

Feature Fusion ResNet18 and VGG16 for ALL classification

Kritsadakorn Khongsep¹, Buntueng Yana^{1*}, Sitthichat Meekhwan¹, and Tanawat Palamee¹

¹Department of Electrical Engineering, School of Engineering, University of Phayao, Thailand
kritsadakorn.mark@gmail.com, mr.buntueng@gmail.com*, mr.sitthichat.n@gmail.com, beetanawat7@gmail.com

Abstract

Acute Lymphoblastic Leukemia is a common blood cancer in children. Early detection is important, as it can reduce the risk of death rate and improve survivor rate. In this study, we propose a classification model for ALL by using fusion of two pre-trained convolution neural network models like ResNet18 and VGG16. Both models are used to extract feature from images. The extracted features then combine them with the attention mechanism and pass through to linear layers to predict the result. The CNMC-2019 public dataset is used in this study. We do data augmentation to make the balanced dataset with 14,544 blood smear images. The model is evaluated using five-fold cross-validation. The proposed fusion model achieved a strong performance, with an accuracy of 92.16%, precision of 94.59%, recall of 89.44%, F1-score of 91.94%, and an AUC of 0.9697. These results show that our proposed fusion model can effectively classify ALL images without a complex structure model. This approach offers a practical solution for developing medical image classification systems using pre-trained models.

Keywords: acute lymphoblastic leukemia, feature fusion, deep learning, convolutional neural network

1. Introduction

Leukemia type of cancer that affects the white blood cells, where abnormal cells increase uncontrollably. It disrupts the balance of blood cells in the body, weakens the immune system and increasing the risk of infections because of the lack of healthy blood cells. Acute lymphoblastic leukemia (ALL) is the most common type of childhood cancer globally, accounting for about 25% of cancer diagnoses in children under 15 years old [1, 2]. Early detection is important because it can improve the chances of successful treatment and recovery while reducing the risk of severe complications or mortality.

Various methods, including blood tests, biopsies, genetic tests, CT scans, MRI scans, and others [3], can detect leukemia. However, the process for identifying abnormal cell still needs a manual process by medical experts. This process can be error, as it depends on the skills and experience of the person. Some mistakes may happen because certain details are too small to be noticed by the human eye. For the detection process, machine learning can help with this task. It can see details that humans can't see through the analysis of blood smear images. This helps reduce errors in the detection process.

Image detection tasks utilize machine learning techniques such as You Only Look Once (YOLO), Convolution Neural Network (CNN), Visual Geometry

Group (VGG), Region-based Convolutional Neural Network (R-CNN), Long-Short Memory (LSTM), and Residual Net (ResNet). Among these, CNNs are among the most widely used and effective techniques for image recognition and detection. The advantage of CNN is that it can understand the structure and position of objects in an image while preserving the spatial relationships between pixels, which makes CNN very good at working with images [4]. However, CNN also has some limitations. It usually needs very large, balance datasets to work well, to avoid overfitting or bias. Model performance can drop when the dataset is noisy [5].

Pre-trained base-CNN models are deep learning models that someone previously trained. These models have already learned to detect a wide range of features from images, such as edges, textures, and shapes. Using a pre-trained model can help to improve performance even with a small dataset [6]. Imaging classification tasks widely adopt models like VGG, ResNet, and Inception. However, sometimes using a single pre-trained model will get a low performance model [7].

In this study, we propose a fusion ResNet18 with VGG16 to boost model performance and make it to suitable. By fusion both models, we aim to enhance the accuracy of ALL classification model.

2. Literature Review

This literature review explores various applications of feature fusion techniques in deep learning models. And find new solution for building ALL classification model. In the context of blood-related disease detection, multiple studies have utilized deep learning and feature fusion to improve diagnostic accuracy.

