Blood Cancer Diagnostic System using Deep Learning

[1]Ume-Ezeoke Jenevieve Chika
[2] Nweze, Rosemary Chika
[3]Nkanu, Inyang Unoh
[4]Ogah, Ogbonnaya Sunday
[5]Ugwuanyi Sunday Chinedu

[1]Computer Science Department, Ebonyi State University Abakaliki, Nigeria. genevive0001@gmail.com[1]
[2,5]Computer Science Department, State University of Medical & Applied Sciences, Igbo-Eno, Enugu State, Nigeria. rosemary.nweze@sumas.edu.ng[2], Sunday.ugwuanyi@sumas.edu.ng[5]
[3]ICT Unit, Ministry of Agriculture, Calabar, Cross River State, Nigeria. unoh2007@yahoo.com
[4]Department of Computer Science & Mathematics, Evangel University, Ebonyi State, Nigeria. tedi242010@gmail.com

Abstract

Blood cancer is currently the seventh most common type of cancer and the ninth leading cause of cancer-related deaths in both males and females. In medical diagnosis, time and accuracy are critical factors. Traditional pathological diagnostic methods are often expensive, time-consuming, and prone to human error. As a result, the use of deep learning algorithms for detection and diagnosis is becoming increasingly prevalent due to their high performance and accuracy. We developed a system that effectively diagnoses two types of blood cancer, leukemia and lymphoma. In order to test the accuracy of lymphoma and leukemia diagnosis, other kinds of cells were added into our dataset such as lymphocytes, which are not very easy for doctors to distinguish. Each type of blood cancer has distinct characteristics. Deep learning techniques enable the development of automated and effective systems for their identification. In this study, pre-trained convolutional neural networks (CNNs) including ResNet50 (a 50-layer Residual Network), Google Inception, and VGG16 (a 16-layer model from the Visual Geometry Group) were utilized through transfer learning. These models were fine-tuned using a dataset of blood cell images and trained using an end-to-end method. An ensemble model was developed by averaging the predictions of the three individual classifiers. All models were evaluated, and a comparative analysis was conducted to determine the most effective classifier. This ensemble approach achieved higher accuracy of 91%, sensitivity 100%, f1 score 100% and precision 95% which outperformed the individual models. The final diagnostic system was deployed as an offline application to ensure accessibility for medical personnel.

Keywords: *Blood cancer, Leukemia, Lymphoma, Lymphocytes, Medical diagnosis, Cancer detection, Diagnostic accuracy, Sensitivity, Precision, F1 score*

Technical & Methodological Keywords: *Deep learning, Convolutional neural networks (CNNs), Transfer learning, ResNet50, Google Inception, VGG16, Ensemble model, End-to-end training, Image classification, Model evaluation*

1.0 Introduction

In Watine (2021), medical diagnosis means not only detection or exclusion of disease but also evaluation of disease risk, prognostic assessment and therapeutic monitoring. There was a substantial interest in studying decisions in the early 1970's by experts without the utilization of statistical instruments, and figuring out whether and how such choices could be demonstrated in a personal computer. Specifically, there was an enthusiasm for researching conceptual and symbolic strategies suitable for displaying doctor and other expert decision making (O'Leary, 2013).

As a standard practice in clinical medicine, microscopic examination of peripheral blood plays an important role in the field of diagnosis and control of major diseases. It is uniquely capable of discerning clinically relevant morphologic features of hematopoietic cells, including abnormal white blood cells (WBCs, also known as leukocytes) in lymphoma, leukemia, dysplasia and other diseases (Bain *et al*., 2015). Gold-standard morphologic profiling of blood cells relies heavily on manual smear processing techniques and visual inspection with limitations from quality-control and economic scalability (Durant *et al*., 2017). Blood smear preparation and interpretation are negatively affected by observer bias, slide distribution errors, statistical sampling error, recording error, and also involve labor-intensive processes that require highly skilled technicians. As such, there has been considerable interest in developing systems for automated classification of digital images of peripheral blood smears with high sensitivity and specificity (Ceelie *et al*., 2017). The widespread adoption of electronic medical records contributes to the exponential generation of biomedical data in size, dimension and complexity (Livieris *et al*., 2018).

Due to all the drawbacks of traditional recognition systems, there is an urgent necessity to develop more adaptive and practical high-performance recognition systems.

Casey (2019) views deep learning as an essential branch of Artificial Intelligence (AI) that closely tries to mimic how the human brain works. CNN-motivated deep learning algorithms have become the standard choice for medical image classifications. State-of-the-art CNN-based classification deep learning solutions has the potential not just to aid human experts in image diagnosis but also has the capability to replace human experts (Bennur, 2019).
The deep learning technologies have shown impressive performance in various vision tasks such as image classification, object detection and semantic segmentation. The core of the deep learning technology is automatic feature extraction which is a critical step in image processing. It involves the identification and extraction of significant attributes from an image. In the field of deep learning, Convolutional Neural Networks (CNNs) have achieved excellent performance in image analysis. It makes it possible to avoid complications and error associated with hand-crafted features.

