapping study (SMS) and a systematic literature review (SLR) focused on deep learning techniques for leukemia diagnosis. By analyzing thirty studies, the review highlights a variety of methodologies that demonstrate promising levels of accuracy and sensitivity in detecting leukemia from blood smear images. The findings emphasize the importance of ongoing research to overcome current challenges and enhance the development of deep learning models that are accurate, reliable, and interpretable. Such advancements are essential for improving patient care and supporting informed decision-making in clinical settings. Integrating large language models (LLMs) into healthcare systems holds great potential to further transform diagnostics and treatment. LLMs' ability to adapt to evolving data and contexts can drive advancements in disease prediction, personalized care, and operational efficiency. However, addressing challenges like dataset limitations, ethical concerns, and practical implementation hurdles is critical to successfully adopting these technologies. With continued innovation and interdisciplinary collaboration, AI and machine learning can significantly reshape healthcare delivery, leading to improved outcomes and more equitable access to care.

Author Contributions: Conceptualization, R.F.O.K., T.P.T.A. and H.-C.K. methodology, R.F.O.K. and T.P.T.A.; validation, H.-C.K. formal analysis, R.F.O.K. and T.P.T.A.; writing—original draft preparation, R.F.O.K. and T.P.T.A.; writing—review and editing, R.F.O.K. and T.P.T.A.; supervision, H.-C.K. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

AUC	Area Under the Curve
CAS	Computer-Aided System
CBS	Complete Blood Smear
CNFN	Convolutional Neuro-Fuzzy Network
COVID-19	An Infectious Disease Caused by the SARS-CoV-2 Virus
DCNN	Deep Convolutional Neural Network
DL	Deep Learning
DNA	Deoxyribonucleic Acid
DT	Decision Tree
FCH	Fuzzy Color Histogram
GAN	Generative Adversarial Network
GFLOPS	Gigaflops
IoMT	Internet of Medical Things
IoT	Internet of Things
K-NN	K-Nearest Neighbor
LBP	Local Binary Pattern
LD	Linear Discriminant
LeakyReLU	Leaky Rectified Linear Unit
MAP	Mean Average Precision
ML	Machine Learning
RBC	Red Blood Cell

R-CNN	Region-based Convolutional Neural Network
ReLU	Rectified Linear Unit
RGB	Red, Green, Blue
RNA	Ribonucleic Acid
ROC-AUC	Receiver Operating Characteristic—Area Under the Curve
SBILab	Laboratory Name
SGD	Stochastic Gradient Descent
SVM	Support Vector Machine
SVM-Cubic	SVM Cubic Kernel
VGG	Visual Geometry Group
WSDAN	Weakly Supervised Data Augmentation Network
YCbCr	Luminance (Y), Blue-difference Chroma (Cb), and Red-difference Chroma (Cr)

Appendix A. An Overview of Existing Primary Studies

Ref	Year	Dataset	Preprocessing	Model	Evaluation
[7]	2019	ALL-IDB ASH	Data aug: Rotation, height shift, width shift, zoom, horizontal flip, vertical flip, and shearing	CNN Optimizations: SGD and ADAM	88.25% acc binary class. 81.74% acc multi-class classification
[10]	2019	BSMI	Preprocessing: Removing the noise, enhancing the features Segmentation: Gaussian Distribution and K-Means clustering Feature extract: GLCM	CNN	97.3% acc
[12]	2019	C-NMC 2019	Normalization: RGB, ImageNet Resizing: 380×380 Augmentation: Contrast adjustments, brightness correction, horizontal and vertical flips, and intensity adjustments	- Feature extraction: VGG16, MobileNet Hybrid CNN: VGG+ MobileNet	96.17% acc
[13]	2020	BCCD ASH	Transformation: Convert to RGB Resize: 227×227 Augmentation: Translation, reflection, and rotation. Feature extract: AlexNet	Classification: SVM, DT, LD, K-NN Feature extraction fine-tuned AlexNet.	Case1: 99.79% acc Case 2: 100% acc
[15]	2020	C-NMC 2019	Normalization: Stain norm. Resizing: $224 \times 224 \times 3$ Augmentation: Rademacher Mixup	MMA-MTL combines GAC and LAC, PAC, LSA	F1 = 0.9189
[16]	2020	ALL_IBD2, LISC	Not mentioned	AlexNet, GoogleNet, and VGG-16	IDB_2—96.15% acc LISC—80.82%
[18]	2020	ALL_IBD	Augmentation: Rotation, brightness adjustment, contrast adjustment, shearing, horizontal and vertical flipping, translation, and zooming	CNN	95.45% acc
[19]	2020	BCCD	Augmentation and Balancing Resize: 120×160	CNN, Adam optimization	98.31% acc