Hasanaath et al. [8] used CNMC-2019 dataset to present ALL detection model. Features are extracted using InceptionResNetV2, DenseNet121, and VGG16 models. A Global Average Pooling layer concatenates the collected features into a flattened tensor, creating a 1D feature vector with 3072 values per image. Researchers trained and evaluated five classification algorithms using these features. In conclusion, Support Vector Machine performed best with 91.63% testing accuracy, 90.83% precision, 89.48% recall, and 90.14% F1-score.

Similarly, J. Amin et al. [9] developed white blood cell detection model. Images were transformed from RGB to HSV. AlexNet, MobileNetV2, ShuffleNet, and ResNet-18 were pre-trained transfer learning models for feature extraction. Four feature vectors are serially concatenated. On this merged vector, a non-dominated sorting genetic algorithm (NSGA) reduces redundant features and selects the best ones based on fitness functions to select and optimize the final feature set. Then classify with SVM.

*Corresponding Author

The model tested on LISC, ALL_IDB1, and ALL_IDB2 public datasets. The approach had 1.00 accuracy on blast and non-blast cells, 0.9992 accuracy on basophil cells, and 1.00 accuracy on lymphocyte, neutrophil, monocyte, eosinophil, and mixture of these cells.

Anilkumar et al. [10] proposed pre-trained CNNs for leukemia detection such as AlexNet, VGG16, VGG19, ResNet. The study tested each model on ALL_IDB1 and ALL_IDB2. On ALL_IDB1, most models were 100% accurate. However, on ALL_IDB2, they were less accurate. In this more complicated testing scenario, VGG16 and AlexNet performed somewhat worse than deeper residual networks. Despite this, VGG16 showed strong and stable training behaviour across optimization strategies. Beyond hematologic problems, fusion models have shown promise in other medical applications.

Rehman et al. [11] combined ViT and VGG16 to detect architectural distortion in digital mammograms. Gaussian and LoG filters combine smooth and edge details with original images, and histogram equalization boosts contrast and data. VGG16 extracts important visual patterns, while ViT captures complex features to boost performance. Model validation used private PINUM and public DDSM datasets. The model had 0.97 sensitivity, 0.92 F1-score, 0.93 precision, 0.94 specificity, and 0.96 accuracy on PINUM and 0.93 sensitivity, 0.91 F1, 0.94 precision, 0.92 specificity, and 0.95 accuracy on DDSM.

Kaya et al. [12] suggested Fusion-Brain-Net, a hybrid deep learning model for brain tumor classification that combines VGG16, ResNet50, and MobileNetV2. By using complementary network strengths, the model reduces overfitting and improves classification accuracy. The framework incorporates preprocessing, data augmentation, feature fusion, fine-tuning, and classification components. The Br35H dataset includes 3,000 MR images from Figshare (3064), Nickparvar (7,023), and Sartaj (3,264). Across datasets, the refined Fusion-Brain-Net model achieved 97.56% accuracy.

Q. Wang et al. [13] created a CADx for OCT image classification. The model uses a self-supervised vision transformer. Three ViTs extract multi-scale features from images resolutions. Multiple ViT enchanted model performance. Cross-attention and a concatenation layer combine extracted features. A weighted average produces the final prediction in the classification layer. To evaluate the model, they use ten-folds cross-validation. The model achieved an AUC value of 0.9963, 95.89% sensitivity, and 98.23% specificity for the internal dataset. They also use external to validate the model. The model achieved 92.06% sensitivity, and 95.56% specificity. Some studies applied fusion to innovative data types or advanced generative models.

Shi and Yan [14] introduced an algorithm address the issues of imbalance dataset and low performance in single-feature extraction. They use generative adversarial networks and denoising diffusion probabilistic models (DDPMs) to increase the dataset, then they converted one dimension time-series into color image using Markov

transition field (MTF). Next, they use VGG16, ResNet18, and Vision Transformer (ViT) for exact features. They use Moore dataset with contains 12 classes for network traffic in csv or arff file format. In the result their algorithm success to achieve higher accuracy of 93.00% followed by VGG16 and ResNet18 with 92.78%.