2.0 Related works

Several notable deep learning systems have been developed in recent years. The success of this project will depend greatly on the review of existing literatures on deep learning models especially in the areas of diagnosis.

In Anwar and Alam (2020), a convolutional neural network-based learning approach to acute lymphoblastic leukemia detection with automated feature extraction was proposed. The objective of this study is to detect acute lymphoblastic leukemia using a convolutional neural network model. The dataset includes 368 microscopic blood smeared images and 736 after data augmentation (360 blast cells and 376 non-blast cells). A softmax classifier was used to classify the images into blast cells (leukemia) and non-blast cells (not leukemia). The proposed model yields an outstanding accuracy between 99 and 100% in most of the trials. Although a high accuracy was achieved in the classification task and the model eliminated the need for any segmentation or extraction technique, the model

is trained to diagnose only one type of blood cancer (leukemia). Also, limited dataset was used to train the model which could decrease robustness and increase chances of over-fitting.

Similarly, Brancati *et al*. (2019), a deep learning approach for breast invasive ductal carcinoma detection and lymphoma multi-classification in histological images. The goal of this paper is to address two different use cases: the detection of invasive ductal carcinoma in breast histological images and the classification of lymphoma sub-types using deep learning. For invasive ductal carcinoma, total of 162 whole slide images were obtained which when divided into patches representing IDC and non-IDC is 46, 633 and 124, 011, respectively (public dataset). For the lymphoma multi-classification, a total of 374 images were generated, containing 113, 139 and 122 images of chronic lymphocytic leukemia, follicular lymphoma and mantle cell lymphoma respectively. A Supervised Encoder FusionNet (SEF) method based on a Residual CNN (the encoder of FusionNet) and a softmax classifier was used to address the two use cases. The SEF method significantly outperformed the other deep learning approaches with an overall increment of 4.86%, 5.52%, 3.7% and 2.33% in accuracy as compared to FusionNet, Unet, SEU and ResNet-34, respectively. This study also goes a long way in reducing the inter- and intra-reader variability among pathologists and also accelerates the diagnosis process. Nevertheless, there is the possibility of over-fitting because large dataset was used in training the model and no data augmentation was performed on the datasets. Also, the scope is limited to only one type of blood cancer.

Boldu *et* al. (2021), a deep learning model (ALNet) for the diagnosis of acute leukemia lineage using peripheral blood cell images. This study aims to develop a deep learning based system to predict acute leukemia using blood cell images. A set of 731 blood smears containing 16, 450 single-cell images was analyzed. To find the best architecture for acute leukemia classification, VGG16, ResNet101, DenseNet121 and SENet154 (CNNs) were evaluated. Once chosen, a system with two modules working sequentially was configured (ALNet). The proposed system, ALNet, obtained 94.2% accuracy for myeloid leukemia detection and 89.5% accuracy for lymphoid leukemia. This system can distinguish neoplastic (leukemia) and non-neoplastic (infections) diseases and can assist clinical pathologists in the diagnosis of acute leukemia during blood smear review. Nevertheless, by not sourcing data from different institutions, accuracy could be negatively affected. Also, this study was not extended to cover other kinds of white blood cancers, only leukemia.

Saleem *et al*. (2021), in their paper on deep network designed for the segmentation and classification of leukemia using fusion of the transfer learning models proposes a modified deep learning approach for the accurate segmentation of leukocytes and their classification. Two public datasets were utilized; LISC and ALL-IDB. The LISC dataset includes 250 images while the ALL-IDB contains 367 images. Generative Adversarial Network (GAN) was used to preprocess the datasets. Deep models (Darknet-53 and ShuffleNet) were used for classification of white blood cells. Finally, a CNN (ResNet-18) was used for semantic segmentation. Classification accuracy achieved with ALL-IDB and LISC dataset is 100% and 99.70% for leukocytes whereas semantic segmentation achieved 99.10% and 98.60% for average and global accuracy respectively. The increase in size of datasets by using GAN network with selected learning parameter ultimately increases the performance of classification accuracy but this study was not extended to accommodate other types of blood cancer. Also, the use of limited number of dataset and no data augmentation techniques could lead to over-fitting.

In Kassani (2019), a hybrid deep learning architecture for leukemic B-lymphoblast classification, an automated deep learning-based method was proposed to distinguish between immature leukemic blasts and normal cells. The dataset contains 7272 Acute Lymphoblastic

Leukemia cell images and 3389 normal cell images. Two CNN model architectures (VGG16 and MobileNet) were trained. Prediction accuracy of 96.17% was achieved for leukemic B-lymphoblast classification. By proposing a hybrid CNN model that combines low-level features from intermediate layers, high-level discriminative feature maps can be generated for immature leukemic blast classification. On the other hand, this study was not extended to accommodate other types of blood cancer and the proposed hybrid CNN model requires a lot of computational time and memory space.