Ref	Year	Dataset	Preprocessing	Model	Evaluation
[20]	2020	ALL-IDB ASH	Augmentation: Rotation, height shift, width shift, horizontal flip, zoom, and shearing	DCNN: DenseNet-121, ResNet-34	100% acc
[21]	2021	ALL dataset	Augmentation: Horizontal and vertical flipping, random rotational transformations	Sequential CNN	ROC-AUC of 0.97 ± 0.02
[22]	2021	ALL-IDB	Augmentation Resize: $224 \times 224 \times 3$	ShuffleNet	ALLIDB1: 96.97% acc ALLIDB2: 96.67% acc
[23]	2021	C-NMC 2019	Normalize: Difference enhancement-random sampling (DERS) Resize: 224×224	Ensemble Model: ViT, CNN (EfficientNet) Optimizer: Adam	99.03% acc
[24]	2021	ALL-IDB	Augmentation	Hybrid transfer learning: MobileNetV2 and ResNet18	ALLIDB1: 99.39% acc ALLIDB2: 97.18% acc
[25]	2021	C-NMC 2019	Augmentation: Mirroring, rotation, shearing, blurring, salt-and-pepper noise	Neural Network	F1 = 91.2%
[26]	2021	Local dataset from Tehran hospitals (Iran)	Decoding, resizing to 224×224, segmentation using LAB color space and K-means clustering, data augmentation (vertical and horizontal flips), and normalization.	Lightweight CNN models (EfficientNetB0, MobileNetV2, NASNet Mobile)	100% accuracy, sensitivity, and specificity for B-ALL detection and classification
[27]	2022	ALL-IDB	Noise reduction: Average and Laplacian filters Segmentation: Adaptive Region-Groving Algorithm Feature extract: LBP, GLCM, FCH, CNN Augmentation: Rotation, flipping, cropping, displacement	(1): ANN, FFNN, SVM (2) CNN: AlexNet, GoogleNet, ResNet-18 (3) Hybrid CNN-SVM	FFNN—100% GoogleNet—100% ResNet18+ SVM—100%
[28]	2022	ALL-IDB	Augmentation: Rotation and random shifts	deep CNN architecture with squeeze-and-excitation learning	IDB1—100% IDB2—99.98% IDB1+ IDB2—99.33
[29]	2022	ALL-IDB	Augmentation: Vertical and horizontal flips and image rotation. Img. Prep: Noise reduction: selective filtering, unsharp masking Transformation: RGB to L*a*b Resize: 64×64 Segmentation: Color-based clustering, followed by the application of the bounding box method	Convolutional neuro-fuzzy network: fuzzy inference, Takagi-Sugeno-Kang (TSK) fuzzy model	97.31%

