IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 15 January 2023, accepted 5 February 2023, date of publication 14 February 2023, date of current version 22 February 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3245128

SURVEY

Deep Learning for the Detection of Acute Lymphoblastic Leukemia Subtypes on Microscopic Images: A Systematic Literature Review

TANZILAL MUSTAQIM, (Member, IEEE), CHASTINE FATICHAH, (Member, IEEE), AND NANIK SUCIATI, (Member, IEEE)

Department of Informatics, Faculty of Intelligent Electrical and Informatics Technology, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia

Corresponding author: Chastine Fatichah (chastine@if.its.ac.id)

This work was supported by the Directorate General of Higher Education, Ministry of Education and Culture of the Republic of Indonesia for Master's to Doctoral Education Scholarship Program for Excellent Undergraduates (PMDSU) 2021–2024 through the Penelitian Disertasi Doktor (PDD) Scheme under Grant 084/E5/PG.02.00.PT/2022.

ABSTRACT Computer vision research in detecting and classifying the subtype Acute Lymphoblastic Leukemia (ALL) has contributed to computer-aided diagnosis with improved accuracy. Another contribution is to serve as an assistant and second opinion for doctors and hematologists in diagnosing the ALL subtype. Early detection can also rely on computer-aided diagnosis to determine initial treatment. The purpose of this study is to review the progress of research in the detection and classification of ALL subtypes. The method's discussion focuses on the application of deep learning to the domain of object detection and classification. Motivations, challenges, and future research recommendations are thoroughly discussed to improve understanding and progress in this field of study. The study was carried out methodically by analyzing a collection of papers on the detection and classification of ALL subtypes published in science direct, IEEE, and PubMed from 2018 to 2022. The analysis of this paper field is included in the results of the selected paper. The paper selection from among 65 papers was based on inclusion and exclusion methods. Based on research methods and objectives, papers are divided into two large groups. The first group discusses the classification of ALL subtypes, while the second group discusses the detection of ALL subtypes. The discussion of prior research reveals some challenging issues and future work, such as the limited availability of the ALL subtypes dataset, the high computational complexity of the deep learning model, and further exploration of transformers in computer vision as a reference for research gaps that can contribute to future research.

INDEX TERMS Acute lymphoblastic leukemia subtypes, object detection, deep learning, CNN, blood microscopy, blood cancer.

I. INTRODUCTION

Leukemia is a hematological disease characterized by an abnormally large number of white blood cells [1]. Abnormal white blood cells are white blood cells that have not fully developed and thus perform poorly as part of the immune system. Leukemia, also known as blood cancer, is caused by an uncontrollable production of abnormal white blood cells in the spinal cord and lymphatic system [2]. Adults and children are both at risk of developing leukemia. Leukemia is frequently difficult to diagnose because the symptoms are similar to those of other diseases. As a result, early examination and detection are required so leukemia patients can receive appropriate treatment and reduce their risk of death.

The associate editor coordinating the review of this manuscript and approving it for publication was Sudipta Roy.

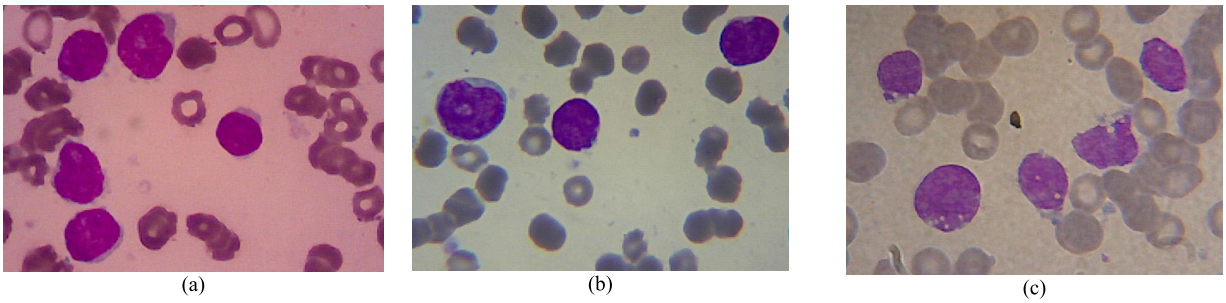

FIGURE 1. Visual examples of morphology of the ALL subtype on multi-cell blood microscopic images: (a) L1; (b) L2 and (c) L3.

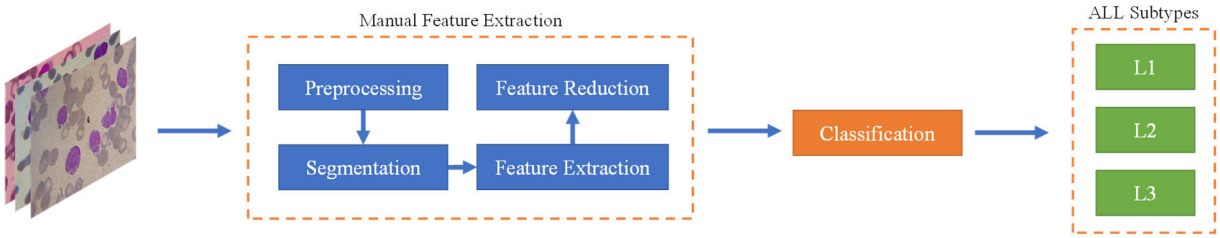

FIGURE 2. An overview of the stages of ALL detection with a machine learning model that requires several stages and manual extraction features.

Acute lymphoblastic leukemia (ALL) is one of the leukemia types [3]. ALL is a form of leukemia that originates in the bone marrow and affects the maturation of B and T lymphocyte cells [4]. Infected white blood cells can rapidly attack the blood and spread to other organs such as the lymph gland, spleen, central nervous system, and liver. The ALL classification is divided into three subtypes, L1, L2, and L3, based on French American British (FAB) [5]. There is no distinction between other L1 cells and small cytoplasm in terms of their morphological characteristics and size. The nucleus of cell L1 has a disc shape and is well organized. L2-type cells differ from other L2 cells in shape and are larger than L1 cells. L2 has an irregularly shaped nucleus and cytoplasm. Typically, the shape and size of L3 cells are identical to those of other L3 cells. The cell nucleus of L3 is round or oval in shape. Fig. 1 depicts a visual representation of the ALL subtype.

In 2016, the World Health Organization (WHO) classified ALL into three types: pre-B-cell ALL, B-cell ALL, and T-cell ALL [6], [7], [8]. Pre-B-cell ALL manifests itself during the early stages of B-lymphocyte cell development in the bone marrow. Adult cases of pre-B-cell ALL have an incidence rate of 75-80%. B-cell ALL appears more on the development of mature lymphocyte cells and has a lower incidence rate of about 3-5% in adults. T-cells ALL appear when the number of white blood cells is high and is associated with the central nervous system. T-cell ALL affects approximately 20-25% of adults.

According to the World Health Organization (WHO), morphological analysis, along with other tests like immunophenotype, cytogenetics, and molecular biology, is critical for the diagnosis and detection of Acute Lymphoblastic Leukemia (ALL) [9]. Morphological analysis is used to differentiate between normal and leukemia cells, primarily from each of the ALL subtypes. Differences in the characteristics of infected cells from the morphological analysis can be used as a guide for diagnosing and determining the type of leukemia.

Several characteristics in the blood image become obstacles during morphological analysis, including blur, noise, variations in blood smear staining, overlapping cells, cell occlusion, and cell size variations [10], [11]. Because blood images are analyzed by a hematologist, manual analysis can directly solve problems. However, a hematologist's morphological analysis of blood cell images performed manually takes a long time, and the results are dependent on expertise, inefficient on large data sets, and subjective [12]. Because the error rate due to hematology bias can reach 30% to 40%, a computer-assisted system is required to analyze morphology for automatic detection of ALL and its subtypes [13]. Identification of ALL subtypes of L1, L2, and L3 with their location in morphological blood cell images is critical because therapeutic diagnosis and treatment are dependent on this differentiation [14].

Machine learning and deep learning models can be used in the ALL subtype detection approach with computer-aided automated systems [15], [16], [17]. As illustrated in Fig. 2, the machine learning model includes several interconnected stages, such as preprocessing, segmentation, feature extraction, and classification [18]. Each stage is critical to the detection of ALL subtypes. If one stage produces suboptimal results, it will impact the other stages, particularly the final stage of ALL subtype detection. Support Vector Machine

FIGURE 3. An overview of the stages of ALL detection with a deep learning model that requires a single learning framework and an automatic extraction feature.

(SVM) [19], Random Forest [20], Ensemble Learning [21], and other machine learning models have been used to detect ALL in microscopic images of multi-cell blood. Each model has its advantages in detecting ALL subtypes. However, the shortcomings of previous studies are that machine learning models mostly only detect between normal classes and ALL classes. Machine learning models are also highly dependent on the success of each stage, so the detection process is still not optimal.

Fig. 3 shows that the deep learning model can overcome the weakness of machine learning model detection by using a single learning framework without the need for separate stages [22], [23], [24]. Deep learning models that use feed forward and backpropagation of deep learning can generalize feature representations in blood microscopic images. The weight features extracted from microscopic images of blood changed during the training model process [25]. The final feature's results are classified directly using the last layer's activation function to match the fit between the prediction and actual classes. Many researchers currently use deep learning models as the foundation of their research because deep learning approaches produce better model performance than machine learning models in terms of evaluation metrics in detecting ALL subtypes [9], [21].

The object detection model is one of the deep learning models. In a single learning framework, the object detection model performs the classification and localization of objects in an image [26]. Researchers are increasingly interested in using object detection models in the medical field because they can provide disease analysis and diagnosis based on morphology and information on medical images that is useful in the initial diagnosis process. The use of the YOLO object detection model by Khandekar et al. to detect ALL from blood microscopic images can distinguish between normal white blood cells and ALL with bounding box information [27]. Classifying and localizing cells in medical images can help doctors make early diagnoses and provide appropriate therapy to reduce disease severity.

A systematic literature review on the detection of ALL subtypes on microscopic blood images will be discussed in this review paper. Previously, several review papers addressed the detection of ALL. Das et al. recently discussed the general use of machine learning and deep learning in detecting ALL [28]. However, the discussion of their paper did not go into great detail about the use of the detection model. Another paper by Hedge et al. discusses the use of image processing methods in general with types of malaria, anemia, leukemia, thrombocytopenia, sickle cell anemia, and others [29]. Their paper does not discuss the use of deep learning, particularly object detection models, and instead focusing on blood cell types, segmentation models, and traditional classification. The paper by Deshpande et al. is nearly identical to the previous paper, but it has discussed about the use of deep learning models [2]. However, the paper does not go into great detail about the use of object detection models, specifically ALL subtype detection. Alsalem et al. [1] review the previous literature on the detection and classification of ALL, including taxonomic information. Their paper, however, does not discuss the object detection model.

The portion of the literature review on the detection and classification of ALL that has not been discussed is a research gap that will be discussed in detail in this review paper. The focus of the discussion is the use of object detection models with ALL subtype objects supported by literature, which can contribute to the development of ALL subtype detection. In addition, a number of research issues and future work in the field of ALL subtype detection will be the discussion topics.

The followings are the discussion topics of discussion papers. Part II describes the Systematic review protocol in detail, including the paper selection process from the search process and the paper criteria selection based on the scope of the research. Section III explains how deep learning is used to detect and classify ALL subtypes. Section IV discusses datasets that researchers frequently use to detect ALL subtypes. Section V discusses the evolution of the object detection model from its inception to its current state in the context of ALL subtype detection. Section VI then delves into the specifics of the research problems discovered in the survey literature, as well as the possibility of future work that can be used to fill research gaps for future research. Finally, section VII of this review paper presents the discussion and conclusions.

II. SYSTEMATIC REVIEW PROTOCOL

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) model [30] is used in this paper to determine which papers meet the criteria. The paper search was conducted manually and automatically, using

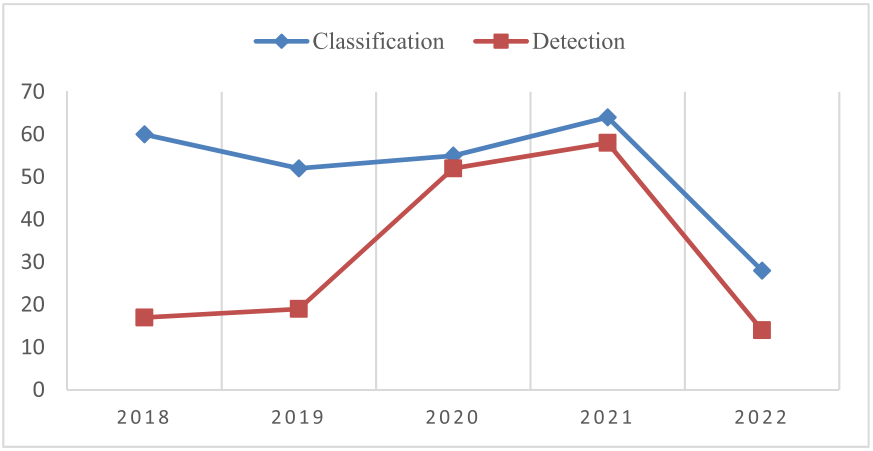

FIGURE 4. Statistical statistics on object classification and detection research fields in 2018-2022 for the identification of acute lymphoblastic leukemia subtypes.

