Enhanced Detection of Acute Lymphocytic Leukemia Using Deep Learning and Hybrid Classifiers on Microscopic Blood Images

H. A. El Shenbary*, Amr T. A. Elsayed, Khaled A. A. Khalaf Allah, Belal Z. Hassan
Department of Mathematics-Faculty of Science, Al-Azhar University Cairo, Egypt

***Abstract*—There is no doubt that a significant number of individuals worldwide suffer from blood cancer. A lot of people are unaware of the dangers associated with this disease, which can be fatal. When diagnosed, patients may feel intense fear and a sense of powerlessness. In addition, due to the rarity of these diseases, patients often struggle to find the necessary help and information. A specific type of blood cancer called acute lymphocytic leukemia (ALL) mainly affects white blood cells and is particularly prevalent in children. Early detection of this disease will improve the chances of recovery. Therefore, it is crucial to have an accurate and dependable method for identifying blood cancers. Deep learning (DL) architectures have garnered significant interest within the computer vision realm. Recently, there has been a strong focus on the accomplishments of pretrained architectures in accurately describing or classifying data from various real-world image datasets. Classification performances of the proposed models are investigated by using Softmax, Support Vector Machine (SVM), and K-Nearest Neighbors algorithm (K-NN) separately on a deep learning neural network (Alexnet and VGG19) to differentiate between the three types of ALL using microscopic images dataset. The experimental results demonstrate that the combination of Alexnet with SVM achieves outstanding classification performance on the leukemia dataset, particularly on the original(unsegmented) data, achieved 97.03% on bengin class, 96.14% on early class, 99.49% on pre class and 99.9% on pro class. This approach achieves higher accuracy levels than practicing physicians.**

***Keywords*—Deep learning; transfer learning; leukemia; Alexnet; VGG19; SVM; K-NN; classification**

I. Introduction

In the last decades, uncontrolled growth of abnormal cells in the leukemia patients and the bone marrow rapidly expanded [1]. A condition distinct from other illnesses because it typically does not create detectable masses through imaging tests like X-Rays [2]. Hematopoietic stem cells, which give rise to all blood cells, go through multiple stages of development before reaching maturity [3].

In leukemia, a specific type of blood cell undergoes rapid and uncontrolled development, leading to the proliferation of these uncontrollable cells known as leukemia cells [4]. These abnormal cells take over the space within the bone marrow which is known as acute lymphoblastic leukemia (ALL) [5]. It is noted that the early detection of ALL can improve the chance of patient survival. Automatic specialized testing, like cytogenetics, immunophenotyping, and morphological cell categorization, can now detect leukemia. Now, due to its great accuracy, the method of observing blood cells using the mentioned processes is advised. The operator's abilities and level of enervation will determine how tough these procedures are [6]. Various deep learning (DL) techniques have demonstrated significant efficacy across diverse domains including automatic recognition, detection, and segmentation. Examples encompass breast cancer [7], skin lesions [8], [9], [10], brain tumor [11], [12], the COVID-19 pandemic [13], [14], [15], and among other areas [16].

Leukemia detection in human blood samples using microscopic images is only appropriate for low-cost and remote diagnosis systems. New-age approaches can help in this situation [17]. Numerous researchers have created algorithms that offer a simple and incredibly accurate technique to detect and classify various types of blood cancer using deep learning, machine learning, and convolution neural networks (CNN) [18]. Authors in [19], [20] used DL for COVID-19 image recognition and classification, also for object detection and recognition. Early detection is the paramount for successful treatment outcomes. Medical advancements have facilitated various screening methods to identify the disease in its nascent stages. These techniques often involve blood tests to analyze abnormalities in white blood cell counts, red blood cell counts, and platelets. Additionally, bone marrow biopsies may be performed to examine the composition of blood cells directly from the source. Advanced imaging technologies such as Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) scans can also aid in detecting leukemia-related abnormalities within the body. Early detection enables prompt intervention, increasing the likelihood of successful treatment and improved patient prognosis. In this type of cancer, the primary impact is on white blood cells. Particularly in children, this illness is frequently identified, making them more susceptible. Therefore, there is a pressing need for a precise and dependable mechanism for identifying such blood cancers in patients. Authors in [21] introduces a system designed to differentiate between the three distinct types of ALL through the implementation of a CNN. The system offers thorough disease diagnosis along with comprehensive information on symptoms and different treatment stages. Through the utilization of a CNN, the system has demonstrated an accuracy rate exceeding 85% in detecting ALL.

The structure of this paper is organized as follows: Section II presents the key standard techniques used for feature extraction, classification, and the Alexnet deep learning approach. Section III describes the datasets employed for training, validation, and evaluation of the proposed method. Section IV

*Corresponding author.

outlines the proposed method along with its results. Lastly, Section V provides the conclusion and discusses future work.

II. Methods and Materials

A. Deep Learning

A novel area within artificial intelligence is known as Deep learning (DL), which focuses on developing techniques enabling computers to learn complex tasks such as perception with high accuracy. It excels in tasks like image classification, object detection, speech recognition, language processing, and vehicle detection. Unlike traditional classification algorithms that rely on ad-hoc methods for feature extraction from images, DL employs a different approach. Instead of manually defining features, DL methods learn to extract relevant features directly from data, leading to improved performance and reduced false alarms.

Several challenges arise in traditional methods:

1) Difficulty in defining general, reliable characteristics corresponding to specific object types.
2) Complex task of determining the optimal combination of attributes to identify each object type.
3) Challenges in developing efficient techniques for translating, rotating, and scaling objects.

These challenges impact segmentation, classification, and detection processes. DL methods address these challenges by leveraging large labeled datasets to identify relevant features and combinations of characteristics for accurate object categorization. They employ a hybrid feature extraction and classification model, enabling classification of a wide range of unseen objects beyond those included in the training set [22].

Convolutional Neural Networks (CNNs) are a prominent DL architecture that utilizes multiple hidden layers to perform extensive computations on input data. Each layer's output serves as input to the subsequent layer as shown in Fig. 1, facilitating hierarchical feature extraction. The final layer outputs class labels based on the training data. During the training stage, the network learns to predict efficiently, and its performance is evaluated during the prediction phase.

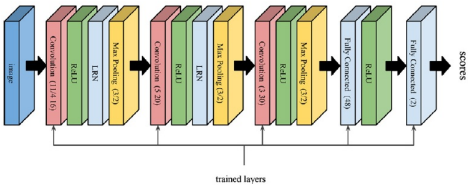

Fig. 1. CNN architecture.

B. Alexnet Deep Learning Neural Network

Alexnet is considered a very important CNN, had a huge performance on the DL Field and it is used in a wide range in computer vision applications [23]. Fig. 2 examines the fundamental structure of Alexnet. It consists of eight layers, with five being convolution layers and the remaining three being fully connected layers. To accelerate the training process, Rectified Linear Unit (ReLU) activation is applied after each convolution and fully connected layer. Dropout is implemented before the first and second fully connected layers. During the training phase, the images are resized to 256 * 256 pixels. A dataset comprising 1.2 million images is utilized for training, while 150,000 images are allocated for testing, and 50,000 images are reserved for network validation.