Ref	Year	Dataset	Preprocessing	Model	Evaluation
[30]	2022	Surabaya dataset	Labeling: YOLO format Resizing: 416 × 416	GhostNet convolution module into the YOLO 4 and 5	F1 = 86.1 to 90
[32]	2022	ALL_IBD	Clustering: K-Means Segmentation: modified Residual U-Net architecture Augmentation: Flipping, rotating, shifting, and scaling	Multistage transfer learning: InceptionV3, Xception, InceptionResNetV2	Detection: 99.60% Classification: 94.67%
[33]	2022	Surabaya dataset	Augmentation: Rotation, zoom, flip	Detection: YOLO 4, 5 Segmentation: Mask R-CNN	F1 = 89.5% maP = 93.2%
[34]	2022	C-NMC 2019	Resize: 300 × 300 × 3 Augmentation: Horizontal flipping and resizing	Inception V3	99% acc.
[35]	2023	Tabriz dataset: 184 (ALL) 469 (AML)	Img. Prep: Resize: 224 × 224, grayscale, normalization, Augmentation: Using GAN	Customized CNN	99.5% acc.
[36]	2023	Tabriz dataset: 184 (ALL) 469 (AML)	Img. Prep: Resize: 224 × 224, grayscale, normalization, Augmentation: Rotation, horizontal and vertical translation	Type-II Fuzzy DCNN	98.8% acc.
[37]	2023	C-NMC 2019	Resize and balance, Gamma transforms, shuffle and split	QCResNet modified ResNet-18, trained with Adam	98.9%acc.
[38]	2023	ALL_IDB ASH C_NMC 2019	ResNet18 for feature extraction	Orthogonal SoftMax Layer (OSL)-Based on Classification	99.39% acc. ALLIDB1, 98.21% acc. ALLIDB2, 97.50% acc. ASH
[39]	2023	Raabin-WBC	Resize: 224 × 224	ViT	99.40%
[40]	2023	C-NMC 2019	Img. Prep: Normalization, noise reduction, resolution standardization, augmentation	ResNet, InceptionNet, MobileNet, EfficientNet	Accuracies: ResNet 95.75% Inception 95.37% MobileNet 94.81% EfficientNet 96.44%
[41]	2023	Raabin Leukemia	Augmentation Img. Prep: Rescaling, brightness adaptation, and discrimination.	Ensemble: VGG16 + ResNet50+, InceptionV3	Accuracies: Ensemble 99.8%, VGG16 98%, ResNet50 90%, InceptionV3 92.5%
[42]	2023	C-NMC 2019	Normalization, resizing to 300 × 300 pixels, edge enhancement, data augmentation with 16 techniques, and image standardization	Ensemble of four CNN models (DenseNet121, Inception V3, Inception-ResNet-v2, and Xception) using a majority voting technique	Sensitivity of 99.4%, specificity of 96.7%, accuracy of 98.5%

Appendix B. The Comparison of Articles Related to the Leukemia Diagnosis Based on DL

Ref	Objective	Advantages	Disadvantages
[7]	Accurately finding different subtypes of leukemia using CNN	Data augmentation Comparison with other models	Limited dataset. Evaluation variability. Generalization to real-world settings. Noise.
[10]	Automated diagnosis of white blood cell cancer diseases	Accurate diagnosis Robust segmentation Efficient feature extraction	Limited dataset. Dependency on image quality. Complexity of implementation
[12]	Automatic leukemic B-lymphoblast classification using a DL-based hybrid method	High accuracy Hybrid approach Addressing limitations	Limited dataset. Computational complexity. Evaluation on a single dataset.
[13]	Automated leukemia detection, using deep learning techniques and transfer learning	Utilizes transfer learning High accuracy Data augmentation comparative analysis	Limited dataset. Lack of detailed analysis.
[15]	Reducing false negatives in the classification of ALL cells	Detecting early-stage leukemia Enhancing the model's ability to discern morphologically similar cells Effective regularization technique	It does not offer detailed insights into the computational complexity and scalability of model and acknowledges the challenge of comparing results due to the held-out test set not being public.
[16]	CAS using CNN for automated WBC detection and classification, particularly for finding ALL	Achieves high accuracy in classifying WBC types Comparison of pre-trained models provides insights into model performance	Misclassifications still occur, particularly for certain WBC types. Limited discussion on preprocessing techniques and model hyperparameters.
[18]	Automated diagnostic system to detect ALL using a CNN	High accuracy Augmentation techniques	Limited discussion on potential challenges and limitations. Dependency on the specific characteristics of the ALL_IDB dataset.
[19]	An automated approach using a lightweight CNN for WBC classification	High-accuracy automated approach reduces the time and effort required for WBC Augmentation and balancing techniques	Limited discussion on potential challenges and limitations. model's performance may vary depending on the dataset and specific application.
[20]	Automated detection and classification of leukemia subtypes using an IoMT framework	Quick, safe, and accurate early-stage diagnosis of leukemia offers the potential for remote diagnosis and medical care, particularly during pandemics like COVID-19	Limited discussion on potential challenges or limitations of the proposed framework.