Pang *et al*. (2017), a novel end-to-end classifier using domain transferred deep convolutional neural networks for biomedical images. This study proposes an end-to-end classifier for all kinds of biomedical images through deep learning and transfer learning. Data was gotten from three well known public biomedical image databases; NEMA-CT database (696 training images and 95 testing images), TCIA-CT database (744 training images and 117 testing images) and OASIS-MRI database (452 training images and 82 testing images). Domain transferred deep convolutional neural network (DT-DCNN) was applied in classification of biomedical images. The classification accuracy of the proposed CNN model architecture for the NEMA-CT, TCIA-CT and OASIS-MRI datasets are 100%, 100% and 93.767%, respectively. Even though this model shows a highly reliable and accurate performance which has been confirmed on several public datasets, it is domain based and so can only be used to classify a particular disease at a time. Moreover, convergence speed for the OASIS-MRI is slow and reflects in the accuracy and limited number of dataset used which could cause over-fitting.

Hasan *et al*. (2020) proposed an accurate recognition of leukemia sub-types by utilizing a transfer learning deep convolutional neural network. The objective of this research is to analyze a leukemia sub-type dataset which has samples from three sub-types of leukemia. The dataset includes 334 images and 3600 after augmentation. CNN architecture (DenseNet-201) was used for classification and an overall accuracy of 99.56% was realized. From the result, this approach can be employed for precise recognition of leukemia. However, this work was not extended to recognize and classify other types of blood cancer and the dataset used is limited since it is a multi-class classification and could lead to over-fitting.

Shafique and Tehsin (2018) proposed an acute lymphoblastic leukemia (ALL) detection and classification of its subtypes using pre-trained deep convolutional neural networks. This research is aimed at utilizing deep convolutional neural network for automated detection of acute lymphoblastic leukemia and classification of its subtypes into 3 classes. The dataset includes 368 images and softmax was used as the classifier. Sensitivity of 100%, specificity of 98.11% and accuracy of 99.50% was achieved for acute lymphoblastic leukemia detection. Sensitivity of 96.74%, specificity of 99.03% and accuracy of 96.06% was achieved for acute lymphoblastic leukemia subtype classification. This automated diagnostic system can help in early diagnosis of leukemia. It is also able to detect and classify three subtypes of acute lymphoblastic leukemia. This approach made use of one deep learning model for classification. There was no comparison with other deep learning architectures to check which model performs better than the others. Also, it can only diagnose one type of blood cancer and limited number of images in the dataset could lead to over-fitting.

Current automated systems basically depend on traditional machine learning and image processing techniques which involves more computational time, relatively high error rate in terms of recognition and mostly suited for small datasets (Yu *et al*., 2017). In addition, deep learning systems are mostly suited for large data with low recognition error rate. With this current trend in artificial intelligence, it is possible to automatically detect and diagnose diseases using deep learning (Lee *et al*., 2017).

3.0 Deep Learning Models

i. ResNet-50**:** ResNet-50 model is a convolutional neural network (CNN) that is 50 layers deep and stacks residual blocks on top of each other to form a network. It is an innovative neural network that was first introduced in computer vision research paper research titled 'Deep Residual Learning for Image Recognition' (He *et al.,* 2016). Residual Networks (ResNet50) are deep convolutional networks where the basic idea is to skip blocks of convolutional layers by using shortcut connections. It was pre-trained for object detection task on the ImageNet dataset. This model was immensely successful, as can be ascertained from the fact that its ensemble won the top position at the ILSVRC 2015 classification competition with an error of only 3.57%. Additionally, it also came first in the ImageNet detection, ImageNet localization, COCO detection, and COCO segmentation in the ILSVRC & COCO competitions of 2015 (Boesch, 2022). The ResNet-50 is used as the first classifier for this research.

ii. VGG-19: Visual Geometry Group Network (VGGNet) was introduced by Karen Simonyan and Andrew Zisserman from the Visual Geometry Group (VGG) of University of Oxford in 2014 to examine the effect of the depth of a convolutional network on the final classification accuracy. This architecture achieved one of the top performances, with better feature extraction in the ILSVRC 2014 challenge, using a 3 × 3 convolution filter size. In this architecture, small filters increase the depth of the network instead of its width, which plays a critical role in gaining higher performance. There are two versions of this architecture: VGG16 and VGG19, with different depth and layers (Simonyan and Zisserman, 2015).

iii. Inception V3: The InceptionV3 architecture, proposed by Szegedy *et al.*2016, won the ImageNet ILSVRC 2014 competition. It is a very deep network, consisting of 22 layers and having fewer parameters than AlexNet. The Inception module is able to extract multi-level features in parallel, making it considerably more efficient and faster (Szegedy *et al.,* 2016). The Inception-v3 Convolutional Neural Network is used as the third classifier for the research work. The Inception V3 has 48 layers and is trained with more than a million images from the dataset. Inception V3 module has pooling layers and convolutional filters and Softmax activation function. Two Dense layers are added at the end. The fine-tuning parameters are same as the previous architecture VGG-19.

iv. Ensemble: The weights of each image classifiers are saved and are ensembled. In this method, the average predictions of all the three image classifiers are taken to make the final prediction.