Based on previous research, we propose a fusion model combining ResNet18 and VGG16 for ALL classification using the CNMC-2019 dataset. VGG16 contributed effectively to feature extraction in a high-performing [8] and have stability and strong performance across various situations [10]. ResNet18 showing high performance for classification white blood cell [9]. We chose ResNet18 and VGG16 because both are convolutional neural network (CNN) models that have strong performance in image classification tasks. This combination was selected to provide a good balance between deep feature extraction and model simplicity. By using these two models together, we aim to improve the performance of ALL classification model.

3. Methodology

This section explain our propose model for enchanted ALL classification model with blood smear images. We use ResNet18 and VGG16 to extract important information from images. The output from both method are fuse to linear layer by using cross-attention. All experiment were doing in local machine dedicated Nvidia GeForce RTX 3080 GPU. Software that use in this study is Nvidia driver 575.57.08, cuda version 12.9, python version 3.10.12, and pytorch version 2.7.1.

3.1 Propose Model

In this work, we investigate three models for classifying blood smear images of ALL. The first model use ResNet18 for feature extraction followed by three linear layers to get the prediction. Second, VGG16 also have the same setting like ResNet18 model. Last, our proposed model that fusion ResNet18 and VGG16. This model use attention mechanisms to fuse the features extracted from both model then passing through linear layers to get final result.

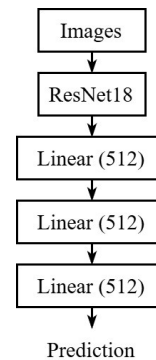


Fig. 1 ResNet18 model structure

We extract features from blood smear images using ResNet18. Three linear layers, each with 512 units, then

process these features to learn complex relationships and make the final prediction, as illustrated in Fig. 1.

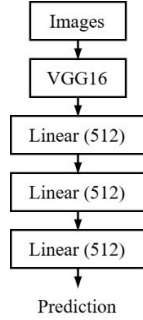


Fig. 2 VGG16 model structure

The second model uses the same structure as the first model, which is shown in Fig. 2. However, feature extraction should use VGG16, not ResNet18.

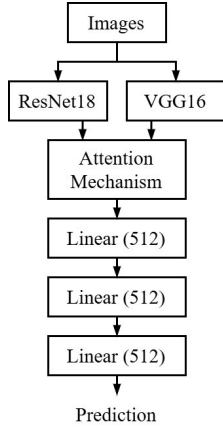


Fig. 3 Fusion model structure

The proposed model for feature extraction uses ResNet18 and VGG16, depicted in Fig. 3. The attention mechanism layer combines the extracted features before passing them to three linear layers to generate the final predictions.

3.2 Dataset CNMC 2019

The CNMC 2019 dataset was collected at the Laboratory Oncology Unit, ALL India Institute of Medical Sciences, New Delhi, India [15]. Researchers prepared slides from patient bone marrow aspirates. Researchers made normal cell slides from control subjects and prepared cancer cell slides from new participants. The collection includes 49 healthy and 69 malignant cell pictures. Author resized photos to 350×350 pixels by padding columns and rows with zero intensity and centering them. The collection contains 7,272 photos for aberrant cell call "ALL" and 3,389 for healthy cell call "HEM" classes. Our dataset contains 14,544 pictures.

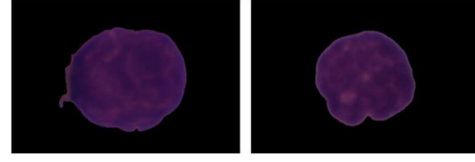


Fig. 4 Blood smear image

Fig. 4 Blood smear image shows example image. Left image represents "ALL" class, right shows "HEM" class. We normalized each image to 0–1 before using it. We also employ data augmentation to balance the dataset. We balanced the dataset with 7,272 photos per class by rotating and mirroring "HEM" images. Our dataset has 14,544 pictures. We employed five-fold cross-validation on our dataset. Splitting the data set with a stratified k-fold.