Ref	Objective	Advantages	Disadvantages
[21]	Analyzing the impact of training set size on CNN performance	Achieve high classification performance even with small training datasets	Primarily focuses on classification performance and does not extensively discuss the generalizability of the proposed method across different datasets or medical imaging tasks. Does not provide detailed information on preprocessing steps or hyperparameter tuning.
[22]	Transfer learning with ShuffleNet	Highly computationally efficient, Outperforming other techniques superior performance over existing methods in terms of various evaluation metrics	Limited discussion on preprocessing techniques.
[23]	Diagnosing ALL using an ensemble model combining ViT and CNN.	High classification accuracy Augmentation and balancing techniques Ensemble model Optimizer	The evaluation is primarily based on accuracy and precision metrics, and more metrics or external validation may offer further insights into the model's performance.
[24]	Aims to address the challenge of achieving correct ALL detection with smaller datasets by using hybrid transfer learning	Excellent performance in ALL detection and classification Achieving accurate results with smaller datasets It offers computational efficiency compared to ensemble transfer learning approaches	While the proposed method generally demonstrates superior performance, there are instances where it delivers slightly poorer performance.
[25]	Effectiveness of a simple multilayer neural network combined with standard image processing feature extraction techniques for lymphocyte classification	Achieves state-of-the-art performance with a fraction of the computational cost required by Convolutional Neural Networks It uses a rich set of features extracted from lymphocyte images, contributing towards accurate classification	While achieving high accuracy, an F1-score of 91.2% may not be accurate enough for disease diagnosis. May still require optimization for real-time applications.
[26]	Develop a lightweight CNN model integrated into a mobile application for real-time diagnosis of B-ALL using blood smear images	Achieved 100% accuracy, sensitivity, and specificity, with deployment in a mobile application for accessible and efficient diagnosis	Limited to a single dataset and may not generalize well to broader datasets or diverse clinical conditions.

Ref	Objective	Advantages	Disadvantages
[27]	Application of both traditional ML and DL models for early detection of leukemia	High classification accuracy Hybrid systems combine the strengths of CNN for feature extraction and SVM for classification	CNN models require a large dataset to avoid overfitting, and this is one of the limitations of this study because the data set is insufficient to train CNN models. The data augmentation technique overcomes this limitation.
[28]	Deep CNN architecture with squeeze-and-excitation learning for diagnosing leukemia	High classification accuracy Utilizes data augmentation to overcome limited dataset size Incorporates squeeze-and-excitation learning to enhance feature learning discriminability	The proposed method may require further validation on diverse datasets to assess its generalizability to different populations or subtypes of leukemia.
[29]	Classification of ALL using a deep convolutional neuro-fuzzy network	Achieves high accuracy Effective use of fuzzy reasoning rules and deep learning for automated prognosis Utilizes data augmentation to address limited training samples	Specific challenges in finding useful characteristics and categorization due to the small dataset.
[30]	It introduces the GhostNet convolution module to replace the standard convolution module in the backbone of YOLOv4 and YOLOv5 models	The modification reduces computational complexity, number of parameters, and GFLOPS values without significant loss in detection accuracy Object detection time is improved, making the model lighter and faster than the original YOLO models	This study's evaluation is based on a specific dataset, which may limit the proposed method's generalizability to other datasets or medical imaging tasks. Further research is needed to explore the robustness and scalability of the modified YOLO models across different medical imaging datasets and applications.
[32]	Multistage transfer learning (MTL) approach for ALL detection and classification	Achieves high accuracy in leukemia detection and classification Utilizes a MTL approach to leverage knowledge from pre-trained models and improve classification performance. Incorporates comprehensive preprocessing	The computational and time-intensive nature of the proposed method may limit its scalability and applicability in real-time clinical settings.