According to Kasani *et al.* (2020), transfer learning is a common strategy in deep learning, where a large dataset from a source task is used for training another target task, leading to not only overcoming the problem of small datasets but also accelerating the learning process and improving the accuracy. Transfer learning approach enables one to adopt a pre-trained network that has already learned a rich set of low-level features from layers that are closer to the input image. Though the dataset is not the same, the low-level features produced by the source CNN are mostly of general shapes, for example, edges, contours and curves, which are similar to the low-level features of the target dataset, while high-level features at the final layers concentrate on complex class-level characteristics, which are needed to differentiate between classes. With the use of transfer learning, training of a CNN is a practical strategy with promising results and significantly more cost-effective than training from scratch. Previous studies also showed that transfer learning also has the potential to prevent over-fittings (Lu *et al.,* 2019).

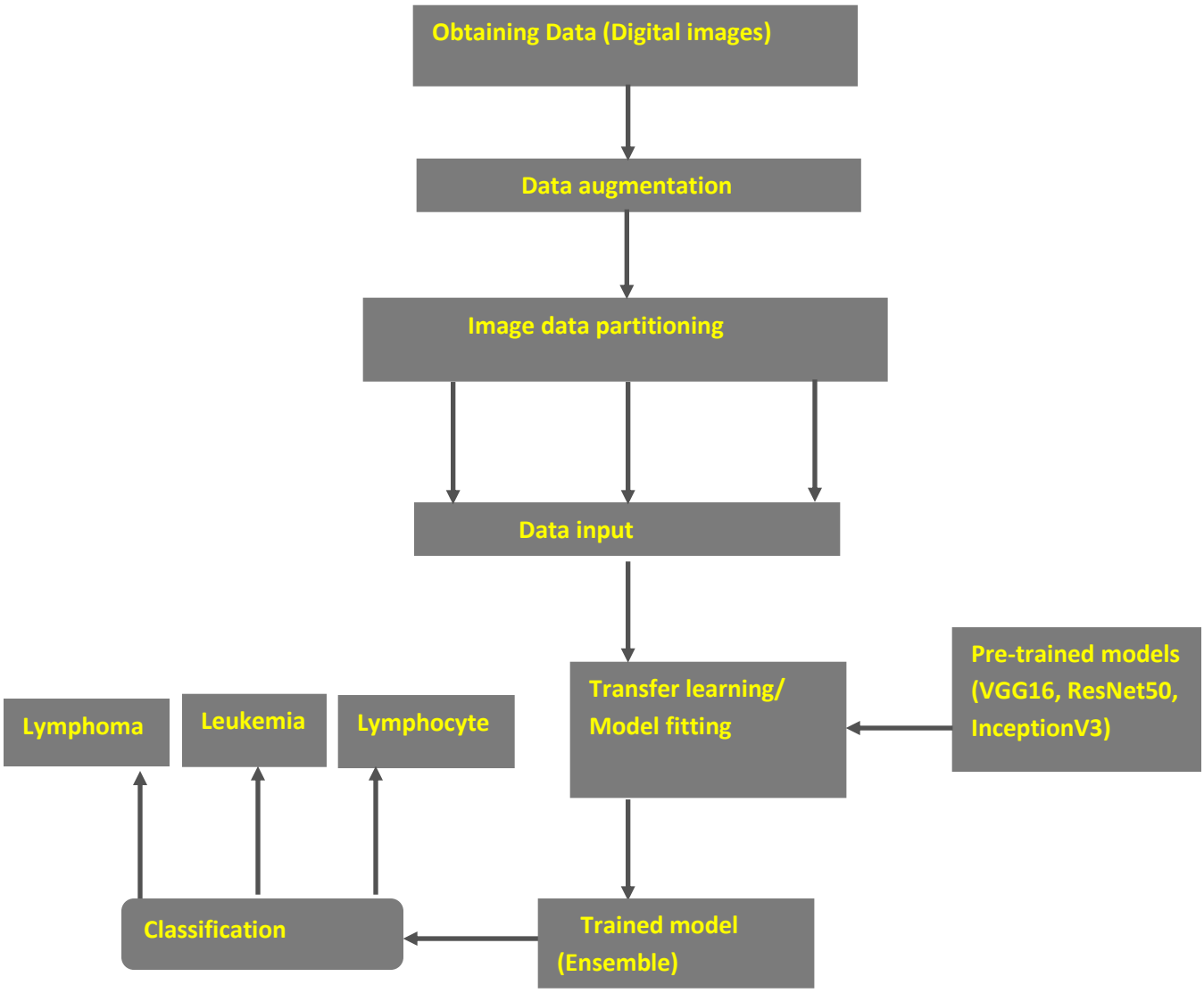

Figure 1: Detailed work flow diagram of blood cancer diagnostic system

4.0 Dataset

The images used in training the deep learning model were gotten from kaggle.com. They were prepared in the bone marrow laboratory of Taleqani Hospital (Tehran, Iran). The acquired data consists of labeled 2231 blood smear microscopic images, categorized into leukemia/ lymphoma/ lymphocyte. Another datasets include 1538 images of malignant cancer and 693 of benign non-cancerous images. The third dataset consists of 3256 PBS images from 89 patients suspected of ALL, whose blood samples were prepared and stained by skillful laboratory staff. The last dataset is divided into two classes of benign and malignant. The former comprises of hematogones and the latter is the ALL group with three subtypes of malignant lymphoblasts: Early Pre-B, Pre-B, and Pro-B ALL. All the images were taken by using a Zeiss camera in a microscope with a 100x magnification and saved as JPG files. A specialist using the flow cytometry tool made the definitive determination of the types and subtypes of these cells.