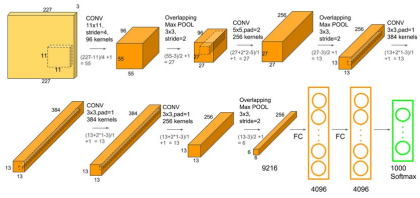

Fig. 2. Alexnet architecture.

C. Visual Geometry Group Deep Learning Neural Network

The Visual Geometry Group (VGG) 19 model, a convolutional neural network (CNN), represents a significant milestone in the field of deep learning and computer vision. Developed by the Visual Geometry Group at the University of Oxford, VGG19 is renowned for its remarkable performance in image classification tasks. With its deep architecture comprising 19 layers, including 16 convolutional layers and 3 fully connected layers, as shown in Fig. 3. VGG19 demonstrates exceptional ability in extracting intricate features from images, enabling accurate categorization across diverse datasets [24].

Fig. 3. VGG19 architecture.

D. Support Vector Machine

Support Vector Machine (SVM) is a powerful supervised machine learning algorithm primarily used for classification tasks, but it can also be employed for regression and outlier detection. SVM aims to find the best possible decision boundary (known as a hyperplane) that separates data points of different classes in such a way that the margin between the two classes is maximized. This is achieved by forming a set of support vectors [25]. SVM works by constructing hyperplanes to separate two classes of raw data for classification [26], [27]. The goal of SVM is to identify the optimal hyperplane, which maximizes the distance between the elements of the training data. For example, if we have two groups of data, finding two hyperplanes, as shown in Fig. 4. It is evident that Fig. 4(b), which has a larger margin, provides better accuracy compared to Fig. 4(a).

E. K-Nearest Neighbors

K-Nearest Neighbors (KNN) is a straightforward, yet effective algorithm for image classification. It classifies an image by considering the majority label of its *k* nearest neighbors in the dataset. Each image is represented as a feature vector,

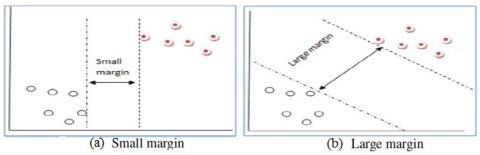

Fig. 4. Support vector machine.

which can be derived from pixel values or higher level features like edges or textures. When a new image is introduced, the algorithm calculates the distance, often using Euclidean distance, between the new image and all images in the training set, determining the *k* closest neighbors. The image is then assigned to the category that appears most frequently among those neighbors. KNN is easy to implement and understand, though its accuracy relies on selecting the right *k* value and distance measure [28], [29].

KNN is particularly well suited for smaller datasets in image classification, since it does not require a training phase, making it a type of lazy learning algorithm. However, its performance declines with larger datasets because of the computational expense of calculating distances for each image in the dataset. Additionally, KNN is sensitive to irrelevant or redundant features, which can reduce its accuracy. Despite these drawbacks, KNN serves as a strong baseline for image classification before turning to more advanced methods like deep learning.

III. Proposed Method and Performance Metrics

A. Proposed Method

Leukemia presents a significant challenge in healthcare due to its complexity and varied subtypes. Traditional diagnostic methods have limitations in accuracy and speed, necessitating the exploration of advanced technologies like deep learning and transfer learning. The proposed method explores the development of a hybrid method incorporating the deep learning architectures Alexnet and VGG19, coupled with transfer learning, and employing classifiers such as SVM and kNN for leukemia prediction. Deep Learning and Transfer Learning: Deep learning models, particularly CNNs, have demonstrated remarkable performance in image classification tasks. Alexnet and VGG19 are popular CNN architectures known for their effectiveness in feature extraction and classification. By leveraging these architectures, we aim to extract meaningful features from leukemia images, facilitating accurate prediction. Transfer learning enhances model performance by transferring knowledge gained from pretrained models to new tasks. In the context of leukemia prediction, transfer learning allows us to leverage the learned features from large datasets like ImageNet and adapt them to the specific characteristics of leukemia images, thereby improving prediction accuracy. The hybrid method begins with pre-processing leukemia images to enhance quality and remove noise. Subsequently, the pre-processed images are fed into the Alexnet and VGG19 architectures for feature extraction. Transfer learning is applied by fine-tuning the pre-trained models on the leukemia dataset, allowing them to adapt to the task at hand while retaining valuable learned features.

Feature vectors extracted from the deep learning models are then inputted into SVM and kNN classifiers for prediction. SVM is chosen for its ability to handle high dimensional data and delineate complex decision boundaries, while kNN is selected for its simplicity and effectiveness in classification tasks. Fig. 5 shows a flow chart of the proposed method. Furthermore, the flexibility of the hybrid method allows for customization and adaptation to different leukemia subtypes, thereby enhancing its applicability in clinical settings.

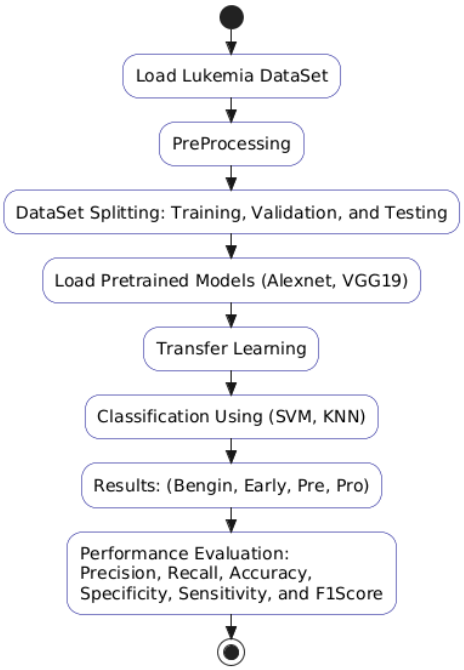

Fig. 5. Flowchart of the proposed method.

B. Performance Metrics

To evaluate the proposed method, a lot of performance parameters were used.