3.3 ResNet18

ResNet is a deep-learning architecture that has residual connections, which allows the model to learn the difference (or residual) between the input and the output. These connections let the input skip certain layers and go directly to the output. ResNet18 is a model that contains 18 layers, include convolutional layers for feature extraction, batch normalization layers for training stability, and fully connected layers for making the final prediction.

3.4 VGG16

VGG16 is a pre-trained model that designed to reduce complex structures by use the set of convolution kernels with 3×3 layers. These filters help extract important features from the input image. Each group of convolution layers followed by max pooling layers to down sampling. VGG16 contains 16 layers, 13 layers for Convolution layers and 3 layers for fully connected layers at the end for making prediction.

4. Evaluation Metrics

We evaluate our model by using accuracy, precision, recall, F1-score and training time. To calculate value of these metrics, We use True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). Equations can be written as follows.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{Tp}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 \text{ score} = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

5. Result

We evaluate the performance of all models using a test set and five-fold cross-validation. We show the average results for each model in Table 1.

Table 1 Result

Model	ResNet18	VGG16	Resnet+VGG16
Accuracy	0.8902	0.9053	0.9216
Precision	0.8820	0.9215	0.9459
Recall	0.9018	0.8863	0.8944
F1-score	0.8915	0.9034	0.9194
AUC	0.9548	0.9582	0.9697

The fusion model achieves accuracy 0.9216, follow by VGG16 with accuracy 0.9053, and ResNet18 with accuracy 0.8902. Fusion model achieved higher precision with 0.9459, F1-score with 0.9194, and AUC with 0.9697. In recall, ResNet18 has a higher score with 0.9018, followed by a fusion model with 0.8944.

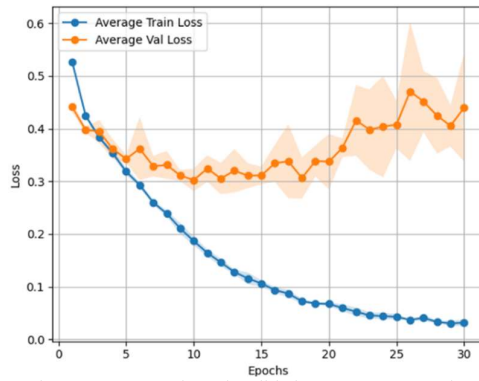


Fig. 5 Average Train and Validation Loss over Epochs

Fig. 5 Average Train and Validation Loss over Epochs displays the training and validation losses of the fusion model. Training loss steadily decreasing, while validation loss slightly decreases in early epochs and hit the lower around epochs 10–15. After epochs 15 the curve is start increased, showing the overfitting in the later epochs.

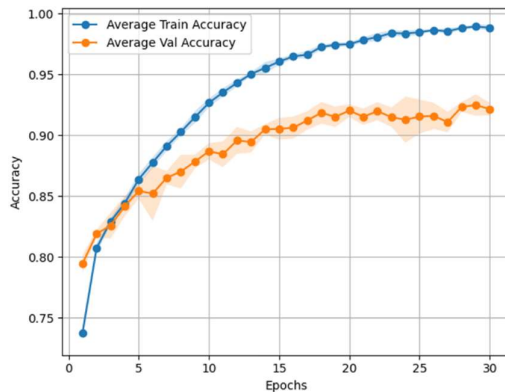


Fig. 6 Average Train and Validation Accuracy over epochs

From Fig. 6 Average Train and Validation Accuracy over epochs, training accuracy is significantly increased

over epochs. However, validation accuracy is gradually increases before being stable in epochs 20 to later.