TABLE 1. Research paper inclusion/exclusion criteria for the proposed study.

Inclusion	Exclusion
English-language research articles	Non-English research papers are excluded.
Acute lymphocytic leukemia and subtypes articles.	Non-main-topic papers are excluded.
Comprehensive articles with clear evidence.	Articles without clear results or comparison are excluded
Articles published between 2018 and 2022	Duplicate articles are excluded

three database sources: IEEE, Science Direct, and Pubmed. The search process is based on predetermined keywords to obtain paper criteria relevant to the scope of research.

The keyword used is "('acute lymphoblastic leukemia' OR 'acute lymphocytic leukemia' OR 'acute lymphoblastic leukemia subtype' OR 'acute lymphocytic leukemia subtype') AND ('detection' OR 'classification') AND ('deep learning') – segmentation." Keywords are arranged based on the main search criteria of the ALL subtype, followed with the criteria that must include classification or detection using deep learning, and the topics whose main discussion is segmentation are excluded because they are outside the scope of the research.

Search keywords are also derived from research questions developed in accordance with the Patient, Intervention, Comparison, and Outcome (PICO) framework [31]. The following are the specifics of the review of the research question paper:

1) RQ -1: Which datasets are frequently utilized in ALL detection research?
2) RQ-2: What are the differences between deep learning models for detecting and classifying ALL subtypes on blood smears?
3) RQ -3: What types of deep learning models are used for ALL detection as detection objects?
4) RQ-4: What are the challenges and opportunities for ALL detection object research?

The duration of the search process is between 28 May 2022 and 17 July 2022. The results of the paper submission are then discussed exhaustively with all authors to determine which papers are suitable for use as sources of review. Fig. 4 demonstrates that research in the field of medical object detection, including the detection of acute lymphoblastic leukemia, increases annually. Object detection data in 2022 is still limited due to the fact that the time range for searching papers does not encompass the entirety of 2022.

The results of the paper search from each database were combined and chosen using the inclusion/exclusion method of PRISMA model, as shown in Table 1. The inclusion selection process involves entering a paper into the review process based on predetermined criteria. In contrast, the exclusion selection process involves removing a paper from the review process because it does not meet the criteria. As depicted in Fig. 5, the paper selection process consists of several stages, including a check for duplication, the suitability of the title and abstract, the suitability of the paper's discussion, and an overall assessment of the paper. The initial batch of 525 papers was checked for duplication and reduced to 515 papers. The next stage was to ensure that the title and abstract were appropriate for the scope of the research, which resulted in a total of 317 papers. The stage of checking the full paper discussion by reviewing the discussion and method sections yielded a total of 92 papers. The final stage was to check the entire paper for clarity of the experiment process and metric evaluation, resulting in a total of 65 papers. The final number of papers served as the foundation for the discussion of this review paper.

III. DATASET FOR DETECTION AND CLASSIFICATION OF LEUKEMIA

According to previous research, researchers use multiple types of datasets, including public and private datasets. A public dataset refers to a dataset that can be accessed

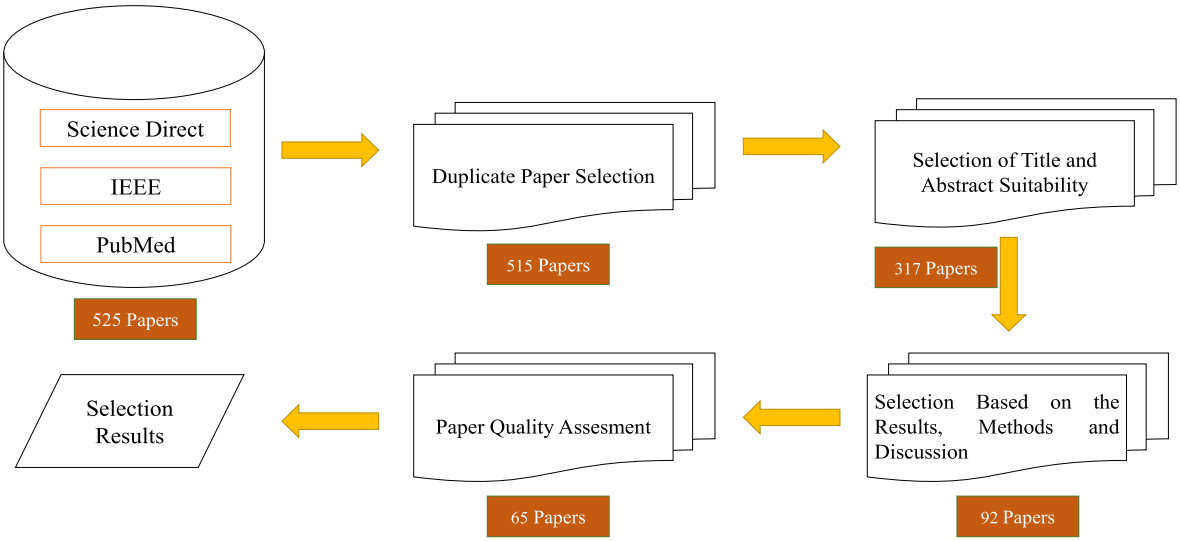

FIGURE 5. Diagram of the paper selection process.

and downloaded by the general public, allowing researchers to directly use it to evaluate the developed method [32]. Meanwhile, private datasets refer to datasets generated by researchers who typically collaborate with health institutions such as hospitals and health laboratories to collect the appropriate dataset [33].

The blood microscopy image dataset was acquired by taking images of a blood smear under a microscope [34]. The blood smear is the process of analyzing blood by placing drops of blood on a slide, spreading it out, and staining it, followed with a hematologist's examination using a microscope. The staining process can utilize either Leishman stain or Wright-Giemsa stain [35].

In blood microscopic images used for research, there are two types of cells: single cells and multi-cells. Single cells are typically cropped from multi-cell images, such as the ALL-IDB2 dataset obtained from ALL-IDB1 [36]. Single-cell images only contain one cell information, whereas multi-cell images contain multiple cell information based on the image taken from the blood smear [37]. Tables 2 and 3 how details of the evaluation metrics obtained in previous studies using object detection and classification models. A sample dataset of multi-cell and single-cell microscopic blood images is shown in Fig. 6.

A. MULTI-CELL DETAIL EXPLANATION

Images captured through a microscope containing multiple cells, including red blood cells, platelets, etc. are called multi-cell images. Different staining processes and varying degrees of sharpness present challenges for multi-cell images [2]. American Society of Hematology (ASH) and ALL-IDB1 are examples of multi-cell image datasets.

The ASH dataset is a publicly accessible dataset that can be downloaded directly from the ASH website's image databank [4]. The ASH dataset contains microscopic images of blood smears with various conditions and distinct hematological and pathological cell information because it was compiled by a large number of contributors. Consequently, the analysis process utilizing the ASH dataset can minimize bias towards one or more image conditions based on real-world conditions. However, because the ASH dataset does not yet provide ground truth for research purposes in the object detection domain, researchers require assistance from hematologists in grouping and creating ground truth from the available images.

The ALL-IDB1 dataset includes ALL subtype information from L1, L2, and L3 as well as available ground truth data for object detection research [17]. In this instance, the research process can be conducted immediately without the need for manually grouping and creating ground truth images of microscopic blood smears. However, because the ALL-IDB1 dataset is compiled by a limited number of contributors, the conditions of variation and hematological and pathological information in the ALL-IDB1 dataset are less comprehensive than in the ASH dataset.

Hamza et al. used the ALL-IDB1 dataset to test the biLSTM model for classification [15]. Then in 2021, Dash and Meher conducted research utilizing the same dataset to evaluate a lightweight model [12]. The used model is a hybrid of MobilenetV2 and ResNet18.

B. SINGLE-CELL DETAIL EXPLANATION

There are various single-cell image types, including raw crop versions of multi-cell datasets and segmented single-cell images. The raw cropped image still contains the background and other cell information [38]. Distinct from single image cells have been segmented prior to processing. The image only contains information about the cytoplasm and nucleus, allowing researchers to use the dataset without preprocessing. ALL-IDB2 and C-NMC 2019 are the single cell datasets available to the public. ALL-IDB2 is a collection of normal

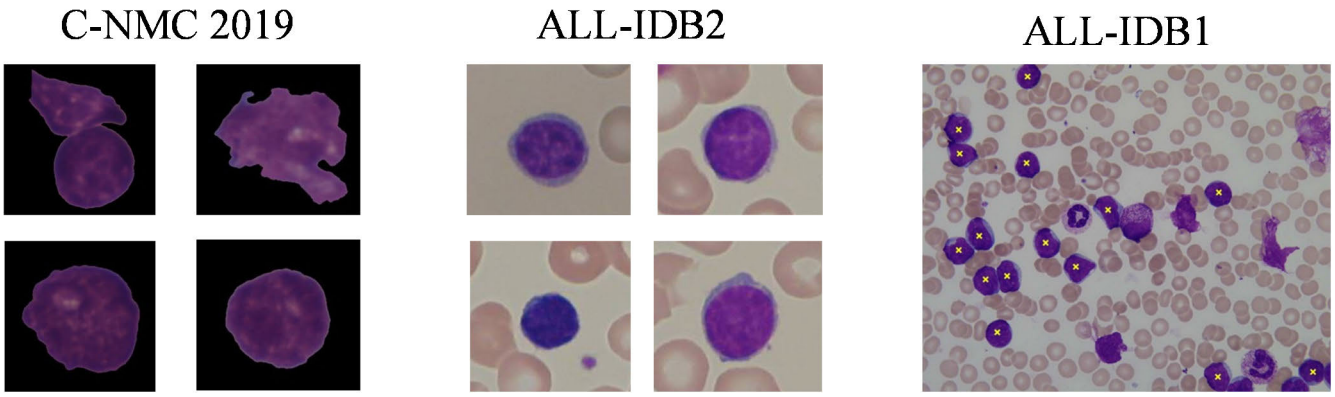

FIGURE 6. Multi-cell (ALL-IDB1) and single-cell (C-NMC 2019 and ALL-IDB2) blood microscopic image dataset samples.

and blast cells from the ALL-IDB1 dataset that have been cropped. ALL-IDB2 includes 260 images, of which

fifty percent are of lymphoblasts [39]. ALL-IDB2 images share the same gray level characteristics as ALL-IDBI images. The dataset is publicly accessible and free to use. The C-NMC 2019 dataset contains images of single cells from 118 individuals, including 49 healthy controls and 69 cancer patients [40]. Due to the different sizes of the cell images, a constant size of 350 × 350 is achieved for each image by filling the columns and rows with zero intensity after centering each cell within the image. Jha and Dutta [41] used ALL-IDB2 as a test dataset for their hybrid active contour and fuzzy c-means segmentation model [41]. The segmentation results are then classified using a modified version of a deep CNN model. Magpantay et al. [26] used the YOLOv3 model for classification of ALL in the C-NMC 2019 dataset.

IV. DEEP LEARNING FOR THE DETECTION AND CLASSIFICATION OF LEUKEMIA

Utilizing deep learning models for the detection of ALL subtypes has become a popular topic of discussion amongst academics. Deep learning models have an advantage over conventional machine learning models due to their ability to recognize pattern representations in the data [42]. Deep learning models can perform feature extraction without the need for manual feature extraction. The process of feature extraction is taken automatically from the data based on the available pattern representation [43]. Moreover, deep learning models can produce model performance with a high level of generalization compared to conventional machine learning models and are very effective for analyzing large datasets [33].

Convolutional neural network (CNN) is the deep learning model frequently used by researchers for analyzing microscopic blood images [44], [45]. CNN is composed of convolutional layer blocks that extract pattern representations. The convolution process with filter displacement on the input image permits the generation of a feature map that indicates the location and pattern of objects within the input image. CNN's innovation of automatically examining multiple filters in parallel in the model generates object-specific features in the input image [46].

CNN can be used to classify and detect microscopic blood image objects by applying a deep learning model [24]. Classification is the process of matching objects in an image to their actual class, with the prediction scope encompassing the entire input image. Contrary to object detection, which performs classification and localization of each object in the input image containing multiple objects, multiple object detection performs classification and localization of each object in the input image containing more than one object. This fundamental difference can be used as a reference for identifying ALL subtypes [27]. The results of previous studies of object classification and detection models using state-of-the-art models for the detection of ALL and ALL subtypes are shown in Table 4.