True_positive (T_P), False_positive (F_P), True_negative (T_N) and False_negative (F_N) are the main components used for evaluating the performance metrics. T_P is described as the total number of images predicted as class1 and they are already found in class1. T_N is described as the total number of images predicted as class1 and they are not already found in class1. F_P is described as the total number of images not predicted as class1 and they are already belong to class1. F_N is described as the total number of images incorrectly predicted as class1. The following equations define Recall, Precision, Specificity, Sensitivity, Accuracy, and FScore [see Eq. (1), Eq. (2), Eq. (3), Eq. (4), Eq. (5), and Eq. (6)], respectively.

$$Recall = \frac{T\_P}{T\_P + F\_N}. \tag{1}$$

$$Precision = \frac{T\_P}{T\_P + F\_P}. \tag{2}$$

$$Specificity = \frac{T\_N}{T\_N + F\_P}. \tag{3}$$

$$Sensitivity = \frac{T\_P}{T\_P + F\_N}. \quad (4)$$

$$Accuracy = \frac{T\_P + T\_N}{T\_P + T\_N + F\_P + F\_N}. \quad (5)$$

$$FScore = \frac{2 * T\_P}{(2 * T\_P) + (F\_P + F\_N)}. \quad (6)$$

C. Dataset

A precise diagnosis of acute lymphoblastic leukemia (ALL), a malignancy with a high prevalence, necessitates invasive, costly, and time consuming diagnostic procedures. Peripheral blood smear (PBS) pictures used for ALL diagnosis are crucial in the first cancer screening process to distinguish cancer cases from non-cancer cases. Due to the non-specific character of ALL indications and symptoms, misinterpretation is a common concern when these PBS images are analyzed by laboratory users. The bone marrow laboratory at Taleqani Hospital in Tehran, Iran, is where the photographs for this collection were created. This dataset included 3256 PBS pictures from 89 individuals who were thought to be ALL and whose blood samples were properly processed and stained by knowledgeable laboratory personnel. This dataset is separated into the benign and malignant classes. Hematogones are included in the former. The Early Pre-B, Pre-B, and Pro-B ALL subtypes of malignant lymphoblasts, which make up the latter ALL group. All of the photos were taken with a Zeiss camera at a 100x magnification in a microscope, and they were all saved as JPG files. The specific types and subtypes of these cells were identified by a professional using the flow cytometry instrument. We also offer segmented photos after color thresholding-based segmentation in the HSV color space [30].

Fig. 6 shows a sample of the original dataset images from all of the mentioned classes types. Fig. 7 view a sample of

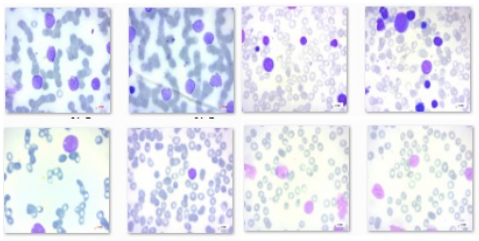

Fig. 6. Sample of the original dataset.

the segmented dataset images from all of the mentioned class types.

IV. Results and Discussion

A. Alexnet and SVM

The results presented in the Table I reflect the performance of the Alexnet combined with SVM on the original leukemia dataset, evaluated across four different classes: Benign, Early, Pre, and Pro. Overall, the model demonstrates outstanding classification performance, with all classes achieving 99% or

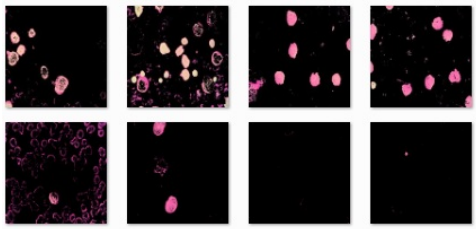

Fig. 7. Sample of the segmented dataset.

100% accuracy. The Pro class exhibits perfect classification with 100% across all metrics, indicating the model's high confidence and consistency in identifying this class. The Pre and Early stages also show near-perfect performance, with only a 1% deviation in precision or Fscore, suggesting that the model effectively captures relevant patterns even in intermediate stages of leukemia progression.

The Benign class, while still performing excellently, shows slightly lower recall and sensitivity values at 97%, which might indicate a small number of false negatives — cases where benign samples were misclassified. However, its 100% specificity suggests no benign cases were incorrectly identified as other leukemia stages, a critical factor in clinical decision making. These results suggest that the model is particularly reliable for detecting advanced leukemia stages, while still performing robustly on earlier and non-cancerous samples. Fig. 8 shows the confusion matrix results for the desired method. The high performance across all metrics and classes highlights the suitability of Alexnet with SVM as a strong classification pipeline for medical image analysis in leukemia detection.

TABLE I. Performance Metrics of Alexnet and SVM on Original Data

Class	Recall	Precision	Accuracy	Specificity	Sensitivity	Fscore
Benign	97%	98%	99%	100%	97%	98%
Early	99%	99%	99%	100%	99%	99%
Pre	100%	99%	100%	100%	100%	99%
Pro	100%	100%	100%	100%	100%	100%

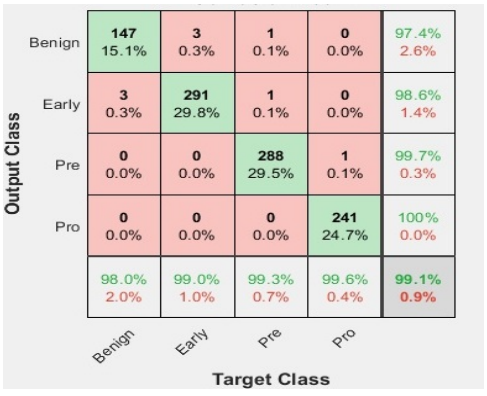

Fig. 8. Confusion matrix of Alexnet and SVM on original data.

When applied to the segmented leukemia dataset, the Alexnet with SVM model continues to deliver strong classification performance, though with slightly lower metrics compared to the original unsegmented data, as shown in Table II. The

Pro class still achieves excellent results, with perfect precision and specificity, and a high Fscore of 98.1%, indicating that segmentation did not significantly impact the model's ability to detect advanced leukemia. Fig. 9 shows the confusion matrix on the segmented data for the same method. The Pre and Early classes also maintain high accuracy and balanced sensitivity and precision, suggesting reliable detection of intermediate disease stages. However, the Benign class shows a noticeable drop in both precision (89.0%) and Fscore (90.2%), possibly due to segmentation affecting the distinctiveness of benign features. Overall, while segmentation may introduce some variability in early and benign classifications, the model remains highly effective, particularly in identifying more advanced leukemia stages.

TABLE II. PERFORMANCE METRICS OF ALEXNET AND SVM ON SEGMENTED DATA

Class	Recall	Precision	Accuracy	Specificity	Sensitivity	Fscore
Benign	91.4%	89.0%	96.9%	97.9%	91.4%	90.2%
Early	94.9%	96.2%	97.3%	98.4%	94.9%	95.6%
Pre	98.3%	95.3%	98.1%	98.0%	98.3%	96.8%
Pro	96.3%	100.0%	99.1%	100.0%	96.3%	98.1%

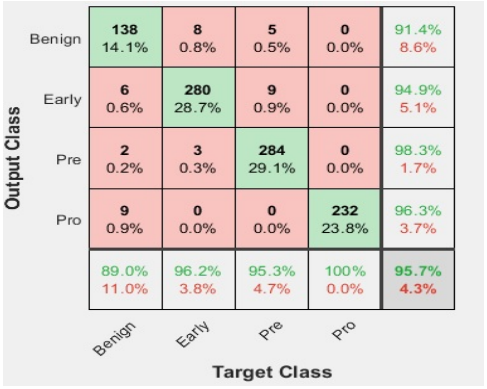

Fig. 9. Confusion matrix of Alexnet and SVM on segmented data.