Three types of malignant lymphoma are represented in the third dataset:

i. CLL (chronic lymphocytic leukemia);

ii. FL (follicular lymphoma);

iii. MCL (mantle cell lymphoma).

The ability to distinguish classes of lymphoma from biopsies sectioned and stained with Hematoxylin/Eosin (H+E) would allow for more consistent and less demanding diagnosis of this disease. Only the most expert pathologists specializing in these types of lymphomas are able to consistently and accurately classify these three lymphoma types from H+E-stained biopsies. The standard practice is to use class-specific probes in order to distinguish these classes reliably. This dataset is a collection of samples prepared by different pathologists at different sites. There is a large degree of staining variation that one would normally expect from such samples.

5.0 Methods

5.1 Data Preprocessing

Some of the images collected could have non-uniform illumination across all regions and some could be blurred due to focusing problem. These problems can be solved by employing pre-processing tasks such as the rolling-ball background subtraction method and using a sharpening filter to increase the contrast for the images (Cai and Verbeek, 2019). Generally, some of the pre-processing tasks performed on blood cancer images include data normalization, data augmentation, scaling, staining, cropping, resizing and image registration. However, not applying these pre-processing tasks on the raw images may distract the classification model which can lead to high misclassification rate (Murtaza *et al*., 2019). This stage is required to increase the overall performance of the model and to avoid over-fitting (Tessema *et al*., 2021).

The performance of deep learning models may degrade due to imbalanced datasets or small numbers of training data. This normally occurs where data dimensionality is much larger than the number of samples or due to an imbalance in the number of training samples of each class. In order to avoid over-fitting of the classification model which usually arise from small sized and/or imbalanced data, different strategies of data augmentation were employed, such as horizontal and vertical flips, contrast adjustments and brightness correction to enlarge the dataset. Employing data augmentation techniques enables the model to yield better performance compared to non-augmented data. Data augmentation is employed as a pre-processing step to avoid over-fitting as well as boost classification performance.

Before the augmented data is further processed, some morphological features of these blood cancer types (lymphoma and leukemia) are first extracted from the images (feature extraction) in Google colaboratory (CoLab) environment which is a free Jupiter notebook interactive development environment. These features include:

i. Contour: For lymphoma, the nuclei are large and often rounded in contour with a prominent, often irregular nuclear membrane, pale chromatin and usually one prominent eosinophilic nucleolus. Mononuclear variants of reed-sternberg cells (lymphoma cells) are characterized with a rounded contour or oblong nucleus with large inclusion-like nucleoli. For leukemia, the nuclei are eccentric and often bilobed with a folded contour or a squishy contour and a reniform nuclear shape.
ii. Color**:** In lymphoma, reed-sternberg cells may have condensed cytoplasm and pyknotic reddish nuclei called mummified cells. However, in leukemia, the blasts are large cells with abundant cytoplasm, light gray to deep blue (purple) with possible rare azurophilic granules.
iii. Size: In lymphoma, classical diagnostic reed-sternberg cells are large (15 to 45 micrometers). In leukemia, blasts are often larger than a lymphocyte and at least the size of a monocyte. They are usually medium to large cells (14 to 18 micrometers).
iv. Edges**:** In the case of lymphoma, reed-sternberg cells appear as large cells of delicate edge with a perinuclear clearing (halo), resembling a viral inclusion. They are also mostly found at the edge of the smear. For leukemia, the edge of the nucleus is rounded with a delicate lacy chromatin with one or more large prominent nucleoli. Most times, they present themselves as large cells with the same chromatin pattern but with convoluted/ indented nuclear shape.

5.2 Feature extraction and Classification:

According to Phung and Rhee (2019), convolutional neural network consists of five different layers: an input layer, a convolution layer (Acevedo *et al*., 2019), a pooling layer, a fully-connected layer and an output layer. These layers are divided into two parts: feature extraction and classification layers. Feature

extraction consists of an input layer, a convolution layer, and a pooling layer, while classification consists of a fully-connected layer and an output layer. Feature extraction facilitates the accurate classification of blood cancer by presenting the dimensionally reduced significant features (Praveena and Singh, 2020). In computer vision, a feature is a measurable piece of data in an image, which is unique to that specific object. It can be a distinct color in an image or a specific shape such as a line, edge, corner, or an image segment. A good feature is used to distinguish objects from one another.

Table 1: Some notable key contributions of different deep learning models for blood cancer classification.