A. CLASSIFICATION OF ALL

The objective of the classification of white blood cell objects is to identify the type of ALL subtype cells from a single input of microscopic blood images [4]. The detection process employs the appearance of white blood cell morphological characteristics in microscopic images of blood. Morphological differences in the nucleus and cytoplasm become an automatic reference for feature extraction by deep learning models [47]. Laosai and Chamnongthai [19], [48] classified white blood cells into ALL and AML subtypes using imaging flow cytometry as a preliminary processing step [48]. The final stage of research is classified using a four-layer CNN model. The author uses Cluster of Differentiation (CD Markers) as the final validation of the ALL subtype detection process resulted in the deep learning model [19].

Deep learning models are frequently used by researchers to classify white blood cells as healthy or infected with ALL in addition to detecting ALL subtypes. In 2020, Safuan et al. used the Alexnet, Googlenet, and VGG16 models to classify white blood cells and distinguish between healthy and leukemia cells [49]. The author compared the three models

TABLE 2. List of studies using multi-cell datasets to classify and detect all and all subtypes.

Literature	Type	Dataset	Image Type	Specifity (%)	Sensitivity (%)	Precision (%)	Accuracy (%)	F1 Score (%)	folder (%)
Khandekar *et al.* [27]	YOLOv4			. -	92	-	-	92	95.57
	Faster R-CNN			-	-	-	-	-	96.5
Talukdar *et al.* [70]	EfficientDet D3			-	-	-	-	-	91.3
	CenterNet Hourglass			-	-	-	-	-	95
Revanda *et al.* [53]	Mask R-CNN			-	79.41	79.91	-	80.07	81.96
	Mask R-CNN + HE			-	79.48	77.88	-	77.08	60.59
	Mask R-CNN + DHE			-	72.90	73.68	-	74.09	75.82
	Mask R-CNN + CLAHE			-	71.10	73.04	-	74.09	78.53
	Mask R-CNN + EFF			-	81.61	85.17	-	83.72	83.65
Das *et al.* [50]	VGG16	ALL-IDB1	Multi cell	88.36	82.34	85.43	85.94	83.84	-
	VGG16 + RF			91.36	85.79	89.76	89.62	87.73	-
	VGG16 + SVM			95	87.33	92.03	91.48	89.62	-
	ResNet50			90.73	96.92	93.85	95.08	95.36	-
	ResNet50 + RF			98	93.48	97.14	95.72	95.27	-
	ResNet50+ SVM			97	96.19	95.72	95.68	95.95	-
	MobileNetV2			97	96.19	95.72	95.68	95.95	-
	MobileNetV2 + RF			98	96.19	97.33	96.29	96.76	-
	MobileNetV2+ SVM			99	100	98.57	99.39	99.26	-
	VGG16			82	88.13	83.04	85.22	85.51	-
	VGG16 + RF			86.05	87.53	86.42	86.92	86.97	-
	VGG16 + SVM			89.68	90.24	89.99	90	90.11	-
	ResNet50			95.26	88.63	96.33	92.05	92.32	-
	ResNet50 + RF			84.95	95.34	87.05	90.26	91	-
	ResNet50+ SVM			85.95	97.95	87.93	92.05	92.67	-
	MobileNetV2			89.10	95.95	90.02	92.56	92.89	-
	MobileNetV2 + RF	ALL-IDB2	Single cell	92.63	96.95	93.58	94.87	95.24	-
	MobileNetV2+ SVM			97.37	98.95	97.61	98.21	98.28	-
Rastogi *et al.* [23]	GLCM + XGB			-	84.4	93.1	86.5	88.5	-
	LBP + Gabor + GLCM + XGB			-	84.6	73.3	76.9	78.6	-
	Pretrained VGG16 + XGB			-	84.6	78.6	80.77	81.4	-
	Pretrained Xception + XGB			-	80.8	84	82.69	82.3	-
	LeuFeatx + ETC			-	98.69	94.28	96.15	96.44	-
	LeuFeatx + RF			-	97.21	94.94	95.76	95.99	-
	LeuFeatx + XGB			-	94.96	91.62	92.67	91.62	-
	LeuFeatx + SVM			-	96.6	90.63	92.3	93.54	-

TABLE 3. List of studies using single-cell datasets to classify and detect all and all subtypes.

Literature	Type	Dataset	Image Type	Specifity (%)	Sensitivity (%)	Precision (%)	Accuracy (%)	F1 Score (%)	folder (%)
Khandekar *et al.* [27]	YOLOv4			. -	96	-	-	92	98.57
Magpantay *et al.* [26]	YOLOv3			-	-	-	-	-	99.8
Princess *et al.* [112]	Resnet-50			93.49	76.55	-	81.12	-	-
	Densenet-201			92.35	76.25	-	80.61	-	-
	WideResnet50-2			93.35	75.97	-	80.56	-	-
Mathur *et al.* [113]	ResNet-18	C-NMC 2019	Single cell	-	84.93	65.68	-	80.56	-
	EfficientNet			-	90.45	65.74	-	84.73	-
	WideResNet-50			-	85.63	69.40	-	76.67	-
	WSDAN			-	88.65	72.82	-	86.38	-
	ResNet-18 + GAC			-	93.11	61.64	-	87.62	-
	ResNet-18 + LAC			-	93.89	73.60	-	87.67	-
	ResNet-18+ Mixup			-	83.44	75.15	-	86.05	-
	ResNet-18+ Mixup + MTL			-	93.85	82.11	-	91.89	-
Jawahar *et al.* [22]	AlexNet			-	75	83	90	78	-
	VGG16			-	61	92	88	74	-
	GoogleNet			-	60	89	87	71	-
	ResNet-50			-	81	78	89	79	-
	ALNett			-	98	93	91	96	-

to determine which model performs best. Alexnet is the best model for detecting leukemia in terms of accuracy and computation time, according to the author's research. In 2021, Das et al. conducted tests and comparisons of deep learning models and conventional machine learning models to determine a lightweight model with the shortest computation time and the highest evaluation metrics [50]. The ShuffleNet model outperforms the ResNet 50, VGG19, and other models in the author's research.

B. CLASSIFICATION AND LOCALIZATION OF ALL

ALL subtype cell detection simultaneously classifies and localizes multi-objects on a single input of multi-cell blood microscopy images. The morphology of white blood cells is one of the distinguishing characteristics between L1, L2, and L3 in the FAB classification system. Using the fourth version of the You Only Look Once (YOLO) object detection model developed by Bochkovskiy et al. [51] in reference to the initial model developed by Redmon et al. [52], the 2021 study by Khandekar et al. [27] classified and localized each white blood cell according to its actual class, namely the leukemia class or the normal class.

Revanda et al. [53] conducted an additional study that specifically classified and localized white blood cells into ALL subtypes. The researchers used the Mask-RCNN model to perform segmentation instances of ALL subtypes with varying preprocess image enhancement types. The subtype of ALL depicted in the image was determined by a majority vote. If what is detected differs between L1 and L2, the dominant calculation is executed. If the L3 subtype appears, the image is determined to be of the L3 type. This is based on the recommendations of hematologists regarding the calculation of ALL subtype cells [54].

The majority of previous studies used CNN-based deep learning models capable of extracting features from images for classification and localization. Previous research findings can classify and localize ALL subtype objects with high evaluation values for metrics such as accuracy, precision, recall, and mAP. However, previous research has not paid much attention to the model's resource complexity, such as the GFLOPS value and the number of parameters. The dominant model continues to produce a model that is resource-intensive, making it unsuitable for embedded devices and computers with limited specifications.

TABLE 4. List of studies on all and all subtype classification and detection.

Literature References	Research Type	Remarks
Safuan *et al.* [49]	Classification	The research compared three pre-trained models: alexnet, googlenet, and vgg16. The classification procedure consists of the classification of wbc types using the lisc dataset and the classification of all using the ALL-IDB2 dataset.
Aftab *et al.* [3]	Classification	Leukemia detection is done using Apache Spark BigDL library and GoogleNet deep transfer learning on microscopic images of human blood cells. BigDL outperformed Keras in efficiency and accuracy.
Princess *et al.* [112]	Classification	This study used DenseNet201 for "deep" and Wide-ResNet-50-2 for "wide" architectures with ResNet50 as the Baseline. Transfer learning reduces training steps in this study.
Dash & Meher[32]	Classification	This paper proposes a lightweight, computationally efficient SqueezNet classifies malignant and benign with promising results. Channel shuffling and pointwise-group convolution speed it up.
Akanksha *et al.* [114]	Classification	This article covers deep learning blood cell counting and ALL detection. Otsu's median filters and RGB segmentation preprocess the images. Then, scale-invariant feature transform (SIFT) extracts the features and applies them to recurrent neural network (RNN) classification.
Ullah [115]	Classification	To overcome deficiency training samples and visual flaws that distinguish leukemia from normal, this study proposed an augmented image enhanced bagging ensemble learning with an elaborately developed training subset.
Genovese *et al.* [38]	Classification	This study proposed transfer learning based on a histopathological dataset for ALL detection, in which a CNN is trained on a histopathology dataset to classify tissue types and then fine-tuned on the ALL database to detect the presence of lymphoblasts.
Mathur *et al.* [113]	Classification	This paper proposed a new architecture called the Mixup Multi-Attention Multi-Task Learning Model (MMA-MTL), which uses Pointwise Attention Convolution Layers and Local Spatial Attention blocks to capture global and local features at the same time.
Genovese *et al.* [116]	Classification	An adaptive unsharpening method was proposed in this paper to improve blood sample images. The method employs image processing techniques and deep learning to normalize the radius of the cell, estimate focus quality, adaptively improve image sharpness, and then perform classification.
Al-Qudah *et al.* [82]	Classification + Localization	This paper used Locality Sensitive Hashing (LSH) to create a novel balanced dataset of 2500 synthetic blood smears. This dataset, automatically annotated during generation, will be made public for research and covers 17 essential blood cell categories.
Khandekar *et al.* [27]	Classification + Localization	The paper suggested using AI to automate blast cell detection for ALL diagnosis. The system predicts leukemic cells from microscopic blood smear images using object detection. The paper implemented YOLOv4 for cell detection and classification.
Su *et al.* [10]	Classification + Localization	The paper created a high-quality Bone Marrow Aspirate Smear Image Dataset (BMASID) with 230 labeled images. The paper proposed a benchmark-trained Cell Recognition Network (CRNet). CRNet uses a cell detector and classifier to identify and classify cells in bone marrow aspirate images.
Talukdar *et al.* [70]	Classification + Localization	This study compares Faster R-CNN, EfficientDet D3, and CenterNet Hourglass object detection models and their inference results. The three models were compared using COCO evaluation metrics to determine the best model performance for the task.
Magpantay *et al.* [26]	Classification + Localization	This study proposed a computer-assisted diagnosis method employing a deep learning approach based on transfer learning. YOLOv3 is used to train a deep learning model that classifies ALL and normal cells.
Revanda *et al.* [53]	Classification + Localization	This paper uses Mask R-CNN to classify acute lymphoblastic leukemia in white blood cell microscopy images. Mask R-CNN was used to detect white blood cell microscopy images using transfer learning. This paper used contrast enhancement to fix the lighting in stained white blood cell microscopy images.

V. OBJECT DETECTION BASED ON DEEP LEARNING FOR LEUKEMIA DETECTION

Based on the workflow of object detection in the image, object detection models based on CNN can be divided into two types: two-stage detectors and one-stage detectors [55]. A two-stage detector model has a separate module to produce object input in the form of a proposal region and then perform the object classification stages separately [56]. Furthermore, the one-stage detector uses a direct learning framework to perform object classification and localization in the image [57]. Because the model structure of a one-stage detector is simpler than that of a two-stage detector, it is more commonly used in real-time object detection.

Another object detection model that performs better in terms of accuracy and computational efficiency than the CNN-based model is the transformer-based model. In the field of Natural Language Processing (NLP), the transformer model has evolved into a cutting-edge model. Transformer is a deep learning model that employs an attention mechanism that assigns different weights to different parts of the input data. Vision transformers (ViT) are transformer models in the field of computer vision that were published in 2021 by Google Research Brain Team researchers Dosovitskiy et al. [58]. The ViT model employs a series of image patches as input, as well as a series of word embeddings. The main distinction between the ViT and CNN models is in the feature extraction process flow; CNN uses an array of pixels, whereas ViT divides the image into visual tokens. ViT divides the image into fixed-size patches, embedding is implemented for each image patch, and a positional embedding sequence is included as input to the transformer model. DeTR [59] and Swin Transformer [60] are transformer-based object detection models

There aren't many object detection models used for ALL and ALL subtype analysis and detection. YOLOv2, YOLOv3, YOLOv4, and Mask R-CNN models were used. Other models for detecting ALL and ALL subtypes have not been explored further, so there are still opportunities to develop the domain of research methods in the future. The presentation of cancer detection case studies other than ALL is added to the object detection model that has not been used for detection of ALL and ALL subtypes to add insight and serve as a starting point for future research.