B. Transfer Learning of Alexnet

Table III shows results from applying transfer learning using Alexnet on the leukemia dataset demonstrate highly effective classification across all four classes. The model achieves consistently high recall, precision, and Fscore values, with accuracy ranging from 98.36% to 99.08%. The Pro class stands out with the highest overall accuracy (99.08%) and excellent precision and Fscore, indicating that the fine-tuned Alexnet is particularly effective in detecting advanced leukemia. The Pre and Early classes also show strong and balanced performance, reflecting the model's ability to generalize across different stages of the disease. Fig. 10 shows the confusion matrix of the transfer learning of Alexnet. Even the Benign class, which is typically more challenging to classify due to subtle features, achieves a solid 94.08% recall and 95.33% precision. These results highlight the strength of transfer learning in leveraging pre-trained features from Alexnet, improving classification accuracy while minimizing training time and data requirements in medical image analysis.

When transfer learning using Alexnet is applied to the segmented leukemia dataset, the model maintains strong overall performance, particularly in detecting the Pro class, which

TABLE III. PERFORMANCE METRICS OF TRANSFER LEARNING OF ALEXNET ON ORIGINAL DATA

Class	Recall	Precision	Accuracy	Specificity	Sensitivity	Fscore
Benign	94.08%	95.33%	98.36%	99.15%	94.08%	94.70%
Early	97.28%	97.28%	98.36%	98.83%	97.28%	97.28%
Pre	98.26%	97.24%	98.67%	98.84%	98.26%	97.75%
Pro	97.94%	98.35%	99.08%	99.45%	97.94%	98.14%

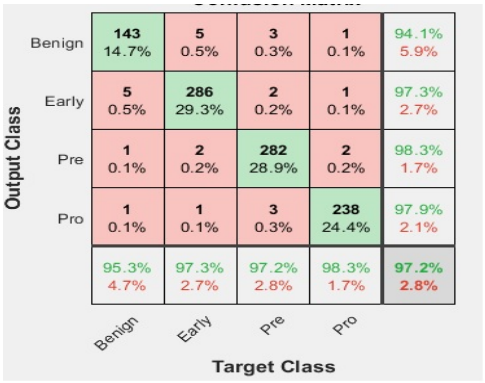

Fig. 10. Confusion matrix of transfer learning using Alexnet on original data.

achieves the highest accuracy (99.28%) and Fscore (98.54%), as shown in Table IV. However, a noticeable performance drop is observed in the Benign class, where recall decreases to 82.78%, indicating that the model struggles more to identify benign samples correctly after segmentation. Despite this, its high precision (90.58%) and specificity (98.42%) suggest that false positives are minimal. The Early and Pre classes show balanced and reliable metrics, with Fscores above 93%, demonstrating that the model remains effective for intermediate disease stages. These results suggest that while segmentation may introduce challenges in distinguishing non-cancerous cases, transfer learning with Alexnet still provides robust classification, especially for more advanced stages of leukemia, as shown in Fig. 11.

TABLE IV. PERFORMANCE METRICS OF TRANSFER LEARNING OF ALEXNET ON SEGMENTED DATA

Class	Recall	Precision	Accuracy	Specificity	Sensitivity	Fscore
Benign	82.78%	90.58%	96.00%	98.42%	82.78%	86.51%
Early	93.22%	92.91%	95.80%	96.92%	93.22%	93.06%
Pre	95.85%	91.12%	96.00%	96.07%	95.85%	93.42%
Pro	97.93%	99.16%	99.28%	99.73%	97.93%	98.54%

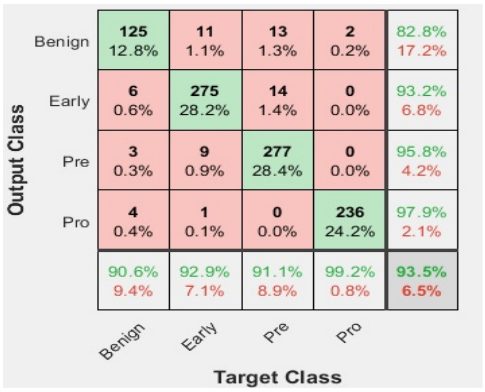

Fig. 11. Confusion Matrix of transfer learning of Alexnet on segmented data.

C. Alexnet

The performance of Alexnet on the original leukemia dataset shows a strong and promising classification ability, particularly for more advanced disease stages. The Pro class achieves perfect recall (100%) and nearly perfect precision (99.59%), with an overall accuracy of 99.90%, demonstrating that the model reliably identifies all Pro samples with minimal error. Similarly, the Pre class also performs exceptionally well, with a near-perfect Fscore of 99.13%, indicating balanced and accurate detection at this stage. The Early class, while slightly lower, still maintains a strong performance with a recall of 95.59% and an Fscore of 94.16%, showing that the model is effective even at earlier stages of leukemia, as shown in Table V and Fig. 12.

However, the Benign class displays comparatively lower performance, with a recall of 87.42%, suggesting a higher rate of false negatives, some benign cases are being misclassified as leukemia. However, the high precision (92.96%) and specificity (98.79%) imply that when the model predicts a case as benign, it is likely correct. This pattern is consistent with the general challenge in distinguishing benign from early pathological features in medical imaging. Overall, Alexnet demonstrates excellent performance on the original data, especially in identifying disease progression stages, making it a reliable model for the classification of leukemia.

TABLE V. PERFORMANCE METRICS OF ALEXNET ON THE ORIGINAL DATA

Class	Recall	Precision	Accuracy	Specificity	Sensitivity	Fscore
Benign	87.42%	92.96%	97.03%	98.79%	87.42%	90.10%
Early	95.59%	92.76%	96.41%	96.77%	95.59%	94.16%
Pre	98.96%	99.31%	99.49%	99.71%	98.96%	99.13%
Pro	100.00%	99.59%	99.90%	99.86%	100.00%	99.79%

Fig. 12. Alexnet confusion matrix results on the original data.

The results of applying Alexnet to the segmented leukemia dataset demonstrate high and consistent classification performance across all classes, with particularly strong results for the Pro class. The Pro class achieves the highest metrics overall, including 98.76% for recall, precision, sensitivity, and Fscore, along with an impressive accuracy of 99.39%, indicating the model's excellent ability to detect advanced-stage leukemia even after segmentation. The Pre and Early classes also show balanced performance, with Fscores of 96.68% and 95.77%, respectively, reflecting the model's robustness in identifying intermediate stages with high reliability as shown in Table VI and Fig. 13.