Author (Reference)	Feature extraction method/ classifier	Accuracy
Kasani *et al*., 2020	Ensembled CNN (NASNetLarge & VGG19)	96.58%
Anwar and Afrin, 2020	10 layers CNN	99 - 100%
Mao *et al*., 2016	DCNN	97% f1 score
Brancati *et* al., 2019	SEF based on residual CNN & softmax classifier	97.6%
Boldu *et al*., 2021	CNN (ALNet)	89.5 - 94.2%
Saleem *et al*., 2021	CNN (DarkNet-53, ShuffleNet & ResNet-18)	99.70 - 100%
Kassani, 2019	CNN (VGG16 & MobileNet)	96.17%
Pang *et al*., 2016	5 layers DT-DCNN	93.77 - 100%
Hasan *et al*., 2020	CNN (DenseNet-201)	99.56%
Shafique and Tehsin, 2018	CNN (AlexNet)	96.06 - 99.50%
Genovese *et al*., 2021	CNN (VGG16)	96.84%
Kasani *et al*., 2020	Ensembled CNN (NASNetLarge & VGG19)	96.58%
Ridoy and Islam, 2020	8 layers CNN (LeNet-5)	92.15 - 98.31%
Banik *et al*., 2020	5 Layers CNN	96%
Das and Meher, 2021	CNN (MobilenetV2 and ResNet18).	97.18 – 99.39%
Qiao *et al*., 2021	3 layers CNN	99.21% Sensitivity
Tsykunov *et al*., 2020	CNN (ResNet34).	92.42%
Zhao *et al*., 2020	Lightweight CNN	86.29%
Khandekar *et al*., 2021	CNN (YOLOV4)	95.57 – 98.57% MAP
Anilkumar *et al*., 2021	DCNN (pre-trained CNN; AlexNet and custome-made DL network; LeukNet.)	94.12%

5.3 Model Training

The dataset was split into 3 separate segments for training, validation, and testing. The training data (80%) is for model learning, validation data (10%) is for model evaluation and model parameters tuning. Test data (10%) is for the final evaluation of our model. Deep learning frameworks and libraries were used in the training to achieve good results. The training involves object detection using blood smear images. This was carried out with python language using Keras library which uses Tensorflow at the backend. The data is transformed from 3D to 4D tensor with shape 1, 120, 120, 3. The augmented data is processed through the pretrained image classifiers ResNet-50, Inception V3 and VGG-19. Final layers of all the three image classifiers were fine-tuned by adding two dense layers. Global Average Pooling and softmax were used as the active function. Adam optimizer with learning rate of 0.001 is used which has the fastest convergence and lower loss during the training. The images were processed in the batch of 8 and the loss function being categorical cross-entropy. All classifiers were trained for 10 epochs. The Inception V3 achieved an accuracy of 85%, ResNet-50 model trained with an accuracy of 86.19% and VGG-19 with training accuracy of 90.06% which is the best among the three. The ensemble model which is the average prediction of the three models achieved an accuracy of 91%.

6. Results

The model was evaluated using accuracy, sensitivity/recall, precision and F1 score. These evaluation metrics helps to determine the reliability of the model. The ideal model classifies true positive and true negative efficiently.

i. Accuracy

Model	ResNet-50	VGG-19	Inception V3	Ensemble
Accuracy in %	**86.19**	**90.06**	**85.08**	**91**

Although all the four models performed well, from the above chart it is obvious that the VGG-19 classifier performed well in terms of accuracy and ensemble method performed slightly better.

ii. Sensitivity or Recall or True positive rate

Model	ResNet-50	VGG-19	Inception V3	Ensemble
Sensitivity	**0.88**	**0.96**	**0.87**	**1**

Sensitivity in this research work denotes what proportion of images that had blood cancer being classified as blood cancer. Sensitivity is given by the equation: Sensitivity = TP / (TP + FN)

iii. Precision

Model	ResNet-50	VGG-19	Inception V3	Ensemble
Precision	**0.91**	**0.92**	**0.90**	**0.95**

Precision = TP/ (TP+FP)

Precision is the measure that tells us what proportion of images that we classified as having blood cancer, actually had blood cancer. Ensemble model has precision value of 0.95 followed by VGG-19 with precision value of 0.92.

iv. F1 score

Model	ResNet-50	VGG-19	Inception V3	Ensemble

F1 score	**0.90**	**0.94**	**0.88**	**1**

F1 score is a measure of test's accuracy that considers both the precision and recall of the test to compute the score. It is given by the formula: F-measure = 2*Recall*Precision / Recall + Precision

v. Confusion Matrix

Figure 2 below shows the confusion matrix of the proposed model. The deep colored diagonal boxes in our confusion matrix show the count of True Positive (TP) items in each class.