A. TWO-STAGE DETECTOR

Two-stage detectors operate in two stages. The first stage employs a module known as the regional proposal network (RPN). RPN is used to find the exact location of an object in an image, with the output being a proposal region containing object information. The RPN results are then classified using CNN to determine the existence of objects in the following stage. The two-stage detector produces accurate results, but the constituent models are complex and unsuitable for real-time applications. Fig. 7 depicts the object detection model diagram employing a two-stage detector approach.

1) R-CNN

The region-based convolutional neural network (R-CNN) proposed by Girschik et al. in 2014 employed a different strategy for locating proposal regions [61]. The technique employed is selective search [62], which extracts only the proposal region by restricting the number of regions to 2000. In the second stage of regional classification, the addition of 2000 regions is anticipated to reduce the complexity of the model. In the second stage, the results of the 2000 region proposal are input into the AlexNet model [63] to generate a 4096-dimensional feature vector output. The SVM model is used to classify the results of feature extraction using AlexNet in order to detect the presence of objects in the proposal region. The algorithm then predicts its bounding box using a trained bounding box regressor, which predicts the center, width, and height of the box's coordinates. The weakness of R-CNN is that it still generates a large number of proposal regions, up to 2,000, and is unsuitable for real-time object detection because each image still requires approximately 47 seconds for object detection [64].

Researchers have not investigated the use of R-CNN for the detection of ALL subtypes because a more recent and accurate R-CNN development model already exists. Historically, the R-CNN model has been used to detect and classify breast mammogram tumors [65]. Breast tissue and fibroglandular segmentation modules are incorporated into a modified stratified region-based convoluted tissue using the method.

2) SPP-NET

The operation of SPP-net is nearly identical to that of R-CNN, but there is a minor difference in the initial stage. In SPP-net, the convolution process is performed first, followed with a selective search on the map feature results [66]. This is done to account for the constraints imposed by the image's input size and aspect ratio variations. The SPP-Net model has the advantage of being faster than the R-CNN model at the same level of accuracy. Another advantage is that it can handle images of any shape or aspect ratio, so object deformation due to input warping does not affect model performance. Due to the similarity between the SPP-net and R-CNN models, the SPP-net model has the same flaws as the R-CNN model, including a high level of model complexity and unsuitability for real-time object detection.

SPP-Net possesses intriguing advantages over R-CNN, but researchers have not examined it for the detection of ALL. Earlier researchers employed SPP-Net to detect glial tumors on MRI images [67]. SPP-net can utilize feature maps of images of various sizes, generate subsamples to produce fixed-length sets, and classify them.

3) FAST R-CNN

Fast R-CNN was proposed by Girschick et al. in 2015 to overcome the shortcomings of the R-CNN and SPP-net models [64]. Fast R-CNN proposes a model capable of object detection within a single learning framework. Almost identical

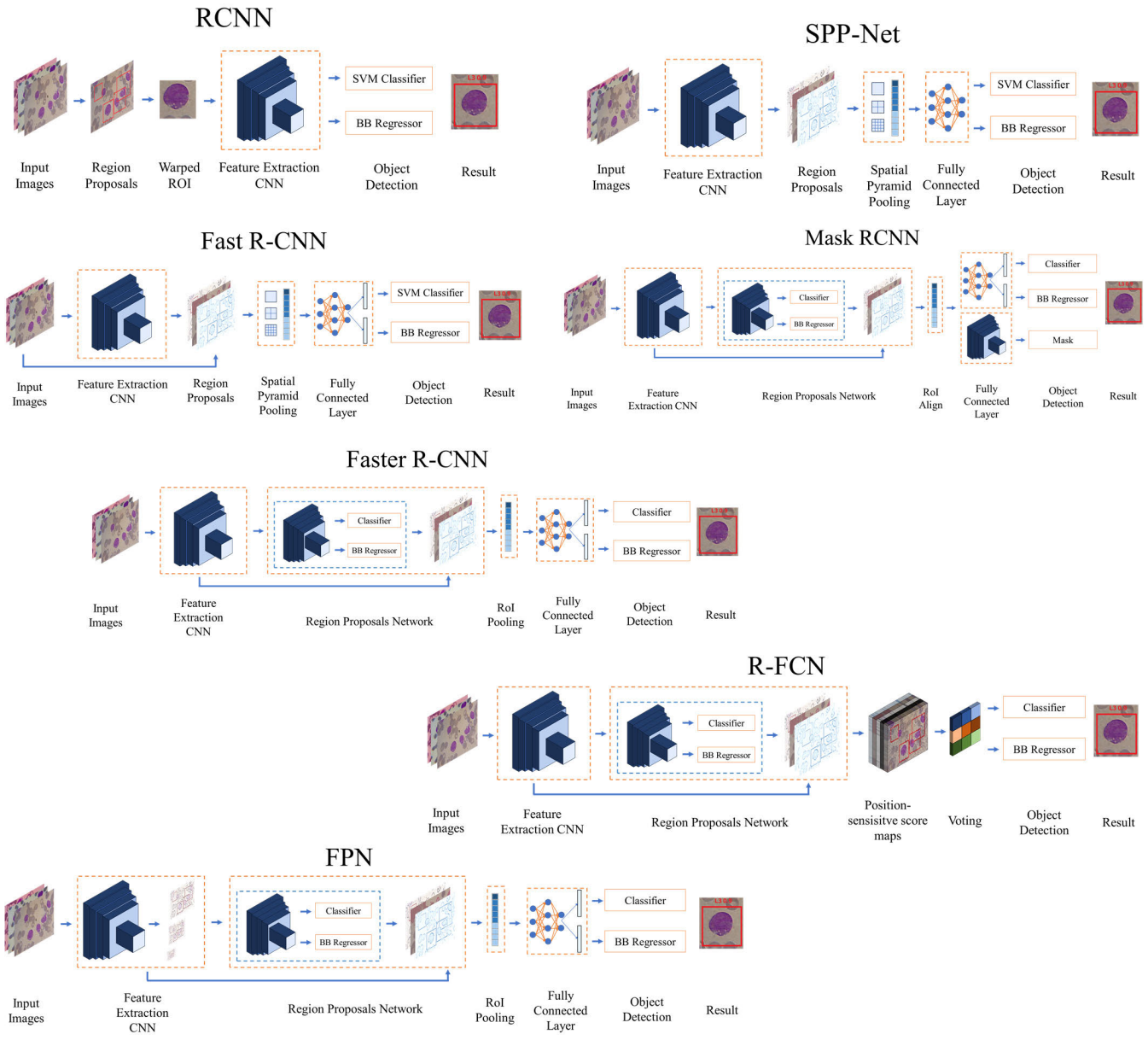

FIGURE 7. **Illustration of object detection model diagram with two-stage detector approach.**

to SPP-net, the convolution process is applied to the input image and mapped to the proposal region at this stage. Fast R-CNN substitutes a polling layer for the SPP model. This layer is connected to a fully connected network and branches into a softmax layer for object classification and a bounding box regressor for calculating object bounding box predictions. This model also modifies the loss bounding box regressor function from L2 to smooth L1 to improve model performance. With greater precision, fast R-CNN can outperform the R-CNN model with 146x faster scatter detection speed. The modification process in the Fast R-CNN model simplifies the model relative to the R-CNN model, and the model's performance capability approaches that of real-time object detection.

Fast R-CNN has previously been compared to other R-CNN models for the detection of brain tumors [68]. Fast-RCNN has a quicker execution time than the original R-CNN model utilizing Alexnet backbone. Other backbones have a longer execution time than Alexnet.

4) FASTER R-CNN

Faster R-CNN compensates for the inefficiency of Fast R-CNN, which generates region proposals slowly. Ren et al. [69] proposed a fully convoluted network as an RPN in the Faster R-CNN model to generate a proposal region. In addition, it is distinguished by the presence of an anchor box as a substitute for SPP to accommodate different object

sizes. R-CNN anchor box uses multiple bounding boxes with different aspect ratios and traces them backwards to determine the location of the object. Images sent to CNN are initially used to generate a set of feature maps. This information is sent to the RPN, which creates bounding boxes and sorts the data into groups. Then, the selected proposals are mapped back to the previous CNN layer's feature map in the RoI pooling layer. Finally, the proposal is delivered to the layer with complete connectivity. The results are then transmitted to the classifier and regressor bounding boxes. The only difference between Faster R-CNN and Fast R-CNN is the use of RPN as a regional proposal module. A Faster R-CNN can improve the accuracy of the previous model by 3%, and the speed of object detection can reach 5 frames per second.

Previous research compared Faster R-ability CNN's to detect cells in the ALL-IDB1 dataset [70]. In terms of average precision, R-CNN can outperform EfficientDet D3 and Centernet Hourglass. However, its average recall value is inferior to that of CenterNet Hourglass.

5) FPN

Lin et al. [71] proposed a Feature Pyramid Network (FPN) model with a top-down architecture and lateral connections for constructing high-level semantic features at various scales. FPN consists of two paths: a bottom-up path that uses ConvNet to determine the feature hierarchy at various scales, and a top-down path that converts coarse feature maps at higher levels into high-resolution features. Connecting these paths laterally via a 1×1 convolution operation increases the semantic information in the feature. FPN can convey a high level of meaning across all scales, reducing the likelihood of detection errors. Researchers developed the benefits of FPN to segment multiple human organs, including lung nodules. 3D Feature Pyramid Network (3DFPN) is an example of a developed method for increasing nodule detection sensitivity by employing multi-scale features to increase nodule resolution and top-down parallel paths for transiting high-level semantic features to complement low-level common features [72].

6) R-FCN

Dai et al. [73] proposed a Region-based Fully Convolutional Network (R-FCN), emphasizing a balanced distribution of computations within the model. In contrast, the model whose architecture is comparable to R-CNN utilizes two distinct stages and a substantial number of resources. R-FCN prioritizes position-sensitive score maps in order to overcome translation-invariance in image classification and translation-variability in detection objects. On this sensitive score map, the relative location of the subject is recorded, which is then added together to determine its precise location. The R-FCN accomplishes this by dividing the desired area into a grid and assigning a score to each cell based on how likely it is to be in its designated location. The object class is then estimated by averaging the scores. The R-FCN detector is comprised of four convolution networks that operate in

detection. ResNet-101 [74] was initially utilized to extract feature maps from images. The intermediate output (Conv4 layer) is sent to the Regional Proposal Network (RPN) to find the RoI proposal, while the final output is sent to the classifier and regressor after passing through the convolution layer. The classification layer makes predictions using the generated position-sensitive map and the proposed RoI, while the regression network provides bounding box attributes.

A previous study [75] used R-FCN (region-based full convolution tissue) to detect laryngeal lesions. The classification based on an image depicts a single organ and a plain background. Target detection is capable of identifying multi-organ targets in complex environments. With sufficient training data, it is possible to identify complex scene targets.

7) MASK R-CNN

Mask R-CNN is a variant of Faster R-CNN that incorporates a module for instance segmentation [76]. At the end, a new branch is added to classify each pixel based on the actual segmentation of classes. The R-CNN mask model is nearly identical to the Faster R-CNN mask model, with the exception of using the RoIAlign Layer instead of the RoIPool layer to improve pixel-by-pixel classification performance and prevent misalignment caused by spatial quantization. ResNeXt-101 [77] was selected as the backbone and the Pyramid Network (FPN) feature for greater accuracy and speed. CNN's R-loss function is updated with mask loss and employs five anchor boxes with a 3:2 aspect ratio, similar to FPN. Mask R-CNN training is identical to faster R-CNN training in general. Mask R-CNN outperforms the best available architecture with a single model. It also adds the ability to group instances through additional computations. It is simple to train, adaptable, and effective in a variety of situations, such as key point detection, human pose prediction, etc. However, its performance remains slower than in real-time.

Previous research has used Mask R-CNN on microscopic images of white blood cells to classify acute lymphoblastic leukemia, which can support the diagnosis process effectively and efficiently [53]. The research implemented Mask R-CNN with transfer learning to suit the task of segmenting instances from microscopic images of white blood cells. In order to address the issue of poor lighting in microscopy images of stained white blood cells, the researchers added a contrast-enhancement process to the image dataset. To evaluate the proposed method, the experimental procedure employs actual hospital data sets.

B. ONE-STAGE DETECTOR

A one-stage detector differs from a two-stage detector in that there is no separate process for locating candidate objects. The process of determining candidate objects at the one-stage detector stage is carried out directly from the process of dividing the image into a certain grid size, and each grid

is responsible for finding objects. The presence of anchors aids the process of finding objects so that they can be found correctly. A one-stage detector has a very fast object detection speed, allowing it to detect objects in real-time. Fig. 8 depicts the object detection model diagram employing a one-stage detector approach.