Ref	Objective	Advantages	Disadvantages
[33]	Use of object detection and instance segmentation techniques, specifically comparing YOLO and Mask R-CNN, for the detection of ALL subtypes	YOLOv4 shows superior performance in detecting ALL subtypes compared to YOLOv5 and Mask R-CNN Data augmentation techniques are employed to increase the model's ability to recognize new variant representations and reduce overfitting	The YOLOv4 model, despite its higher performance, still shows recall values and precision below 90%, with some ALL-subtype cell objects detected as false positives or not detected at all. Mask R-CNN shows the lowest performance among the compared models.
[34]	A transfer learning approach based on the Inception V3 architecture for diagnosing leukemia	High accuracy, precision, and F1-score in classifying ALL from microscopic images Image preprocessing technique	Computational resources are needed for training and fine-tuning the model, particularly with complex architectures like Inception V3.
[35]	DL model for the automatic diagnosis of acute leukemia using images of lymphocytes and monocytes	High accuracy, novel dataset and GAN augmentation, Tversky loss function improves diagnosis	Limited dataset.
[36]	A type-II fuzzy deep network for feature extraction and classification	High-accuracy architecture improves feature extraction and classification accuracy	The computational efficiency of the type-II fuzzy function is lower compared to Relu and Leaky Relu.
[37]	QCResNet for the classification of ALL	High accuracy, converges faster compared to other models	Limitations of the dataset itself may affect the generalizability of the model.
[38]	Classifying and detecting leukemia using DL, specifically addressing the challenges of overfitting with small medical image datasets	Faster testing time, ResNet18 and OSL enhance classification performance and computational efficiency	It does not discuss specific preprocessing techniques.
[39]	Automate the classification of WBCs from blood films using the ViT model to capture long-range dependencies and global context.	Achieves high accuracy with low training cost due to pre-training on large datasets End-to-end transformer structure eliminates the need for feature engineering. Potential for clinical applications due to high classification accuracy	The model's resilience to image variations was not verified. Computational and memory requirements are higher than those of other models.

Ref	Objective	Advantages	Disadvantages
[40]	Analysis of ALL Detection Methods Using DL	Each model has its unique strengths, such as robust feature learning in ResNet, efficient multi-scale feature extraction in InceptionNet, reduced computational cost in MobileNet, and a favorable tradeoff between accuracy and efficiency in EfficientNet	Computational demands, potential overfitting, longer training times, and increased complexity.
[41]	To improve the accuracy and efficiency of leukemia detection using ensemble learning and pre-trained CNN models on blood microscopic images	High accuracy, detailed methodology, and experimental results provided for reproducibility and validation, Ensemble model using pre-trained models	Require longer execution time compared to individual models. Small dataset. Reliance on pre-trained models may limit adaptability to specific dataset characteristics.
[42]	Classify B-lymphoblast cells from normal B-lymphoid precursors using an ensemble CNN model with majority voting.	High accuracy (98.5%) and robust performance for distinguishing cancerous and normal cells	It is computationally expensive due to multiple CNN models in the ensemble.

References

1. Huang, J.; Chan, S.C.; Ngai, C.H.; Lok, V.; Zhang, L.; Lucero-Prisno, D.E., III; Xu, W.; Zheng, Z.J.; Elcarte, E.; Withers, M.; et al. Disease burden, risk factors, and trends of leukaemia: A global analysis. *Front. Oncol.* **2022**, *12*, 904292. [CrossRef] [PubMed]

2. MSD Manual. Acute Lymphoblastic Leukemia (ALL). 2023. Available online: https://www.msdmanuals.com/professional/hematology-and-oncology/leukemias/acute-lymphoblastic-leukemia-all?ruleredirectid=745 (accessed on 8 November 2024).

3. Medicine, Y. Diagnosing Leukemia. Available online: https://www.yalemedicine.org/conditions/leukemia-diagnosis (accessed on 10 November 2024).

4. Cruz-Benito, J. Systematic Literature Review & Mapping. 2016. Available online: https://zenodo.org/records/165773 (accessed on 10 November 2024).