The Benign class, which is often more challenging to differentiate due to its less distinct features, records a recall of 92.05% and a slightly lower precision of 89.68%. While these values are still strong, the slightly reduced precision suggests occasional misclassification of benign samples as leukemia. However, the specificity remains high at 98.06%, indicating the model rarely mislabels diseased samples as benign. Overall, segmentation does not appear to significantly hinder Alexnet's performance and may even enhance class separation in some cases, allowing the model to maintain high accuracy and generalization across all stages of leukemia.

TABLE VI. PERFORMANCE METRICS OF ALEXNET ON THE SEGMENTED DATA

Class	Recall	Precision	Accuracy	Specificity	Sensitivity	Fscore
Benign	92.05%	89.68%	97.13%	98.06%	92.05%	90.85%
Early	95.93%	95.61%	97.44%	98.09%	95.93%	95.77%
Pre	95.85%	97.54%	98.05%	98.98%	95.85%	96.68%
Pro	98.76%	98.76%	99.39%	99.59%	98.76%	98.76%

Fig. 13. Alexnet confusion matrix results on the segmented data.

D. Alexnet and KNN

When Alexnet is combined with KNN and applied to the original leukemia dataset, the model demonstrates high classification performance, particularly for advanced leukemia stages. The Pro class achieves perfect recall (100%) and a strong precision of 96.02%, resulting in a high Fscore of 97.97% and overall accuracy of 98.98%. Similarly, the Pre class shows excellent results with 97.23% recall, 99.65% precision, and an Fscore of 98.42%, indicating that the model is highly effective at detecting and distinguishing advanced and intermediate disease stages in the unsegmented data, as shown in Table VII and Fig. 14.

In contrast, the Benign class exhibits lower recall (76.82%) despite very high precision (99.15%) and specificity (99.88%), suggesting that many benign cases are misclassified as leukemia. This imbalance lowers the Fscore to 86.57%, pointing to difficulty in recognizing non-cancerous samples accurately. The Early class also shows a gap between recall (97.29%) and precision (88.04%), reflecting a tendency toward false positives. These results suggest that while the Alexnet and KNN combination is highly reliable for detecting later stages of leukemia in the original dataset, its performance for benign and early-stage classification is limited by class similarity and KNN's sensitivity to overlapping feature distributions.

TABLE VII. PERFORMANCE METRIC OF ALEXNET AND KNN ON THE ORIGINAL DATA

Class	Recall	Precision	Accuracy	Specificity	Sensitivity	Fscore
Benign	76.82%	99.15%	96.31%	99.88%	76.82%	86.57%
Early	97.29%	88.04%	95.18%	94.27%	97.29%	92.43%
Pre	97.23%	99.65%	99.08%	99.85%	97.23%	98.42%
Pro	100.00%	96.02%	98.98%	98.64%	100.00%	97.97%

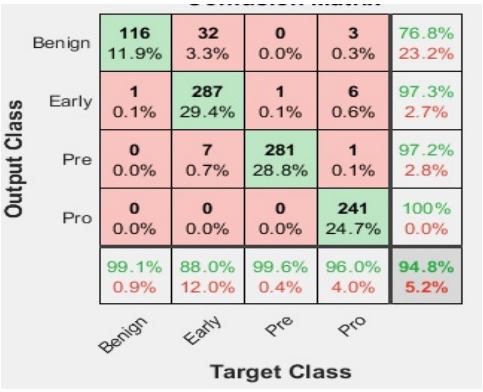

Fig. 14. Confusion matrix results of Alexnet and KNN on the original data.

When applying the combination of Alexnet and KNN to the segmented leukemia dataset, the model continues to perform well overall, particularly in identifying the more advanced disease stages. The Pre and Pro classes show strong and consistent results, with recall values of 97.58% and 97.93%, respectively, and Fscores above 96%. These high metrics demonstrate the model's robustness in capturing critical patterns in segmented images that correspond to later stages of leukemia, even when using a non-parametric classifier like KNN after feature extraction from Alexnet, as shown in Table VIII.

However, the Benign class again poses challenges, as reflected by its low recall of 76.16% and the lowest Fscore (83.33%) among all classes, as shown in Fig. 15. While the precision is relatively high at 92.00%, the model misses a significant portion of benign cases, indicating a tendency to misclassify them as malignant. The Early class shows a more balanced performance, with a recall of 94.92% and an Fscore of 92.26%, suggesting reasonable reliability for early detection. Overall, while segmentation does not drastically improve performance for the benign class, the combination of Alexnet and KNN still yields high classification accuracy and specificity across most classes, particularly for detecting leukemia progression.

TABLE VIII. PERFORMANCE METRIC OF ALEXNET AND KNN ON THE SEGMENTED DATA

Class	Recall	Precision	Accuracy	Specificity	Sensitivity	Fscore
Benign	76.16%	92.00%	95.29%	98.79%	76.16%	83.33%
Early	94.92%	89.74%	95.18%	95.30%	94.92%	92.26%
Pre	97.58%	96.58%	98.26%	98.54%	97.58%	97.07%
Pro	97.93%	95.55%	98.36%	98.50%	97.93%	96.72%

E. VGG19 and SVM

When VGG19 is combined with an SVM classifier and applied to the original leukemia dataset, the results indicate consistently high performance across all classes, with particularly strong outcomes in the more advanced stages of the

Fig. 15. Confusion matrix results of Alexnet and KNN on the segmented data.

disease. The Pro class achieves perfect recall (100%) and a high precision of 98.77%, leading to an excellent Fscore of 99.38% and overall accuracy of 99.69%. Similarly, the Pre class performs exceptionally well with 98.96% recall and an Fscore of 98.79%, showcasing the model's effectiveness in distinguishing leukemia stages with subtle differences, as shown in Table IX.

The model also demonstrates solid performance on the Early and Benign classes. The Early class achieves a recall of 96.95% and a precision of 95.97%, indicating reliable early detection, which is critical for timely treatment. Fig. 16 shows that the Benign class, often harder to distinguish from early malignancies, still achieves a recall of 90.73% and precision of 95.14%, with a strong Fscore of 92.88%. These results reflect the strength of transfer learning with VGG19 in feature extraction and the discriminative power of SVM in handling complex decision boundaries, leading to robust classification performance on unsegmented medical image data.

TABLE IX. PERFORMANCE METRICS OF VGG19 AND SVM ON THE ORIGINAL DATA

Class	Recall	Precision	Accuracy	Specificity	Sensitivity	Fscore
Benign	90.73%	95.14%	97.85%	99.15%	90.73%	92.88%
Early	96.95%	95.97%	97.85%	98.24%	96.95%	96.46%
Pre	98.96%	98.62%	99.28%	99.42%	98.96%	98.79%
Pro	100.00%	98.77%	99.69%	99.59%	100.00%	99.38%

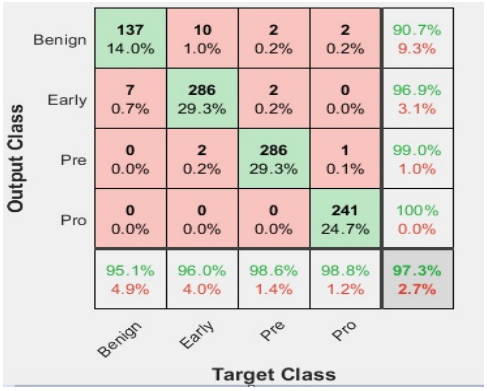

Fig. 16. Confusion matrix of VGG19 and SVM on the original data.