1) YOLO

You Only Look Once (YOLO) detects and predicts image objects using a single learning framework and a regression approach. YOLO partitions the input image into a S × S cell grid [52]. Each grid cell predicts a boundary box and an object that indicates whether or not it contains an item. Each grid pixel predicts the conditional probability that an object class exists. YOLO predicts five parameters for each bounding box, including x, y, w, h, and confidence. The coordinates indicate the center of the (x,y) cell's bounding box. The values of X and y range from 0 to 1. It is anticipated that the width and height of the bounding box will comprise a small portion of the image. The confidence score indicates whether or not the bounding box contains an object and its precision. Zero confidence occurs when the bounding box is empty. If the bounding box contains an object, the confidence score is the intersection of the prediction with the underlying truth. YOLO computes the probabilities of network cell class membership. The class probability depends on the object of the grid cell. YOLO estimates a set of class probabilities per grid cell despite the bounds box. YOLO estimates each grid cell's parameters. YOLO is significantly faster and more accurate than two-stage detector models. However, the accuracy of localization for small or grouped objects and cell boundaries still presents significant limitations.

Previous research proposed a YOLO-based CAD system with four major stages: mammogram preprocessing, feature extraction utilizing multi-convolution deep layers, mass detection with confidence models, and mass classification utilizing fully connected neural networks [78]. YOLO was trained using a set of training mammograms containing information on mass and type ROI. A trained YOLO-based CAD system can detect and classify benign or malignant masses. The proposed system appears feasible as a CAD system capable of simultaneous detection and classification. This system also treats difficult breast cancer cases, such as chest muscle masses or congested areas.

2) YOLOV2

YOLOv2, also known as YOLO9000, is a real-time model for locating objects with a single stage. It is superior to YOLOv1 [52] in numerous ways, including the use of Darknet-19 as the backbone, batch normalization, the use of high-resolution classifiers, the utilization of anchor boxes to predict bounding boxes, and more. DarkNet-19 has replaced the primary structure of GoogleNet [79]. It employs numerous impressive techniques, such as Batch Normalization [80] to improve convergence, joint training of classification and detection systems to increase detection classes, removal of fully connected layers to increase speed, and the use of learned anchor boxes to improve memory and priority. YOLOv2 also used WordNet [81] to combine hierarchical classification and detection datasets. Even if the hyponym is incorrectly classified, this WordTree can be used to predict a greater conditional probability of a hypernym. This boosts the overall performance of the system. The new YOLOv2 architecture makes it simpler to choose a model based on its speed and precision and has fewer parameters.

In a previous study [82], the YOLOv2 model was used to identify blood cells in microscopic images of multicellular blood. The YOLOv2 model is modified through random resizing of the network input. The used dataset is a synthetic dataset derived from ALL-IDB1. The modified YOLOv2 model's performance is superior to the original YOLOv2 model.

3) YOLOV3

In terms of speed, accuracy, and class specificity, there are substantial differences between YOLOv3 and earlier versions [83]. The differences between YOLOv2 and YOLOv3 in terms of precision, speed, and architecture are significant. YOLOv2 was released in 2016 [79], two years prior to the release of YOLOv3. YOLOv2 employs Darknet-19 for the backbone feature extractor, whereas YOLOv3 has switched to Darknet-53. Joseph Redmon and Ali Farhadi, the creators of YOLO, were also responsible for developing the Darknet-53 backbone. Darknet-53 is more powerful than Darknet-19 due to its 53 convolution layers, as opposed to the previous total of 19, and it is also more efficient than competing backbones (ResNet-101 or ResNet-152). YOLOv3 is quick and accurate about mean precision mean value (mAP) and intersection value above union (IOU). It detects objects significantly faster than other detection methods while maintaining the same level of performance. While being trained, the new YOLOv3 uses binary entropy cross-loss and an independent logistic classifier to make class predictions. Due to this modification, it is now possible to train YOLOv3 models with complex datasets, such as the Open Images Dataset (OID) created by Microsoft. The OID has dozens of overlapping labels for photographs in its collection, including ''men'' and ''people.'' YOLOv3 employs a multilabel technique, which enables classes to be more specific and multiple instances per bounding box. In contrast, YOLOv2 employs softmax, a mathematical function that transforms a number vector into a probability vector, where the probability of each value is proportional to its relative scale in the vector. In other words, the implementation of softmax YOLOv2 determines the probability of each value.

Previous research [26] proposed the automatic detection of ALL and healthy cells to make up for an expert's lack of

manual analysis. YOLOv3 is the used model, which employs a transfer learning strategy to generate low loss values and high mAP evaluation values. The evaluation results

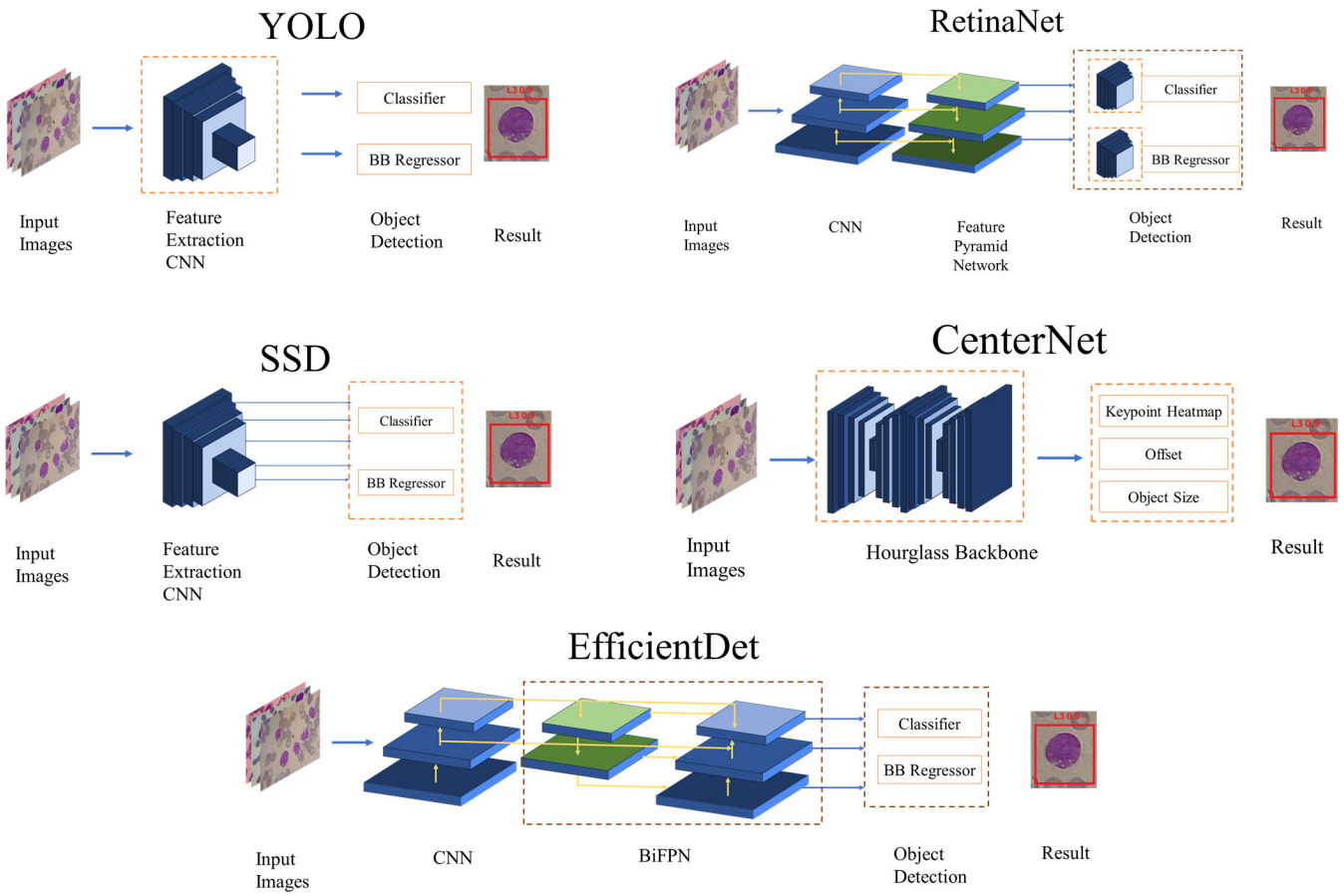

FIGURE 8. Illustration of object detection model diagram with one-stage detector approach.

indicate that the YOLOv3 model can differentiate between healthy and ALL cells.

4) YOLOV4

YOLOv4 incorporates a large number of innovative ideas to produce object detectors that are quick to train and compatible with existing production systems [51]. This is accomplished by employing the "Bag of Freebies" and "Bag of Specials" strategies, both of which only prolong the time spent on training and have no bearing on the time required to draw conclusions. YOLOv4 employs numerous data augmentation strategies, regularization techniques, class label alignment, CIoU-loss [84], Cross mini-Batch Normalization (CmBN) [85], Self-adversarial training, and a Cosine annealing scheduler [86] to enhance training. What included in the network and referred to as "Bag of Specials" are procedures that affect only inference time. Among these techniques are Mish activation [87], Cross-stage Partial Connections (CSP) [88], SPP-Block [66], PAN Path Aggregated Block [89], and Multi-input weighted residual connection (MiWRC). Moreover, genetic algorithms are employed in the search process of hyper parameters. It has an ImageNet CSPNetDarknet-53 backbone, SPP and PAN block necks, and a YOLOv3 detection head. YOLOv4 can be easily trained on a single GPU, whereas most other known detection techniques require the use of multiple GPUs. It is twice as quick as EfficientDet while delivering comparable performance.

Previous research [27] proposed an object detection method that predicts leukemia cells from microscopic blood smear images. The model used for cell detection and classification is YOLOv4. Each cell is labeled as either a blast cell (ALL) or a healthy cell (HEM) during the detection process, which is defined as a binary problem. Object Detection is trained and evaluated using the ALL IDB1 and C-NMC 2019 datasets. Based on images of microscopic blood smears, the proposed algorithm for detecting blast cells can be used as an additional tool during Leukemia pre-screening.

5) YOLOV5

YOLOv5 was released just one month after YOLOv4 and outperformed previous YOLO detectors in various ways [90]. YOLOv5 is currently only available as a GitHub repository and has not been published as a peer-reviewed research paper. The primary distinction between YOLOv5 and YOLOv4 is that YOLOv5 adds an anchor box selection procedure to the model [91]. YOLOv5 is the most recent and most

sophisticated version of the YOLO object detection series. YOLOv5 can automatically determine which anchor box is optimal for the training data set and use it [92]. The fact that the YOLOv5 model has been used in a number of applications with positive results is beginning to lend the model credibility, despite the absence of an official paper. The YOLOv5-V6.0 model is the most recent version of the YOLOv5 model, with a faster inference speed of 1666 fps.

YOLOv5 was utilized to detect and count individual blood cells in previous studies [93]. YOLOv5 has a high level of mAP and a quick inference time, allowing it to be used for real-time detection, according to the evaluation results. The YOLOv5 model can perform better than the SSD model. Experiments indicate that the YOLOv5-based method for blood cell detection and counting has the potential to replace and supplement artificial cell counting in clinical settings.

6) SSD

The Single Shot MultiBox Detector (SSD) [94] achieves the same level of precision as advanced two-stage detectors like the Faster R-CNN without sacrificing real-time performance. The SSD is constructed with an additional structure atop the VGG-16 to improve its performance. This extra convolution layer is added at the very end of the model and gradually decreases in size. The shallower grid layer is responsible for adjusting the default grid and aspect ratio, while the deeper grid layer detects small objects when the image characteristics are not too grainy. Similar to Multibox [95], SSD trains the network by associating each ground-truth box with the default box containing the highest overlapping jaccard. SSDs rely on a weighed amount of localization and loss of trust for model training, similar to DPM. To achieve the desired outcome, less than maximum emphasis is applied. SSDs outperform advanced networks such as YOLO and Faster R-CNN in terms of speed and accuracy, but identifying small objects is challenging. This issue was finally resolved by employing a more sophisticated backbone topology, such as ResNet, and making additional adjustments.

SSD analysis was formerly used to detect cervical cancer cells [96]. The SSD network incorporates both positive and negative characteristics to counteract the insensitivity to small objects. Additionally, a center loss function has been added to account for situations in which intra-class differences are greater than inter-class differences. The optimized SSD network proposed achieved greater precision and mAP (mean Average Precision) than YOLO and traditional SSD, respectively. The addition of complementary features improves the network's sensitivity and overall precision. In addition, the proposed SSD network could be used to classify cells for the automatic early detection of cervical cancer.

7) RETINANET

A single-stage detector called RetinaNet is utilized to demonstrate its efficacy; it predicts objects by sampling the input image densely in terms of position, size, and proportion [97]. The primary network is ResNet, which has been enhanced by the Feature Pyramid Network (FPN) and consists of two identical subnets: a classification subnet and a bounding box regressor. Each of the FPN's layers is embedded within a subnet, which enables the network to recognize objects at multiple scales. It is the responsibility of the classification subnet to predict the object scores at each location, while the box regression subnet is responsible for regressing the offset at each anchor back to the truth. Both networks are relatively straightforward, fully connected networks that share similar parameters. This model employs a class-independent bounding box regressor, which was found to be as effective as the more general approach. Training for RetinaNet is straightforward, fast-paced, and simple to implement. In terms of accuracy and processing speed, RetinaNet outperforms two-stage detectors. By introducing new loss functions, RetinaNet has also advanced the state of the art regarding the optimization of object detectors.