5. Theodore Armand, T.P.; Nfor, K.A.; Kim, J.I.; Kim, H.C. Applications of Artificial Intelligence, Machine Learning, and Deep Learning in Nutrition: A Systematic Review. *Nutrients* **2024**, *16*, 1073. [CrossRef] [PubMed]

6. Hosseini, S.H.; Monsefi, R.; Shadroo, S. Deep learning applications for lung cancer diagnosis: A systematic review. *Multimed. Tools Appl.* **2024**, *83*, 14305–14335. [CrossRef]

7. Ahmed, N.; Yigit, A.; Isik, Z.; Alpkocak, A. Identification of Leukemia Subtypes from Microscopic Images Using Convolutional Neural Network. *Diagnostics* **2019**, *9*, 104. [CrossRef]

8. Labati, R.D.; Piuri, V.; Scotti, F. All-IDB: The acute lymphoblastic leukemia image database for image processing. In Proceedings of the 2011 18th IEEE International Conference on Image Processing, Brussels, Belgium, 11–14 September 2011; pp. 2045–2048. [CrossRef]

9. The American Society of Hematology. Available online: http://www.hematology.org (accessed on 15 October 2024).

10. Agrawal, R.; Satapathy, S.; Bagla, G.; Rajakumar, K. Detection of White Blood Cell Cancer using Image Processing. In Proceedings of the 2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN), Vellore, India, 30–31 March 2019; pp. 1–6. [CrossRef]

11. Gupta, A.; Gupta, R. ALL Challenge Dataset of ISBI 2019 [Data Set]. The Cancer Imaging Archive. 2019. Available online: https://www.cancerimagingarchive.net/collection/c-nmc-2019/ (accessed on 10 November 2024).

12. Kassani, S.H.; Kassani, P.H.; Wesolowski, M.J.; Schneider, K.A.; Deters, R. A Hybrid Deep Learning Architecture for Leukemic B-lymphoblast Classification. In Proceedings of the 2019 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, Republic of Korea, 16–18 October 2019; pp. 271–276. [CrossRef]

13. Loey, M.; Naman, M.; Zayed, H. Deep Transfer Learning in Diagnosing Leukemia in Blood Cells. *Computers* **2020**, *9*, 29. [CrossRef]

14. Blood Cell Images. Available online: www.kaggle.com/paultimothymooney/blood-cells (accessed on 19 October 2024).

15. Mathur, P.; Piplani, M.; Sawhney, R.; Jindal, A.; Shah, R.R. Mixup Multi-Attention Multi-Tasking Model for Early-Stage Leukemia Identification. In Proceedings of the ICASSP 2020—2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–8 May 2020; pp. 1045–1049. [CrossRef]

16. Safuan, S.N.M.; Tomari, M.R.M.; Zakaria, W.N.W.; Othman, N.; Suriani, N.S. Computer Aided System (CAS) of Lymphoblast Classification for Acute Lymphoblastic Leukemia (ALL) Detection Using Various Pre-Trained Models. In Proceedings of the 2020 IEEE Student Conference on Research and Development (SCOReD), Batu Pahat, Malaysia, 27–29 September 2020; pp. 411–415. [CrossRef]

17. Rezatofighi, S.H.; Soltanian-Zadeh, H. Automatic recognition of five types of white blood cells in peripheral blood. *Comput. Med. Imaging Graph.* **2011**, *35*, 333–343. [CrossRef]

18. Anwar, S.; Alam, A. A convolutional neural network–Based learning approach to acute lymphoblastic leukaemia detection with automated feature extraction. *Med. Biol. Eng. Comput.* **2020**, *58*, 3113–3121. [CrossRef]

19. Ridoy, M.A.R.; Islam, M.R. An Automated Approach to White Blood Cell Classification Using a Lightweight Convolutional Neural Network. In Proceedings of the 2020 2nd International Conference on Advanced Information and Communication Technology (ICAICT), Dhaka, Bangladesh, 28–29 November 2020; pp. 480–483. [CrossRef]

20. Bibi, N.; Sikandar, M.; Ud Din, I.; Almogren, A.; Ali, S. IoMT-Based Automated Detection and Classification of Leukemia Using Deep Learning. *J. Healthc. Eng.* **2020**, *2020*, 6648574. [CrossRef]