When applying VGG19 with SVM to the segmented leukemia dataset, the model continues to demonstrate strong

classification performance, especially in the advanced classes, as viewed in Fig. 17. The Pro class achieves outstanding metrics with 97.51% recall, 98.74% precision, and a high Fscore of 98.12%, alongside an impressive 99.08% overall accuracy. The Pre class also performs robustly, with 95.50% recall and 97.87% precision, resulting in an Fscore of 96.67%, indicating the model's effectiveness in identifying subtle pathological features within segmented regions, as shown in Table X.

For the Early class, the recall of 95.59% and precision of 91.86% reflect good early-stage detection performance, which is vital for clinical diagnosis and intervention. The Benign class, while showing relatively lower recall (86.09%) compared to other classes, maintains decent precision (87.25%) and a respectable Fscore of 86.67%. Overall, the results indicate that VGG19's deep features, combined with the SVM classifier, are highly capable even on segmented images, although segmentation may slightly affect the model's ability to capture benign patterns as clearly as it does with original images.

TABLE X. PERFORMANCE METRICS OF VGG19 AND SVM ON THE SEGMENTED DATA

Class	Recall	Precision	Accuracy	Specificity	Sensitivity	Fscore
Benign	86.09%	87.25%	95.90%	97.70%	86.09%	86.67%
Early	95.59%	91.86%	96.11%	96.33%	95.59%	93.69%
Pre	95.50%	97.87%	98.05%	99.13%	95.50%	96.67%
Pro	97.51%	98.74%	99.08%	99.59%	97.51%	98.12%

Fig. 17. Confusion matrix results of VGG19 and SVM on the segmented data.

F. VGG19 and KNN

Using VGG19 with KNN on the original leukemia dataset yields high classification performance for the disease-related classes, though with slightly reduced performance for benign cases. The Pro class shows excellent recall (99.59%) and a strong Fscore (96.97%), indicating high sensitivity to advanced stage leukemia detection. Similarly, the Pre class achieves robust results with 97.92% recall and 97.25% precision, leading to a high Fscore of 97.59% and an overall accuracy of 98.57%, as shown in Table XI.

The Early class also performs well, with 96.27% recall and 91.91% precision, producing a solid Fscore of 94.04%. However, the Benign class lags behind with a recall of only 78.15%, despite an excellent precision of 96.72%. This discrepancy suggests that while the model is very confident when it predicts a sample as benign, it may miss a significant number of actual benign cases, potentially due to overlapping visual features with early leukemia stages. Overall, as shown in Fig. 18, the VGG19-KNN combination performs reliably on original data, particularly for detecting leukemia progression stages, but shows some limitations in distinguishing benign samples.

TABLE XI. PERFORMANCE METRICS OF VGG19 AND KNN ON THE ORIGINAL DATA

Class	Recall	Precision	Accuracy	Specificity	Sensitivity	Fscore
Benign	78.15%	96.72%	96.21%	99.52%	78.15%	86.45%
Early	96.27%	91.91%	96.31%	96.33%	96.27%	94.04%
Pre	97.92%	97.25%	98.57%	98.84%	97.92%	97.59%
Pro	99.59%	94.49%	98.46%	98.10%	99.59%	96.97%

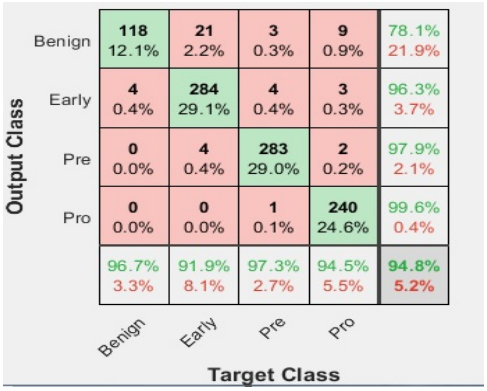

Fig. 18. Confusion matrix results of VGG19 and KNN on original data.

As shown in Table XII, the combination of VGG19 and KNN on the segmented data produces good overall classification results for leukemia detection, particularly for advanced stages, though performance for the Benign class remains relatively low.

The Pro and Pre classes achieve high performance across all metrics. The Pro class attains 96.68% recall and 95.88% precision, resulting in a strong Fscore of 96.28% and an impressive overall accuracy of 98.16%. Similarly, the **Pre** class shows a recall of 95.85% and precision of 96.52%, with an Fscore of 96.18% and overall accuracy of 97.75%, confirming the model's reliability in identifying more developed leukemia stages.

The Early class also performs reasonably well, with 91.86% recall and 88.85% precision, yielding a solid Fscore of 90.33%. However, the Benign class demonstrates noticeably weaker performance, with a recall of just 75.50% and an Fscore of 78.08%. This indicates that while benign samples are occasionally classified correctly, the model tends to confuse them with leukemia stages—likely due to subtle differences in segmented cellular structures.

In summary, as viewed in Fig. 19, VGG19 with KNN on segmented data excels at detecting leukemia stages but continues to struggle with benign classification, reflecting a need for enhanced feature differentiation in non-malignant samples.

G. Transfer Learning of VGG19

The results of applying transfer learning using VGG19 on the original data demonstrate strong performance in classifying

TABLE XII. PERFORMANCE METRICS OF VGG19 AND KNN ON THE SEGMENTED DATA

Class	Recall	Precision	Accuracy	Specificity	Sensitivity	Fscore
Benign	75.50%	80.85%	93.44%	96.73%	75.50%	78.08%
Early	91.86%	88.85%	94.06%	95.01%	91.86%	90.33%
Pre	95.85%	96.52%	97.75%	98.54%	95.85%	96.18%
Pro	96.68%	95.88%	98.16%	98.64%	96.68%	96.28%

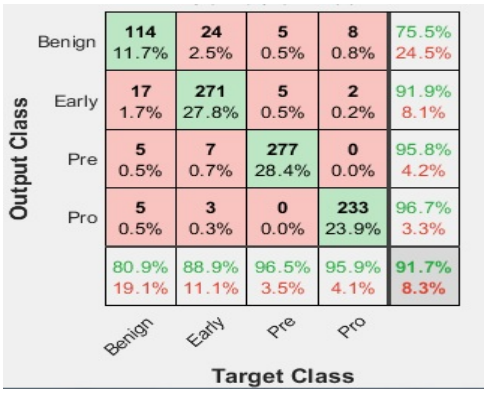

Fig. 19. Confusion matrix results of VGG19 and KNN on the segmented data.

leukemia stages, particularly in advanced cases. The model shows excellent recall and precision for Pre-leukemia and Pro-leukemia classes, with Fscores of 98.25% and 96.59%, respectively, indicating highly reliable detection. The Early stage class also performs very well, achieving a recall of 97.97% and an Fscore of 93.68%. This reflects the model's strong ability to detect various leukemia stages, making it potentially effective for clinical diagnostic support, as shown in Table XIII and Fig. 20.