Previous research suggested using the RetinaNet deep learning network, which is used to recognize and classify objects in microscopic images, as an alternative automated software-based method for counting blood cells accurately [98]. During network training, red blood cells, white blood cells, and platelets are automatically identified and counted. On examining the smear image, it was determined that the trained model has general capabilities. Comparing the results of testing the performance of the proposed method to those obtained by other authors dealing with cell counting demonstrates that object detection and labeling with RetinaNet can provide an advantage when counting objects.

8) CENTERNET

CenterNet represents objects as points as opposed to the conventional bounding box [99]. The object is approximated as a single point in the center of the bounding box by CenterNet. The FCN processes the input image and generates a heat map with peaks corresponding to the object's center. The feature extractor network is the pre-trained Stacked Hourglass-101 from ImageNet, which has three heads: a heatmap head for detecting object centers, a dimension head for estimating item sizes, and an offset head for correcting object point offsets. During training, the feature extractor receives the loss of multitasking from all three heads. The output head offset is used to calculate the object's position during inference, after which a box is created. Postprocessing does not require non-maximum suppression (NMS) because predictions are points and not bounding boxes. CenterNet offers a novel perspective with a more precise approach and concludes in less time than its predecessors. It offers a high level of accuracy for applications such as 3D object detection, key point estimation, posture, instant segmentation, and orientation detection, etc. However, a different backbone architecture is required because generic architectures that work well with other detectors perform poorly with them and vice versa.

Researchers previously employed Centernet to identify leukocyte subtypes [100]. This study presents a method for

avoiding the identification of small items and model defects. On the ROI sub-images, the enhanced CenterNet model was applied to achieve more precise WBC localization and classification.

9) EFFICIENTDET

EfficientDet advances the concept of a scalable detector by enhancing its precision and efficiency [101]. It adds BiFPN and model scaling, as well as efficient multi-scaling capabilities. BiFPN is a two-way feature pyramid network with learnable weights that can relate any size of input features. This improves NAS-FPN [102], which requires extensive training and has a complex network, by removing one input node and adding an additional lateral connection. This reduces inefficient nodes and increases the fusion of high-level features. EfficientDet offers compounding coefficients that can be employed to enhance all dimensions of backbone networks, BiFPN networks, class/box networks, and resolutions. In contrast to previous detectors that were upgraded with larger, deeper backbones or stacked FPN layers, these detectors have no such enhancements. EfficientDet's backbone network is EfficientNet [103], a feature extraction network comprised of multiple sets of BiFPN layers stacked sequentially. The output of the final BiFPN layer is sent to the class and box prediction network. This model is trained with SGD optimizer and synchronized batch normalization, and it employs vortex activation [104] rather than standard ReLU activation, which is more distinguishable, efficient, and effective. EfficientDet outperforms previous detectors in terms of efficiency and precision while being more compact and less computationally expensive.

EfficientDet has previously been utilized for breast cancer detection using an ultrasound image dataset [105]. EfficientDet was retrained using an exclusive public data set on breast cancer ultrasound and transfer learning strategies. EfficientDet has made significant advances in general image recognition tasks, with particular advantages for locating and identifying tumor regions simultaneously.

VI. CURRENT CHALLENGES AND FUTURE WORK

A. CHALLENGES

Previous research indicates that researchers face a number of obstacles when attempting to detect ALL on blood microscopy images. Utilized datasets and deep learning models pose difficulties for researchers. The dataset containing ALL information, particularly the ALL subtype, is still small and a problem for researchers, as deep learning is effective only with large datasets [28]. The condition of the images in the dataset is another research obstacle that must be overcome [106]. As a result of the presence of overlapping cells, noise, and intra-variation, the model is ineffective for ALL detection. Illumination variations of the cytoplasm and nucleus must be handled with care because they significantly impact the model's performance [107]. In addition, the object detection deep learning model used by previous researchers is so complex that it is inefficient in terms of energy consumption and embedded device utilization [108]. The use of deep learning to detect ALL in microscopic blood images is still far from perfect and leaves a great deal of room for future research advancement.

1) DATASET OF ALL SUBTYPES THAT ARE STILL PUBLICLY LIMITED

The publicly accessible dataset is still dominated by the ALL dataset, which is divided into two classes, healthy and ALL [109]. Previous research on ALL subtype detection utilized private datasets from health institutions or laboratories [19]. Datasets are essential to research because they are used to evaluate the performance of a model. To overcome this, researchers and related institutions must collaborate to make ALL subtype datasets publicly accessible. With the availability of the ALL subtype dataset to the public, there will be a great deal of interest from researchers, resulting in a rapid expansion of the ALL subtype problem domain. In the future, the results can be directly applied to aid physicians or hematologists in analyzing ALL subtypes, allowing for more accurate subtype detection.

2) MORPHOLOGICAL VARIATIONS, DISTRIBUTION, AND CONDITION OF MICROSCOPIC IMAGES OF BLOOD

In the ALL dataset, microscopic image conditions frequently contain a great deal of noise, cells that overlap, and varying illumination and staining [110]. Detecting ALL presents a difficulty for researchers. In order to improve the performance of the model, researchers must first conduct a morphological analysis. This stage has a significant impact on the detection success of ALL [11]. A large number of previous researchers employed diverse methods with variable outcomes [53]. This essential analysis process provides an opportunity for future researchers to establish an efficient and universal method applicable to variations in the morphological appearances of microscopic blood images.

3) THE COMPLEXITY OF THE MODEL IS HIGH

Numerous deep learning models employed by earlier researchers have a high complexity value based on the number of parameters and resources employed [111]. This in itself is a challenge, as not every region or health agency is equipped with computers with high specifications and qualified resources [36]. A simple model with an accuracy result that does not deviate from the state-of-the-art model or even surpasses it becomes a challenge for researchers so that the model has a high level of application in the real world [108]. Several studies have centered on creating models that are easy to use on mobile and embedded devices. However, there is still room for improvement in the resulting performance to surpass the state-of-the-art model in ALL detection.

B. FUTURE WORKS

As previously discussed, researchers have modified the deep learning model multiple times to address the challenges of

ALL detection. Nonetheless, there is room for additional research to enhance performance and make the ALL detection process more precise and potent, as outlined below:

- Collaboration between researchers and health institutions to procure public datasets contains information on ALL subtypes. The dataset is freely available to the public and can be used as a benchmark for model performance.
- The combination of datasets from each ALL dataset in order to reduce model bias in detecting ALL.
- Generation of a new dataset from the previous ALL dataset using the Generative Adversarial Network model, which related experts have validated.
- Modification of the object detection model to reduce its complexity and allow it to be used on embedded devices.
- Further investigation of transformer-based object detection models in order to improve ALL detection performance.

VII. DISCUSSION AND CONCLUSION

This study provides a comprehensive review of the use of a deep learning-based object detection model for detecting ALL subtypes in microscopic blood images. The object detection model and dataset used by previous researchers are reviewed. Papers on the detection or classification of ALL, particularly the ALL subtype, were sought for review in this study from 2018 to 2022. The filtering results include 65 papers from journals and conferences. This study includes a paper conference because there is a lot of valuable information, particularly research contributions that can affect the progress of research on detecting ALL subtypes. The paper review process reveals that there are still many opportunities for further research in ALL subtype detection because few papers specifically discuss ALL subtypes compared to general ALL detection. This study focuses on discussing the detection of ALL subtypes and general ALL detection to gain insight into the models and datasets used by previous researchers. Many models used by researchers are based on two-stage and one-stage detectors. Researchers are motivated to use the one-stage detector model because it has a faster detection capacity, a simpler complexity than the two-stage detector model, and the ability to detect objects approaching object detection in real-time. On the other hand, the two-stage detector model has the advantage of greater accuracy while being slower than the one-stage model. The accumulation of information from previous studies discussed in this paper will aid the research community in determining the research position and next steps to detect ALL subtypes.

The challenges in research on the detection of ALL subtypes motivate researchers to improve the model's performance by making certain modifications or techniques. With the challenge of attracting researchers' interest in producing better models in the past and can be used as a tool to assist hematologists or doctors in diagnosing ALL. The performance of an object detection model based on CNN has dramatically improved for detection of ALL subtypes. Recent developments that use transformers as model compilers may be an alternative choice because transformers have become state-of-the-art models in the field of natural language processing (NLP). The discussion of challenges and future work in this paper can be used as a reference for researchers to contribute to developing specific models or techniques for the detection of ALL subtypes in the future.

REFERENCES

[1] M. A. Alsalem, A. A. Zaidan, B. B. Zaidan, M. Hashim, H. T. Madhloom, N. D. Azeez, and S. Alsyisuf, "A review of the automated detection and classification of acute leukaemia: Coherent taxonomy, datasets, validation and performance measurements, motivation, open challenges and recommendations," *Comput. Methods Programs Biomed.*, vol. 158, pp. 93–112, May 2018, doi: 10.1016/j.cmpb.2018.02.005.

[2] N. M. Deshpande, S. Gite, and R. Aluvalu, "A review of microscopic analysis of blood cells for disease detection with AI perspective," *PeerJ Comput. Sci.*, vol. 7, pp. 1–27, Apr. 2021, doi: 10.7717/peerj-cs.460.

[3] M. O. Aftab, M. J. Awan, S. Khalid, R. Javed, and H. Shabir, "Executing spark BigDL for leukemia detection from microscopic images using transfer learning," in *Proc. 1st Int. Conf. Artif. Intell. Data Anal. (CAIDA)*, Apr. 2021, pp. 216–220, doi: 10.1109/CAIDA51941.2021.9425264.

[4] K. K. Anilkumar, V. J. Manoj, and T. M. Sagi, "Automated detection of B cell and T cell acute lymphoblastic leukaemia using deep learning," *IRBM*, vol. 43, no. 5, pp. 405–413, 2021, doi: 10.1016/j.irbm.2021.05.005.

[5] N. Ouyang, W. Wang, L. Ma, Y. Wang, Q. Chen, S. Yang, J. Xie, S. Su, Y. Cheng, Q. Cheng, L. Zheng, and Y. Yuan, "Diagnosing acute promyelocytic leukemia by using convolutional neural network," *Clinica Chim. Acta*, vol. 512, pp. 1–6, Jan. 2021, doi: 10.1016/j.cca.2020.10.039.

[6] L. O. Moraes, C. E. Pedreira, S. Barrena, A. Lopez, and A. Orfao, "A decision-tree approach for the differential diagnosis of chronic lymphoid leukemias and peripheral B-cell lymphomas," *Comput. Methods Programs Biomed.*, vol. 178, pp. 85–90, Sep. 2019, doi: 10.1016/j.cmpb.2019.06.014.

[7] S. Wang and G. He, "2016 revision to the WHO classification of acute lymphoblastic leukemia," *J. Transl. Internal Med.*, vol. 4, no. 4, pp. 147–149, Dec. 2016, doi: 10.1515/jtim-2016-0040.

[8] D. A. Arber and A. Orazi, "The updated WHO classification of hematological malignancies: The 2016 revision to the WHO classification of myeloid neoplasms and acute leukemia," *Blood J.*, vol. 2, no. 20, pp. 58–71, 2016, doi: 10.1182/blood-2016-03-643544.

[9] L. Boldú, A. Merino, A. Acevedo, A. Molina, and J. Rodellar, "A deep learning model (ALNet) for the diagnosis of acute leukaemia lineage using peripheral blood cell images," *Comput. Methods Programs Biomed.*, vol. 202, Apr. 2021, Art. no. 105999, doi: 10.1016/j.cmpb.2021.105999.

[10] J. Su, J. Han, and J. Song, "A benchmark bone marrow aspirate smear dataset and a multi-scale cell detection model for the diagnosis of hematological disorders," *Comput. Med. Imag. Graph.*, vol. 90, Jun. 2021, Art. no. 101912, doi: 10.1016/j.compmedimag.2021.101912.

[11] T.-C. Yu, W.-C. Chou, C.-Y. Yeh, C.-K. Yang, S.-C. Huang, F. M. Tien, C.-Y. Yao, C.-L. Cheng, M.-K. Chuang, H.-F. Tien, Q.-Y. Zhang, W.-H. Hsu, and S.-C. Chou, "Automatic bone marrow cell identification and classification by deep neural network," *Blood*, vol. 134, p. 2084, Nov. 2019, doi: 10.1182/blood-2019-125322.

[12] P. K. Das and S. Meher, "An efficient deep convolutional neural network based detection and classification of acute lymphoblastic leukemia," *Expert Syst. Appl.*, vol. 183, Nov. 2021, Art. no. 115311, doi: 10.1016/j.eswa.2021.115311.