21. Schouten, J.P.; Matek, C.; Jacobs, L.F.; Buck, M.C.; Bošnački, D.; Marr, C. Tens of images can suffice to train neural networks for malignant leukocyte detection. *Sci. Rep.* **2021**, *11*, 7995. [CrossRef]

22. Das, P.K.; Meher, S. Transfer Learning-Based Automatic Detection of Acute Lymphocytic Leukemia. In Proceedings of the 2021 National Conference on Communications (NCC), Kanpur, India, 27–30 July 2021; pp. 1–6. [CrossRef]

23. Jiang, Z.; Dong, Z.; Wang, L.; Jiang, W. Method for Diagnosis of Acute Lymphoblastic Leukemia Based on ViT-CNN Ensemble Model. *Comput. Intell. Neurosci.* **2021**, *2021*, 7529893. [CrossRef]

24. Das, P.K.; Meher, S. An efficient deep Convolutional Neural Network based detection and classification of Acute Lymphoblastic Leukemia. *Expert Syst. Appl.* **2021**, *183*, 115311. [CrossRef]

25. De Sant, Y.F.D.; De Oliveira, J.E.M.; Dantas, D.O. Lightweight Classification of Normal Versus Leukemic Cells Using Feature Extraction. In Proceedings of the 2021 IEEE Symposium on Computers and Communications (ISCC), Athens, Greece, 5–8 September 2021; pp. 1–7. [CrossRef]

26. Hosseini, A.; Eshraghi, M.A.; Taami, T.; Sadeghsalehi, H.; Hoseinzadeh, Z.; Ghaderzadeh, M.; Rafiee, M. A mobile application based on efficient lightweight CNN model for classification of B-ALL cancer from non-cancerous cells: A design and implementation study. *Inform. Med. Unlocked* **2023**, *39*, 101244. [CrossRef]

27. Abunadi, I.; Senan, E.M. Multi-Method Diagnosis of Blood Microscopic Sample for Early Detection of Acute Lymphoblastic Leukemia Based on Deep Learning and Hybrid Techniques. *Sensors* **2022**, *22*, 1629. [CrossRef] [PubMed]

28. Bukhari, M.; Yasmin, S.; Sammad, S.; Abd El-Latif, A.A. A Deep Learning Framework for Leukemia Cancer Detection in Microscopic Blood Samples Using Squeeze and Excitation Learning. *Math. Probl. Eng.* **2022**, *2022*, 2801227. [CrossRef]

29. Boreiri, Z.; Azad, A.N.; Ghodousian, A. A Convolutional Neuro-Fuzzy Network Using Fuzzy Image Segmentation for Acute Leukemia Classification. In Proceedings of the 2022 27th International Computer Conference, Computer Society of Iran (CSICC), Tehran, Iran, 23–24 February 2022; pp. 1–7. [CrossRef]

30. Mustaqim, T.; Fatichah, C.; Suciati, N. Modification of YOLO with GhostNet to Reduce Parameters and Computing Resources for Detecting Acute Lymphoblastic Leukemia. In Proceedings of the 2022 International Conference on Advanced Computer Science and Information Systems (ICACSIS), Depok, Indonesia, 1–3 October 2022; pp. 167–172. [CrossRef]

31. Fatonah, N.S.; Tjandrasa, H.; Fatichah, C. Identification of acute lymphoblastic leukemia subtypes in touching cells based on enhanced edge detection. *Int. J. Intell. Eng. Syst.* **2020**, *13*, 204–215. [CrossRef]

32. Maaliw, R.R.; Alon, A.S.; Lagman, A.C.; Garcia, M.B.; Susa, J.A.B.; Reyes, R.C.; Fernando-Raguro, M.C.; Hernandez, A.A. A Multistage Transfer Learning Approach for Acute Lymphoblastic Leukemia Classification. In Proceedings of the 2022 IEEE 13th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), New York, NY, USA, 26–29 October 2022; pp. 488–495. [CrossRef]

33. Fatichah, C.; Suciati, N.; Mustaqim, T.; Revanda, A.R. Detection of Acute Lymphoblastic Leukemia Subtypes using YOLO and Mask R-CNN. In Proceedings of the 2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), Yogyakarta, Indonesia, 8–9 December 2022; pp. 413–418. [CrossRef]