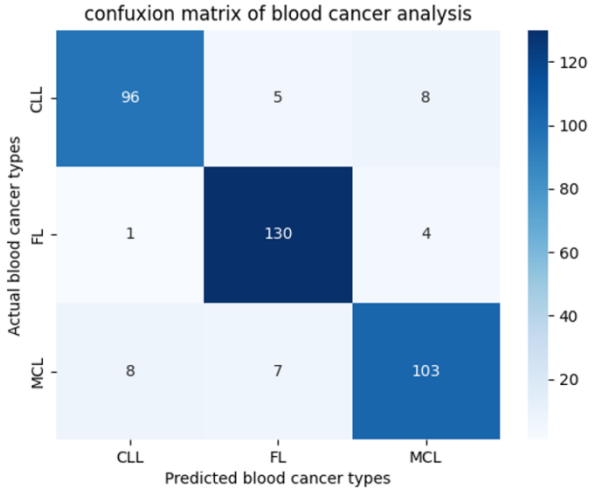

Figure 2: Confusion matrix obtained for the new model

From the above metrics, it is evident that the ensemble architecture outperforms the other deep learning models in image classification. It achieved an accuracy of 91%, sensitivity 100%, f1 score 100% and precision 95%.

7. Conclusion and Recommendations

The system equipped us with the idea of replacing the manual method of diagnosing blood cancer into a computerized approach. Medical personnel are advised to exploit the opportunity which this software offers. There is no gainsaying the fact that reliable software of this sort which can perform the following tasks is needed.

More deliberation can be made which can even lead to the development of better systems like the one which can diagnose any type of disease in the world. In future, it is also recommended to explore the application of our approach to other use cases for histological image analysis and also with a different training strategy that is not end-to-end. For future research directions, it is recommended to employ the ensemble of other CNN architectures to further increase the accuracy and sensitivity values.

References

Acevedo, A., Alferez, S., Merino, A., Puigvi, L. and Rodellar, J. (2019). Recognition of Peripheral Blood Cell Images Using Concolutional Neural Networks. Computer Methods and Programs in Biomedicine, 180(2), 627-638.

Anilkumar, K. K., Manoj, V. J. and Sagi, T. M. (2021). Automated detection of b cell and t cell acute lymphoblastic leukaemia using deep learning. IRBM.

Anwar, S. and Alam, A. (2020). A convolutional neural network–based learning approach to acute lymphoblastic leukaemia detection with automated feature extraction. Medical & biological engineering & computing, 58(12), 3113-3121.

Bain, B. J. (2015). Diagnosis From Blood Smear. New England Journal of Medicine, 5 (498-507), 353.

Banik, P. P., Saha, R. and Kim, K. D. (2020). An automatic nucleus segmentation and CNN model based classification method of white blood cell. Expert Systems with Applications, 149, 113211.

Bennur, A. (2019). Ensemble model for X-Ray Image Classification. Dublin: National College of Ireland. Retrieved March 1, 2022, from http://norma.ncirl.ie/id/eprint/3844

Boesch, G. (2022). Deep Residual Networks - Guide in 2022. Retrieved February 22, 2022, from Viso.ai: https://viso.ai/deep-learning/resnet-residual-neural-network.

Boldú, L., Merino, A., Acevedo, A., Molina, A. and Rodellar, J. (2021). A deep learning model (ALNet) for the diagnosis of acute leukaemia lineage using peripheral blood cell images. Computer Methods and Programs in Biomedicine, 202, 105999.

Brancati, N., De Pietro, G., Frucci, M. and Riccio, D. (2019). A deep learning approach for breast invasive ductal carcinoma detection and lymphoma multi-classification in histological images. IEEE Access, 7, 44709-44720.

Cai, S. and Verbeek, C. (2019). Automatic Thyroid Nodule Recognition and Diagnosis in Ultrasound Image with the YOLOv2 Neural Network. World Journal of Surgical Oncology, 17(1), 1-9.

Casey, K. (2019). How to explain deep learning in plain English. The Interprisers Project , (pp. 1-2).

Ceelie, H., Dinkelaar, R. B. and Van-Gelder, W. (2017). Examination of peripheral blood films using automated microscopy; evaluation of Diffmaster Octavia and Cellavision DM96. Journal of Clinical Pathology, 60 (1), 72-79.

Das, P. K., & Meher, S. (2021). An efficient deep convolutional neural network based detection and classification of acute lymphoblastic leukemia. Expert Systems with Applications, 183, 115311.

Durant, T. J. S., Olsen, E. M., Schulz, W. L. and Torres, R. (2017). Very Deep Convolutional Neural Networks for Morphologic Classification of Erythrocytes. Clinical Chemistry, 63 (12), 1-9.

Genovese, A., Hosseini, M. S., Piuri, V., Plataniotis, K. N. and Scotti, F. (2021). Acute Lymphoblastic Leukemia detection based on adaptive unsharpening and Deep Learning. In ICASSP 2021-2021 IEEE

International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1205-1209). IEEE.

Hasan, M. M., Srizon, A. Y., Sayeed, A. and Hasan, M. A. M. (2020). Accurate Recognition of Leukemia Sub-types by Utilizing a Transfer Learned Deep Convolutional Neural Network. In 2020 11th International Conference on Electrical and Computer Engineering (ICECE) (pp. 427-430). IEEE.

He, K., Zhang, X., Ren, S. and Jian, S. (2016). Deep Residual Learning for Image Recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778). Chengdu, China: IEEE.

Kasani, P. H., Park, S. W. and Jang, J. W. (2020). An aggregated-based deep learning method for leukemic B-lymphoblast classification. Diagnostics, 10(12), 1064.