[13] K. Dese, H. Raj, G. Ayana, T. Yemane, W. Adissu, J. Krishnamoorthy, and T. Kwa, "Accurate Machine-Learning-Based classification of leukemia from blood smear images," *Clin. Lymphoma Myeloma Leukemia*, vol. 21, no. 11, pp. e903–e914, Nov. 2021, doi: 10.1016/j.clml.2021.06.025.

[14] S. Shafique and S. Tehsin, "Acute lymphoblastic leukemia detection and classification of its subtypes using pretrained deep convolutional neural networks," *Technol. Cancer Res. Treat.*, vol. 17, pp. 1–7, Sep. 2018, doi: 10.1177/1533033818802789.

[15] M. A. Hamza, A. A. Albraikan, J. S. Alzahrani, S. Dhahbi, I. Al-Turaiki, M. A. Duhayyim, I. Yaseen, and M. I. Eldesouki, "Optimal deep transfer learning-based human-centric biomedical diagnosis for acute lymphoblastic leukemia detection," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–13, May 2022, doi: 10.1155/2022/7954111.

[16] M. L. Claro, R. D. M. S. Veras, A. M. Santana, L. H. S. Vogado, G. B. Junior, F. N. S. D. Medeiros, and J. M. R. S. Tavares, "Assessing the impact of data augmentation and a combination of CNNs on leukemia classification," *Inf. Sci.*, vol. 609, pp. 1010–1029, Sep. 2022, doi: 10.1016/j.ins.2022.07.059.

[17] L. H. S. Vogado, R. M. S. Veras, F. H. D. Araujo, R. R. V. Silva, and K. R. T. Aires, "Leukemia diagnosis in blood slides using transfer learning in CNNs and SVM for classification," *Eng. Appl. Artif. Intell.*, vol. 72, pp. 415–422, Jun. 2018, doi: 10.1016/j.engappai.2018.04.024.

[18] S. Dasariraju, M. Huo, and S. McCalla, "Detection and classification of immature leukocytes for diagnosis of acute myeloid leukemia using random forest algorithm," *Bioengineering*, vol. 7, no. 4, pp. 1–12, 2020, doi: 10.3390/bioengineering7040120.

[19] J. Laosai and K. Chamnongthai, "Classification of acute leukemia using medical-knowledge-based morphology and CD marker," *Biomed. Signal Process. Control*, vol. 44, pp. 127–137, Jul. 2018, doi: 10.1016/j.bspc.2018.01.020.

[20] S. Mishra, B. Majhi, and P. K. Sa, "Texture feature based classification on microscopic blood smear for acute lymphoblastic leukemia detection," *Biomed. Signal Process. Control*, vol. 47, pp. 303–311, Jan. 2019, doi: 10.1016/j.bspc.2018.08.012.

[21] S. H. Kassani, P. H. Kassani, M. J. Wesolowski, K. A. Schneider, and R. Deters, "A hybrid deep learning architecture for leukemic B-lymphoblast classification," in *Proc. 10th Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, Oct. 2019, pp. 271–276, doi: 10.1109/ICTC46691.2019.8939959.

[22] M. Jawahar, H. Sharen, L. J. Anbarasi, and A. H. Gandomi, "ALNett: A cluster layer deep convolutional neural network for acute lymphoblastic leukemia classification," *Comput. Biol. Med.*, vol. 148, Sep. 2022, Art. no. 105894, doi: 10.1016/j.compbiomed.2022.105894.

[23] P. Rastogi, K. Khanna, and V. Singh, "LeuFeatx: Deep learning–based feature extractor for the diagnosis of acute leukemia from microscopic images of peripheral blood smear," *Comput. Biol. Med.*, vol. 142, Mar. 2022, Art. no. 105236, doi: 10.1016/j.compbiomed.2022.105236.

[24] K. K. Anilkumar, V. J. Manoj, and T. M. Sagi, "Automated detection of leukemia by pretrained deep neural networks and transfer learning: A comparison," *Med. Eng. Phys.*, vol. 98, pp. 8–19, Dec. 2021, doi: 10.1016/j.medengphy.2021.10.006.

[25] A. Acevedo, S. Alférez, A. Merino, L. Puigví, and J. Rodellar, "Recognition of peripheral blood cell images using convolutional neural networks," *Comput. Methods Programs Biomed.*, vol. 180, Oct. 2019, Art. no. 105020, doi: 10.1016/j.cmpb.2019.105020.

[26] L. D. C. Magpantay, H. D. Alon, Y. D. Austria, M. P. Melegrito, and G. J. O. Fernando, "A transfer learning-based deep CNN approach for classification and diagnosis of acute lymphocytic leukemia cells," in *Proc. Int. Conf. Decis. Aid Sci. Appl. (DASA)*, Mar. 2022, pp. 280–284, doi: 10.1109/DASA54658.2022.9765000.

[27] R. Khandekar, P. Shastry, S. Jaishankar, O. Faust, and N. Sampathila, "Automated blast cell detection for acute lymphoblastic leukemia diagnosis," *Biomed. Signal Process. Control*, vol. 68, Jul. 2021, Art. no. 102690, doi: 10.1016/j.bspc.2021.102690.

[28] P. K. Das, V. A. Diya, S. Meher, R. Panda, and A. Abraham, "A systematic review on recent advancements in deep and machine learning based detection and classification of acute lymphoblastic leukemia," *IEEE Access*, vol. 10, pp. 81741–81763, 2022, doi: 10.1109/ACCESS.2022.3196037.

[29] R. B. Hegde, K. Prasad, H. Hebbar, and I. Sandhya, "Peripheral blood smear analysis using image processing approach for diagnostic purposes: A review," *Biocybern. Biomed. Eng.*, vol. 38, no. 3, pp. 467–480, 2018, doi: 10.1016/j.bbe.2018.03.002.

[30] P. S. Fleming, D. Koletsi, and N. Pandis, "Blinded by PRISMA: Are systematic reviewers focusing on PRISMA and ignoring other guidelines?" *PLoS ONE*, vol. 9, no. 5, May 2014, Art. no. e96407.

[31] M. B. Eriksen and T. F. Frandsen, "The impact of patient, intervention, comparison, outcome (PICO) as a search strategy tool on literature search quality: A systematic review," *J. Med. Library Assoc.*, vol. 106, no. 4, p. 420, Oct. 2018.

[32] P. K. Das and S. Meher, "Transfer learning-based automatic detection of acute lymphocytic leukemia," in *Proc. Nat. Conf. Commun. (NCC)*, Jul. 2021, p. 5, doi: 10.1109/NCC52529.2021.9530010.

[33] C. Matek, S. Krappe, C. Münzenmayer, T. Haferlach, and C. Marr, "Highly accurate differentiation of bone marrow cell morphologies using deep neural networks on a large image data set," *Blood*, vol. 138, no. 20, pp. 1917–1927, Nov. 2021, doi: 10.1182/blood.2020010568.

[34] R. B. Hegde, K. Prasad, H. Hebbar, and B. M. K. Singh, "Comparison of traditional image processing and deep learning approaches for classification of white blood cells in peripheral blood smear images," *Biocybern. Biomed. Eng.*, vol. 39, no. 2, pp. 382–392, Apr. 2019, doi: 10.1016/j.bbe.2019.01.005.

[35] A. I. Shahin, Y. Guo, K. M. Amin, and A. A. Sharawi, "White blood cells identification system based on convolutional deep neural learning networks," *Comput. Methods Programs Biomed.*, vol. 168, pp. 69–80, Jan. 2017, doi: 10.1016/j.cmpb.2017.11.015.

[36] S. Nayaki K., J. Denny, M. M. Rubeena, and J. K. Denny, "Cloud based acute lymphoblastic leukemia detection using deep convolutional neural networks," in *Proc. 2nd Int. Conf. Inventive Res. Comput. Appl. (ICIRCA)*, Jul. 2020, pp. 530–536, doi: 10.1109/ICIRCA48905.2020.9183249.

[37] M. J. Ahmed and P. Nayak, "Detection of lymphoblastic leukemia using VGG19 model," in *Proc. 5th Int. Conf. I-SMAC (IoT Social, Mobile, Analytics Cloud) (I-SMAC)*, Nov. 2021, pp. 716–723, doi: 10.1109/I-SMAC52330.2021.9640955.

[38] A. Genovese, M. S. Hosseini, V. Piuri, K. N. Plataniotis, and F. Scotti, "Histopathological transfer learning for acute lymphoblastic leukemia detection," in *Proc. IEEE Int. Conf. Comput. Intell. Virtual Environ. Meas. Syst. Appl. (CIVEMSA)*, Jun. 2021, pp. 1–6, doi: 10.1109/CIVEMSA52099.2021.9493677.

[39] R. D. Labati, V. Piuri, and F. Scotti, "All-IDB: The acute lymphoblastic leukemia image database for image processing," in *Proc. 18th IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2011, pp. 2045–2048, doi: 10.1109/ICIP.2011.6115881.

[40] R. Gupta, S. Gehlot, and A. Gupta, "C-NMC: B-lineage acute lymphoblastic leukaemia: A blood cancer dataset," *Med. Eng. Phys.*, vol. 103, May 2022, Art. no. 103793, doi: 10.1016/j.medengphy.2022.103793.

[41] K. K. Jha and H. S. Dutta, "Mutual information based hybrid model and deep learning for acute lymphocytic leukemia detection in single cell blood smear images," *Comput. Methods Programs Biomed.*, vol. 179, Oct. 2019, Art. no. 104987, doi: 10.1016/j.cmpb.2019.104987.

[42] K. K. Jha and H. S. Dutta, "Nucleus and cytoplasm–based segmentation and actor-critic neural network for acute lymphocytic leukaemia detection in single cell blood smear images," *Med. Biol. Eng. Comput.*, vol. 58, no. 1, pp. 171–186, Jan. 2020, doi: 10.1007/s11517-019-02071-1.

[43] S. Praveena and S. P. Singh, "Sparse-FCM and deep convolutional neural network for the segmentation and classification of acute lymphoblastic leukaemia," *Biomed. Eng.*, vol. 65, no. 6, pp. 759–773, Nov. 2020, doi: 10.1515/bmt-2018-0213.

[44] C.-W. Wang, S.-C. Huang, Y.-C. Lee, Y.-J. Shen, S.-I. Meng, and J. L. Gaol, "Deep learning for bone marrow cell detection and classification on whole-slide images," *Med. Image Anal.*, vol. 75, Jan. 2022, Art. no. 102270, doi: 10.1016/j.media.2021.102270.

[45] M. R. Reena and P. M. Ameer, "Localization and recognition of leukocytes in peripheral blood: A deep learning approach," *Comput. Biol. Med.*, vol. 126, Nov. 2020, Art. no. 104034, doi: 10.1016/j.compbiomed.2020.104034.

[46] H. Kutlu, E. Avci, and F. Özyurt, "White blood cells detection and classification based on regional convolutional neural networks," *Med. Hypotheses*, vol. 135, Feb. 2020, Art. no. 109472, doi: 10.1016/j.mehy.2019.109472.

[47] S. Gehlot, A. Gupta, and R. Gupta, "SDCT-AuxNet: DCT augmented stain deconvolutional CNN with auxiliary classifier for cancer diagnosis," *Med. Image Anal.*, vol. 61, Apr. 2020, Art. no. 101661, doi: 10.1016/j.media.2020.101661.

[48] J. Laosai and K. Chamnongthai, "Deep-Learning-Based acute leukemia classification using imaging flow cytometry and morphology," in *Proc. Int. Symp. Intell. Signal Process. Commun. Syst. (ISPACS)*, Nov. 2018, pp. 427–430, doi: 10.1109/ISPACS.2018.8923175.

[49] S. N. M. Safuan, M. R. M. Tomari, W. N. W. Zakaria, N. Othman, and N. S. Suriani, "Computer aided system (CAS) of lymphoblast classification for acute lymphoblastic leukemia (ALL) detection using various pre-trained models," in *Proc. IEEE Student Conf. Res. Develop. (SCOReD)*, Sep. 2020, pp. 411–415, doi: 10.1109/SCOReD50371.2020.9251000.

[50] P. K. Das, B. Nayak, and S. Meher, "A lightweight deep learning system for automatic detection of blood cancer," *Measurement*, vol. 191, Mar. 2022, Art. no. 110762, doi: 10.1016/j.measurement.2022.110762.

[51] A. Bochkovskiy, C.-Y. Wang, and H.-Y. Mark Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.

[52] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

[53] A. R. Revanda, C. Fatichah, and N. Suciati, "Classification of acute lymphoblastic leukemia on white blood cell microscopy images based on instance segmentation using mask R-CNN," *Int. J. Intell. Eng. Syst.*, vol. 15, no. 5, pp. 625–637, 2022, doi: 10.22266/ijies2022.1031.54.

[54] N. S. Fatonah, H. Tjandrasa, and C. Fatichah, "Identification of acute lymphoblastic leukemia subtypes in touching cells based on enhanced edge detection," *Int. J. Intell. Eng. Syst.*, vol. 13, no. 4, pp. 204–215, 2020, doi: 10.22266/IJIES2020.0831.18.