However, the performance on the Benign class is notably weaker. With a recall of only 73.51%, the model fails to identify a significant portion of benign cases, even though the precision is very high at 98.23%. This indicates that while the model is highly accurate when it predicts a case as benign, it tends to under-predict this class, likely misclassifying benign samples as early-stage leukemia. This imbalance is a concern in clinical settings where distinguishing healthy cases is critical. Overall, the model favors leukemia detection—desirable in screening—but would benefit from further tuning to improve benign classification and reduce unnecessary alarms.

TABLE XIII. PERFORMANCE METRICS OF TRANSFER LEARNING OF VGG19 ON ORIGINAL DATA

Class	Recall	Precision	Accuracy	Specificity	Sensitivity	Fscore
Benign	73.51%	98.23%	95.70%	99.76%	73.51%	84.09%
Early	97.97%	89.75%	96.00%	95.15%	97.97%	93.68%
Pre	97.23%	99.29%	98.98%	99.71%	97.23%	98.25%
Pro	100.00%	93.41%	98.26%	97.69%	100.00%	96.59%

The results in Table XIV, Fig. 21 of applying transfer learning using VGG19 on the segmented data demonstrate consistently high performance in detecting leukemia classes, especially Pre and Pro-leukemia . Both classes exhibit high recall and precision values, with Fscores of 96.90% and 97.10%, respectively. The Early-stage leukemia class also performs well, achieving a recall of 96.61% and an Fscore of 92.83%. These results suggest that VGG19, when fine-tuned on

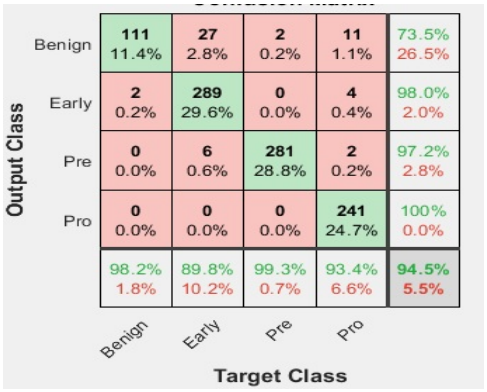

Fig. 20. Confusion matrix results of transfer learning of VGG19 on original data.

segmented data, is particularly effective at identifying various stages of leukemia with minimal misclassification.

However, the Benign class remains a challenge. Although the precision is relatively high at 88.80%, the recall is considerably lower at 73.51%, leading to a modest Fscore of 80.43%. This pattern indicates the model often misclassifies benign samples as leukemia, which could lead to false positives in a clinical setting. Despite this, the overall accuracy remains strong at 94.47%, showing that segmentation helps maintain good classification performance, particularly for critical disease classes, while further optimization is needed to enhance benign detection.

TABLE XIV. PERFORMANCE METRICS OF TRANSFER LEARNING OF VGG19 ON THE SEGMENTED DATA

Class	Recall	Precision	Accuracy	Specificity	Sensitivity	Fscore
Benign	73.51%	88.80%	94.47%	98.30%	73.51%	80.43%
Early	96.61%	89.34%	95.49%	95.01%	96.61%	92.83%
Pre	97.23%	96.56%	98.16%	98.54%	97.23%	96.90%
Pro	97.10%	97.10%	98.57%	99.05%	97.10%	97.10%

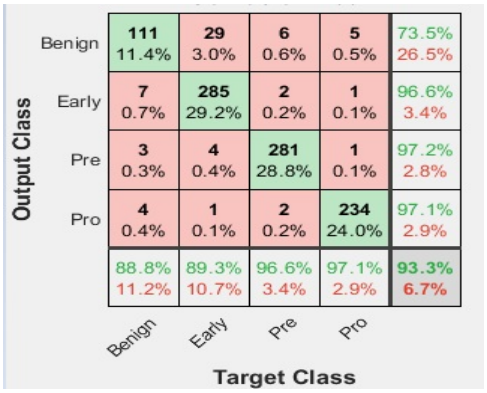

Fig. 21. Confusion matrix results of transfer Learning of VGG19 on the segmented data.

V. CONCLUSION AND FUTURE WORK

The experimental results demonstrate that the combination of Alexnet with SVM achieves outstanding classification performance on the leukemia dataset, particularly on the original (unsegmented) data. This method consistently yields high

accuracy, precision, recall, and Fscores across all leukemia stages, with the Pro and Pre classes often reaching near-perfect metrics. The model's ability to maintain robust performance—even for the challenging Benign class—highlights its effectiveness in distinguishing subtle pathological features. The success of this approach can be attributed to Alexnet's powerful feature extraction capabilities combined with SVM's strong discriminative classification, making it a highly reliable pipeline for leukemia detection in medical imaging.

While other methods, such as transfer learning with VGG19 or hybrid models with KNN, also perform well, they exhibit more variability, particularly in classifying Benign and Early-stage cases. Segmentation introduces additional complexity, sometimes reducing performance, but Alexnet with SVM remains resilient, maintaining high accuracy even on segmented data. These findings suggest that deep feature extraction with a discriminative classifier such as SVM is a superior strategy for leukemia classification, offering a robust and generalizable solution for clinical applications. Future work could explore ensemble techniques or further fine-tuning to enhance benign detection while preserving the model's exceptional performance in identifying advanced leukemia stages.

REFERENCES

[1] S. Mishra, B. Majhi, P. K. Sa, and L. Sharma, "Gray level co-occurrence matrix and random forest based acute lymphoblastic leukemia detection," *Biomedical Signal Processing and Control*, vol. 33, pp. 272–280, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809416302117

[2] M. Zakir Ullah, Y. Zheng, J. Song, S. Aslam, C. Xu, G. D. Kiazolu, and L. Wang, "An attention-based convolutional neural network for acute lymphoblastic leukemia classification," *Applied Sciences*, vol. 11, no. 22, 2021. [Online]. Available: https://www.mdpi.com/2076-3417/11/22/10662

[3] C. Mondal, M. K. Hasan, M. Ahmad, M. A. Awal, M. T. Jawad, A. Dutta, M. R. Islam, and M. A. Moni, "Ensemble of convolutional neural networks to diagnose acute lymphoblastic leukemia from microscopic images," *Informatics in Medicine Unlocked*, vol. 27, p. 100794, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S235291482100263X

[4] N. S. Fatonah, H. Tjandrasa, and C. Fatichah, "Identification of acute lymphoblastic leukemia subtypes in touching cells based on enhanced edge detection," *International Journal of Intelligent Engineering and Systems*, vol. 13, no. 4, p. 204–215, 2020.