Kassani, S. H., Kassani, P. H., Wesolowski, M. J., Schneider, K. A. and Deters, R. (2019). A hybrid deep learning architecture for leukemic B-lymphoblast classification. In 2019 International Conference on Information and Communication Technology Convergence (ICTC) (pp. 271-276). IEEE.

Khandekar, R., Shastry, P., Jaishankar, S., Faust, O. and Sampathila, N. (2021). Automated blast cell detection for Acute Lymphoblastic Leukemia diagnosis. Biomedical Signal Processing and Control, 68, 102690.

Lee, J. G., Jun, S., Cho, Y. W., Lee, H., Kim, G. B. and Kim, N. (2017). Deep Learning in Medical Imaging: General Overview. Korean Journal of Radiology, 18(4), 570-584.

Livieris, I. E., Kanavos, A., Tampakas, V. and Pintelas, P. (2018). An Ensemble SSL Algorithm for Efficient Chest X-Ray Image Classification. Journal of Imaging, 4(7), 95.

Lu, S., Lu, Z. and Zhang, Y. D. (2019). Pathological Brain Detection Based on AlexNet and Transfer Learning. Journal of Computational Science, 30, 41-47.

Mao, Y., Yin, Z. and Schober, J. (2016). A deep convolutional neural network trained on representative samples for circulating tumor cell detection. In 2016 IEEE Winter Conference on Applications of Computer Vision (WACV) (pp. 1-6). IEEE.

Murtaza, P., Sajjad, M., Khan, S. and Wu, W. (2019). Multi-grade Brain Tumor Classification Using Deep CNN with Extensive Data Augmentation. Journal of Computer Science, 30(2), 174-182.

O'Leary, D. E. (2013). Expert Systems. Retrieved September 29, 2016, from Wiley Encyclopedia of Computer Science and Engineering Web site: http://www.ssrn.com

Pang, S., Yu, Z. and Orgun, M. A. (2017). A novel end-to-end classifier using domain transferred deep convolutional neural networks for biomedical images. Computer methods and programs in biomedicine, 140, 283-293.

Phung, V. H. and Rhee, E. J. (2019). A High-Accuracy Model Average Ensemble of Convolutional Neural Networks for Classification of Cloud Image Patches on Small Datasets. Journal of Information and Communication Convergence Engineering., 16(3), 173-178.

Praveena, P. and Singh, J. (2020). Hello World Deep Learning in Medical Imaging. Journal of Digital Imaging, 31(3), 283-289.

Qiao, Y., Zhang, Y., Liu, N., Chen, P. and Liu, Y. (2021). An End-to-End Pipeline for Early Diagnosis of Acute Promyelocytic Leukemia Based on a Compact CNN Model. Diagnostics, 11(7), 1237.

Ridoy, M. A. R. and Islam, M. R. (2020). An automated approach to white blood cell classification using a lightweight convolutional neural network. In 2020 2nd International Conference on Advanced Information and Communication Technology (ICAICT) (pp. 480-483). IEEE.

Saleem, S., Amin, J., Sharif, M., Anjum, M. A., Iqbal, M. and Wang, S. H. (2021). A deep network designed for segmentation and classification of leukemia using fusion of the transfer learning models. Complex & Intelligent Systems, 1-16.

Shafique, S. and Tehsin, S. (2018). Acute lymphoblastic leukemia detection and classification of its subtypes using pretrained deep convolutional neural networks. Technology in cancer research & treatment, 17, 1533033818802789.

Simonyan, K. and Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition . The 3rd International Conference on Learning Representations. San Diego, USA. Retrieved from https://arxiv.org/abs/1409.1556

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z. (2016). Rethinking the Inception Architecture for Computer Vision. Computer Science Conference on Computer Vision and Pattern Recognition (pp. 27-30). Las Vegas, USA: IEEE.

Tessema, H., Prasad, K. and Hebbar, H. (2021). Comparison of Traditional Image Processing and Deep Learning Approaches for Classification of White Blood Cells in Peripheral Blood Smear Images. Biocybersecurity and Biomedical Engineering, 39(2), 382-392.

Tsykunov, D. V., Kolokolnikov, G. A. and Samorodov, A. V. (2020). Application of Deep Learning Methods for Automation of Leukocyte Microscopy. In 2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus) (pp. 2564-2567). IEEE.

Watine, J. (2021). Evaluation of Diagnostic Procedures. Retrieved October 5, 2022, from bmj: http://www.bmj.com.

Yu, H., Li, L., Liu, C., Huang, W., Zhou, J., Fu, W. and Wu, Q. (2017). Factors Associated with Quality of Life of Family Caregivers for Leukemia Patients in China. Health and Quality of Life Outcomes, 15(1), 55.

Zhao, C., Feng, Y., Liu, R. and Zheng, W. (2020). Application of Lightweight Convolution Neural Network in Cancer Diagnosis. In Proceedings of the 2020 Conference on Artificial Intelligence and Healthcare (pp. 249-253). ACM Digital Library.