34. Ahammed, P.; Faruk, M.F.; Raihan, N.; Mondal, M. Inception V3 Based Transfer Learning Model for the Prognosis of Acute Lymphoblastic Leukemia from Microscopic Images. In Proceedings of the 2022 4th International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE), Rajshahi, Bangladesh, 29–31 December 2022; pp. 1–4. [CrossRef]

35. Ansari, S.; Navin, A.H.; Sangar, A.B.; Gharamaleki, J.V.; Danishvar, S. A Customized Efficient Deep Learning Model for the Diagnosis of Acute Leukemia Cells Based on Lymphocyte and Monocyte Images. *Electronics* **2023**, *12*, 322. [CrossRef]

36. Ansari, S.; Navin, A.H.; Babazadeh Sangar, A.; Vaez Gharamaleki, J.; Danishvar, S. Acute Leukemia Diagnosis Based on Images of Lymphocytes and Monocytes Using Type-II Fuzzy Deep Network. *Electronics* **2023**, *12*, 1116. [CrossRef]

37. Fan, Q.; Qu, B.; Teng, J.; Hui, A.; Toe, T.T. Acute lymphoblastic leukemia Classification Based on Convolutional Neural Network. In Proceedings of the 2023 6th International Conference on Artificial Intelligence and Big Data (ICAIBD), Chengdu, China, 26–29 May 2023; pp. 877–883. [CrossRef]

38. Das, P.K.; Sahoo, B.; Meher, S. An Efficient Detection and Classification of Acute Leukemia Using Transfer Learning and Orthogonal Softmax Layer-Based Model. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2023**, *20*, 1817–1828. [CrossRef]

39. Katar, O.; Yildirim, O. An Explainable Vision Transformer Model Based White Blood Cells Classification and Localization. *Diagnostics* **2023**, *13*, 2459. [CrossRef]

40. Talupuri, P.K.; Naseeba, B.; Challa, N.P.; Sathwik, A.S. Analysis of Acute Lymphoblastic Leukemia Detection Methods Using Deep Learning. In Proceedings of the 2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS), Erode, India, 18–20 October 2023; pp. 565–574. [CrossRef]

41. Bhute, A.; Bhute, H.; Pande, S.; Dhumane, A.; Chiwhane, S.; Wankhade, S. Acute Lymphoblastic Leukemia Detection and Classification Using an Ensemble of Classifiers and Pre-Trained Convolutional Neural Networks. *Int. J. Intell. Syst. Appl. Eng.* **2023**, *12*, 571–580. Available online: https://ijisae.org/index.php/IJISAE/article/view/3955 (accessed on 1 January 2020).

42. Ghaderzadeh, M.; Hosseini, A.; Asadi, F.; Abolghasemi, H.; Bashash, D.; Roshanpoor, A. Automated detection model in classification of B-lymphoblast cells from normal B-lymphoid precursors in blood smear microscopic images based on the majority voting technique. *Sci. Program.* **2022**, *2022*, 4801671. [CrossRef]

43. Raina, R.; Gondhi, N.K.; Chaahat Singh, D.; Kaur, M.; Lee, H.N. A Systematic Review on Acute Leukemia Detection Using Deep Learning Techniques. *Arch. Computat. Methods Eng.* **2023**, *30*, 251–270. [CrossRef]

44. Mustaqim, T.; Fatichah, C.; Suciati, N. Deep Learning for the Detection of Acute Lymphoblastic Leukemia Subtypes on Microscopic Images: A Systematic Literature Review. *IEEE Access* **2023**, *11*, 16108–16127. [CrossRef]

45. Abirami, M.; George, G.V.S.; Sam, D. Acute Lymboplastic Leukemia Detection Challenges and Systematic Review. In Proceedings of the 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 23–25 January 2023; pp. 1471–1477. [CrossRef]

46. Theodore Armand, T.P.; Kim, H.C.; Kim, J.I. Digital Anti-Aging Healthcare: An Overview of the Applications of Digital Technologies in Diet Management. *J. Pers. Med.* **2024**, *14*, 254. [CrossRef]