[5] C. Di Ruberto, A. Loddo, and L. Putzu, "Detection of red and white blood cells from microscopic blood images using a region proposal approach," *Computers in Biology and Medicine*, vol. 116, p. 103530, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482519303890

[6] A. Tareef, Y. Song, H. Huang, Y. Wang, D. Feng, M. Chen, and W. Cai, "Optimizing the cervix cytological examination based on deep learning and dynamic shape modeling," *Neurocomputing*, vol. 248, pp. 28–40, 2017, neural Networks : Learning Algorithms and Classification Systems. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231217304290

[7] H. MK, A. TA, and R. S., "Automatic mass classification in breast using transfer learning of deep convolutional neural network and support vector machine," *IEEE region 10 symposium*, 2020. [Online]. Available: http://dx.doi.org/10.1109/TENSYMP50017.2020.9230708

[8] "Dermoexpert: Skin lesion classification using a hybrid convolutional neural network through segmentation, transfer learning, and augmentation," *Informatics in Medicine Unlocked*, vol. 28, p. 100819, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352914821002835

[9] A. Dutta, M. K. Hasan, and M. Ahmad, "Skin lesion classification using convolutional neural network for melanoma recognition," *medRxiv*, 2020. [Online]. Available: https://www.medrxiv.org/content/early/2020/11/26/2020.11.24.20238246

[10] H. MK, D. L, S. PN, T. FI, and M. R., "Dsnet: Automatic dermoscopic skin lesion segmentation," *Computers in biology and medicine*, vol. 12, April, 2020. [Online]. Available: https://doi.org/10.1016/j.compbiomed.2020.103738

[11] F. I. Tushar, B. Alyafi, M. K. Hasan, and L. Dahal, "Brain tissue segmentation using neuronet with different pre-processing techniques," pp. 223–227, 2019.

[12] A. Işın, C. Direkoğlu, and M. Şah, "Review of mri-based brain tumor image segmentation using deep learning methods," *Procedia Computer Science*, vol. 102, pp. 317–324, 2016, 12th International Conference on Application of Fuzzy Systems and Soft Computing, ICAFS 2016, 29-30 August 2016, Vienna, Austria. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S187705091632587X

[13] M. K. Hasan, M. T. Jawad, K. N. I. Hasan, S. B. Partha, and M. M. A. Masba, "Covid-19 identification from volumetric chest ct scans using a progressively resized 3d-cnn incorporating segmentation, augmentation, and class-rebalancing," 2021. [Online]. Available: https://europepmc.org/article/PPR/PPR310774

[14] M. K. Hasan, M. A. Alam, L. Dahal, S. Roy, S. R. Wahid, M. T. E. Elahi, R. Martí, and B. Khanal, "Challenges of deep learning methods for covid-19 detection using public datasets," *Informatics in Medicine Unlocked*, vol. 30, p. 100945, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352914822000922

[15] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, and G. Jamalipour Soufi, "Deep-covid: Predicting covid-19 from chest x-ray images using deep transfer learning," *Medical Image Analysis*, vol. 65, p. 101794, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841520301584

[16] M. S. H. Sunny, A. N. R. Ahmed, and M. K. Hasan, "Design and simulation of maximum power point tracking of photovoltaic system using ann," pp. 1–5, 2016.

[17] Y. Song, L. Zhang, S. Chen, D. Ni, B. Li, Y. Zhou, B. Lei, and T. Wang, "A deep learning based framework for accurate segmentation of cervical cytoplasm and nuclei," *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, IEEE*, p. 2903–2906, 2014.

[18] I. Abunadi and E. M. Senan, "Multi-method diagnosis of blood microscopic sample for early detection of acute lymphoblastic leukemia based on deep learning and hybrid techniques," *Sensors*, vol. 22, no. 4, 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/4/1629

[19] H. A. E. Shenbary, E. A. Ebeid, and D. Baleanu, "Covid-19 classification using hybrid deep learning andstandard feature extraction techniques," *Indonesian Journal of Electrical Engineering and Computer Science(IJEECS*, vol. vol.29, No. 3, pp. 1780-1791, 2023. [Online]. Available: https://doi.org/10.11591/ijeecs.v29.i3.pp1780-1791

[20] M. S. Farag, M. M. M. E. Din, and H. A. E. Shenbary, "Parking entrance control using license plate detection and recognition," *Indonesian Journal of Electrical Engineering and Computer Science(IJEECS*, vol. vol.15, No. 1, pp. 476 -483, 2019. [Online]. Available: https://doi.org/10.11591/ijeecs.v15.i1.pp476-483

[21] P. M. Shafi, V. Bidve, H. Bhapkar, P. Dhotre, and V. B. P. Singh, "Leukemia detection system using convolutional neural networks by means of microscopic pictures," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 31, no. 3, pp. 1616–1623, September, 2023. [Online]. Available: https://doi.org/10.11591/ijeecs.v31.i3.pp1616-1623

[22] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," *In Proceedings of the ieee conference on computer vision and pattern recognition*, vol. pp. 580–587, 2014. [Online]. Available: https://doi.org/10.1109/CVPR.2014.81

[23] A. Krizhevsky, S. Ilya, and H. E. Geoffrey, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. pp. 1097-1105, 2012. [Online]. Available: https://doi.org/10.1145/3065386

[24] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556.*, vol. pp. 1-14, 2014. [Online]. Available: https://doi.org/10.48550/arXiv.1409.1556

[25] C. Cortes and V. Vapnik, "Support vector networks," *Machine Learning*, vol. vol. 20, no. 3, pp. 273-297, 1995.

[26] M. M. MohieEl-din, N. I. Ghali, A. G. Ahmed, and H. A. El-Shenbary, "A study on the impact of wavelet decomposition on face recognition methods," *International Journal of Computer Applications (IJCA)*, vol. vol. 87, no. 3, pp. 14 - 21, Feb. 2014. [Online]. Available: https://doi.org/10.5120/15188-3549

[27] F. Bellakhdhar and K. Loukil, "Face recognition approach using gabor wavelets, pca and svm," *International Journal of Computer Science Issues (IJCSI)*, vol. vol. 10, no. 2, pp. 201-207, 2013.

[28] A. Mucherino, P. J. Papajorgji, and P. M. Pardalos, *k-Nearest Neighbor Classification*. New York, NY: Springer New York, 2009, pp. 83–106.

[29] Y. Wang, Z. Chen, H. Shao, and N. Wang, "A knn-based classification algorithm for growth stages of haematococcus pluvialis," in *2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*, vol. 4, 2021, pp. 6–9.

[30] M. Aria, M. Ghaderzadeh, D. Bashash, H. Abolghasemi, F. Asadi, and A. Hosseini, "Acute lymphoblastic leukemia (all) image dataset," *Kaggle*, 2021. [Online]. Available: https://www.kaggle.com/datasets/mehradaria/leukemia/