[55] S. S. A. Zaidi, M. S. Ansari, A. Aslam, N. Kanwal, M. Asghar, and B. Lee, "A survey of modern deep learning based object detection models," *Digit. Signal Process., Rev. J.*, vol. 126, Jun. 2022, Art. no. 103514, doi: 10.1016/j.dsp.2022.103514.

[56] L. Du, R. Zhang, and X. Wang, "Overview of two-stage object detection algorithms," *J. Phys., Conf. Ser.*, vol. 1544, no. 1, May 2020, Art. no. 012033, doi: 10.1088/1742-6596/1544/1/012033.

[57] F. Sultana, A. Sufian, and P. Dutta, "A review of object detection models based on convolutional neural network," *Adv. Intell. Syst. Comput.*, vol. 1157, pp. 1–16, 2020, doi: 10.1007/978-981-15-4288-6_1.

[58] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16 × 16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.

[59] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *Proc. Eur. Conf. Comput. Vis.*, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds. Cham, Switzerland: Springer, 2020, pp. 213–229, doi: 10.1007/978-3-030-58452-8_13.

[60] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 10012–10022.

[61] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.

[62] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders, "Selective search for object recognition," *Int. J. Comput. Vis.*, vol. 104, no. 2, pp. 154–171, Apr. 2013, doi: 10.1007/s11263-013-0620-5.

[63] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 2, pp. 84–90, Jun. 2012, doi: 10.1145/3065386.

[64] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[65] A. Akselrod-Ballin, L. Karlinsky, S. Alpert, S. Hasoul, R. Ben-Ari, and E. Barkan, "A region based convolutional network for tumor detection and classification in breast mammography," in *Deep Learning and Data Labeling for Medical Applications* (Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)), vol. 10008. Springer, 2016, pp. 197–205, doi: 10.1007/978-3-319-46976-8_21.

[66] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2014, doi: 10.1109/TPAMI.2015.2389824.

[67] A. M. B. Custodio and M. F. Ángel, "Deep learning as a tool for improving efficiency the of glial tumor diagnosis," *Editorial Científica 3Ciencias*, vol. 1, pp. 21–29, 2019, doi: 10.17993/IngyTec.2019.52.

[68] N. Kesav and M. G. Jibukumar, "Efficient and low complex architecture for detection and classification of brain tumor using RCNN with two channel CNN," *J. King Saud Univ., Comput. Inf. Sci.*, vol. 34, no. 8, pp. 6229–6242, Sep. 2022, doi: 10.1016/j.jksuci.2021.05.008.

[69] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 1–9.

[70] K. Talukdar, K. Bora, L. B. Mahanta, and A. K. Das, "A comparative assessment of deep object detection models for blood smear analysis," *Tissue Cell*, vol. 76, Jun. 2022, Art. no. 101761, doi: 10.1016/j.tice.2022.101761.

[71] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 936–944, doi: 10.1109/CVPR.2017.106.

[72] J. Liu, L. Cao, O. Akin, and Y. Tian, "3DFPN-HS: 3D feature pyramid network based high sensitivity and specificity pulmonary nodule detection," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2019, pp. 513–521.

[73] J. Dai, Y. Li, K. He, and J. Sun, "R-FCN: Object detection via region-based fully convolutional networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 1–9.

[74] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[75] B. Luan, Y. Sun, C. Tong, Y. Liu, and H. Liu, "R-FCN based laryngeal lesion detection," in *Proc. 12th Int. Symp. Comput. Intell. Design (ISCID)*, Dec. 2019, pp. 128–131.

[76] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2961–2969.

[77] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1492–1500.

[78] M. A. Al-Masni, M. A. Al-Antari, J. M. Park, G. Gi, T. Y. Kim, P. Rivera, E. Valarezo, S.-M. Han, and T.-S. Kim, "Detection and classification of the breast abnormalities in digital mammograms via regional convolutional neural network," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2017, pp. 1230–1233, doi: 10.1109/EMBC.2017.8037053.

[79] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. 30th IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6517–6525, doi: 10.1109/CVPR.2017.690.

[80] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1026–1034.

[81] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. J. Miller, "Introduction to WordNet: An on-line lexical database," *Int. J. Lexicogr.*, vol. 3, no. 4, pp. 235–244, 1990.

[82] R. Al-Qudah and C. Y. Suen, "Synthetic blood smears generation using locality sensitive hashing and deep neural networks," *IEEE Access*, vol. 8, pp. 102530–102539, 2020, doi: 10.1109/ACCESS.2020.2999349.

[83] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.

[84] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, "Distance-IoU loss: Faster and better learning for bounding box regression," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 7, pp. 12993–13000.

[85] Z. Yao, Y. Cao, S. Zheng, G. Huang, and S. Lin, "Cross-iteration batch normalization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 12326–12335, doi: 10.1109/CVPR46437.2021.01215.

[86] I. Loshchilov and F. Hutter, "SGDR: Stochastic gradient descent with warm restarts," 2016, *arXiv:1608.03983*.

[87] D. Misra, "Mish: A self regularized non-monotonic activation function," 2019, *arXiv:1908.08681*.

[88] C.-Y. Wang, H.-Y. Mark Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, "CSPNet: A new backbone that can enhance learning capability of CNN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 1571–1580, doi: 10.1109/CVPRW50498.2020.00203.

[89] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8759–8768.

[90] G. Jocher, "Ultralytics/Yolov5: V7.0—YOLOv5 SOTA realtime instance segmentation," Zenodo, Nov. 2022, doi: 10.5281/zenodo.7347926.

[91] J. Nelson and J. Solawetz. (2020). YOLOv5 is here. Roboflow. [Online]. Available: https://blog.roboflow.com/yolov5-is-here/

[92] J. Solawetz. (2021). YOLOv5—Complete guide and overview. Roboflow. [Online]. Available: https://blog.roboflow.com/yolov5-v6-0-is-here/
[93] N. Rohaziat, M. R. M. Tomari, and W. N. W. Zakaria, "White blood cells type detection using YOLOv5," in *Proc. IEEE 5th Int. Symp. Robot. Manuf. Autom. (ROMA)*, Aug. 2022, pp. 1–6, doi: 10.1109/ROMA55875.2022.9915690.
[94] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 9905, Dec. 2016, pp. 21–37, doi: 10.1007/978-3-319-46448-0_2.
[95] D. Erhan, C. Szegedy, A. Toshev, and D. Anguelov, "Scalable object detection using deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2147–2154.
[96] D. Jia, J. Zhou, and C. Zhang, "Detection of cervical cells based on improved SSD network," *Multimedia Tools Appl.*, vol. 81, no. 10, pp. 13371–13387, Apr. 2022, doi: 10.1007/s11042-021-11015-7.
[97] X. Zhou, D. Wang, and P. Krähenbühl, "Objects as points," 2019, *arXiv:1904.07850*.
[98] G. Drałus, D. Mazur, and A. Czmil, "Automatic detection and counting of blood cells in smear images using RetinaNet," *Entropy*, vol. 23, no. 11, p. 1522, Nov. 2021.
[99] A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 483–499.
[100] X. Zheng, P. Tang, L. Ai, D. Liu, Y. Zhang, and B. Wang, "White blood cell detection using saliency detection and CenterNet: A two?stage approach," *J. Biophotonics*, 2022, Art. no. e202200174. [Online]. Available: https://onlinelibrary.wiley.com/action/showCitFormats?doi=10.1002%2Fjbio.202200174, doi: 10.1002/jbio.202200174.
[101] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and efficient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 10778–10787, doi: 10.1109/CVPR42600.2020.01079.
[102] G. Ghiasi, T.-Y. Lin, and Q. V. Le, "NAS-FPN: Learning scalable feature pyramid architecture for object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 7036–7045.
[103] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. 36th Int. Conf. Mach. Learn. (ICML)*, Jun. 2019, pp. 10691–10700.
[104] P. Ramachandran, B. Zoph, and Q. V. Le, "Searching for activation functions," 2017, *arXiv:1710.05941*.
[105] R. Du, Y. Chen, T. Li, L. Shi, Z. Fei, and Y. Li, "Discrimination of breast cancer based on ultrasound images and convolutional neural network," *J. Oncol.*, vol. 2022, Mar. 2022, Art. no. 7733583, doi: 10.1155/2022/7733583.
[106] A. M. Patil, M. D. Patil, and G. K. Birajdar, "White blood cells image classification using deep learning with canonical correlation analysis," *IRBM*, vol. 42, no. 5, pp. 378–389, Oct. 2021, doi: 10.1016/j.irbm.2020.08.005.
[107] S. N. M. Safuan, M. R. M. Tomari, and W. N. W. Zakaria, "White blood cell (WBC) counting analysis in blood smear images using various color segmentation methods," *Meas., J. Int. Meas. Confed.*, vol. 116, pp. 543–555, Feb. 2018, doi: 10.1016/j.measurement.2017.11.002.
[108] Y. F. D. De Sant' Anna, J. E. M. De Oliveira, and D. O. Dantas, "Lightweight classification of normal versus leukemic cells using feature extraction," in *Proc. IEEE Symp. Comput. Commun. (ISCC)*, Sep. 2021, pp. 1–7, doi: 10.1109/ISCC53001.2021.9631499.
[109] H. Li, X. Zhao, A. Su, H. Zhang, J. Liu, and G. Gu, "Color space transformation and multi-class weighted loss for adhesive white blood cell segmentation," *IEEE Access*, vol. 8, pp. 24808–24818, 2020, doi: 10.1109/ACCESS.2020.2970485.
[110] A. Anita and A. Yadav, "An intelligent model for the detection of white blood cells using artificial intelligence," *Comput. Methods Programs Biomed.*, vol. 199, 2021, Art. no. 105893, doi: 10.1016/j.cmpb.2020.105893.
[111] A. Genovese, "ALLNet: Acute lymphoblastic leukemia detection using lightweight convolutional networks," in *Proc. IEEE 9th Int. Conf. Comput. Intell. Virtual Environ. Meas. Syst. Appl. (CIVEMSA)*, Jun. 2022, pp. 1–6, doi: 10.1109/civemsa53371.2022.9853691.
[112] D. R. Putri, A. Jamal, and A. A. Septiandri, "Acute lymphoblastic leukemia classification in nucleus microscopic images using convolutional neural networks and transfer learning," in *Proc. 2nd Int. Conf. Artif. Intell. Data Sci. (AiDAS)*, Sep. 2021, pp. 1–6, doi: 10.1109/AiDAS53897.2021.9574176.
[113] P. Mathur, M. Piplani, R. Sawhney, A. Jindal, and R. R. Shah, "Mixup multi-attention multi-tasking model for early-stage leukemia identification," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*. College Park, MD, USA: Univ. of Maryland, May 2020, pp. 1045–1049.
[114] M. N. V. Akanksha, S. Bachu, M. A. Kumar, and N. U. Kumar, "Implementation of potential leukemia detection using recurrent neural networks with blood cell counting," in *Proc. 1st Int. Conf. Electr., Electron., Inf. Commun. Technol. (ICEEICT)*, Feb. 2022, pp. 1–5, doi: 10.1109/ICEEICT53079.2022.9768569.
[115] A. Ullah, "Image analysis of cells acute lymphoblastic leukemia using ensemble learning of deep bagging," in *Proc. IEEE 7th Int. Conf. Comput., Eng. Design (ICCED)*, Aug. 2021, pp. 1–4, doi: 10.1109/ICCED53389.2021.9664867.
[116] A. Genovese, M. S. Hosseini, V. Piuri, K. N. Plataniotis, and F. Scotti, "Acute lymphoblastic leukemia detection based on adaptive unsharpening and deep learning," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2021, pp. 1205–1209, doi: 10.1109/ICASSP39728.2021.9414362.

TANZILAL MUSTAQIM (Member, IEEE) received the bachelor's degree in computer science from Universitas Negeri Semarang. He is currently pursuing the master's and Ph.D. degrees with the Institut Teknologi Sepuluh Nopember. He has 12 journal articles and conference papers in computer science. His research interests include computer vision, deep neural networks, machine learning, and artificial intelligence.

CHASTINE FATICHAH (Member, IEEE) received the Ph.D. degree from the Tokyo Institute of Technology, Japan, in 2012. She is currently an Associate Professor with the Institut Teknologi Sepuluh Nopember, Surabaya. She has published more than 110 computer science-related journal articles and conference papers. Her research interests include artificial intelligence, image processing, and data mining.

NANIK SUCIATI (Member, IEEE) received the master's degree in computer science from the University of Indonesia, Indonesia, in 1998, and the Ph.D. degree in information engineering from Hiroshima University, Japan, in 2010. She is currently an Associate Professor with the Department of Informatics, Institut Teknologi Sepuluh Nopember. She has published more than 50 journal articles and conference papers on computer science. Her research interests include computer vision, computer graphics, and artificial intelligence.