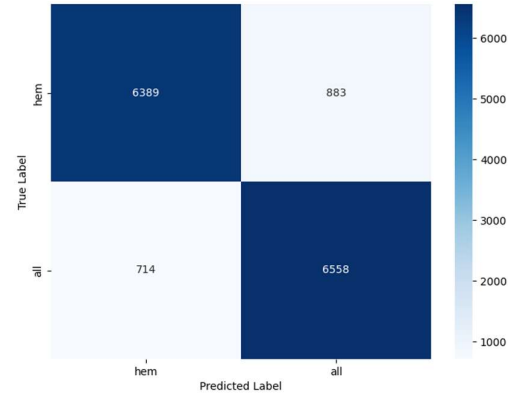


Fig. 7 Confusion matrix of Fusion model

Fig. 7 Confusion matrix of Fusion model is a confusion matrix of fusion model for binary classification. It refers to performing the model to classify ALL.

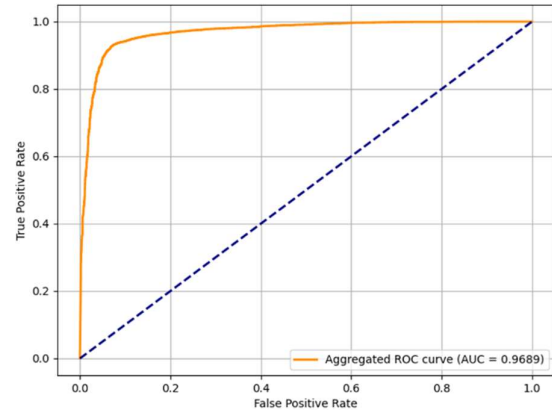


Fig. 8 Fusion model aggregated ROC curve

Fig. 8 shows the classification ability. It showing the model can discriminate between positive and negative class.

6. Conclusions

In this study, we develop a classification model for ALL by combining ResNet18 and VGG16. To enhance model performance, we used data preprocessing techniques to balance the dataset. We also used five-fold cross-validation to evaluate the model performance. Our proposed fusion model can achieve an accuracy of 92.16%, precision of 94.59%, recall of 89.44%, F1-score of 91.94%, and AUC of 0.9697. The result showed the solution to build a simple model with pre-trained. A simple design can build an effective model. By using simple fusion models, we expect our experiment will indicate to build easy solutions to create ALL classification models with simple structures.

References

- [1] Tran, T.H. and S.P. Hunger, *The genomic landscape of pediatric acute lymphoblastic leukemia and precision medicine opportunities*. Seminars in Cancer Biology, 2022. 84: p. 144-152.
- [2] Jiwani, N., et al., *Pattern Recognition of Acute Lymphoblastic Leukemia (ALL) Using Computational Deep Learning*. IEEE Access, 2023. 11: p. 29541-29553.
- [3] Sanjaya, A.D., R. Setiawan, and N.F. Hikmah, *RNA-BioLens: A Novel Raspberry Pi-Based Digital Microscope With Image Processing for Acute Lymphoblastic Leukemia Detection*. IEEE Access, 2025. 13: p. 23618-23628.
- [4] Indolia, S., et al., *Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach*. Procedia Computer Science, 2018. 132: p. 679-688.
- [5] Sriwong, K., K. Kerdprasop, and N. Kerdprasop, *The Study of Noise Effect on CNN-Based Deep Learning from Medical Images*. International Journal of Machine Learning and Computing, 2021. 11: p. 202-207.
- [6] Kartik, K., et al., *Transfer Learning with CNNs in Small ML Datasets: Applying Pre-Trained CNN Models and Fine-Tuning them for Limited Data Scenarios of the Creative Commons Attribution License (CC BY 4.0)*. 2023: p. 1087-1099.
- [7] Stančić, A., V. Vyroubal, and V. Slijepčević *Classification Efficiency of Pre-Trained Deep CNN Models on Camera Trap Images*. Journal of Imaging, 2022. 8, DOI: 10.3390/jimaging8020020.
- [8] Hasanaath, A., et al., *Acute lymphoblastic leukemia detection using ensemble features from multiple deep CNN models*. Electronic Research Archive, 2024. 32: p. 2407-2423.
- [9] Amin, J., et al., *An Integrated Design Based on Dual Thresholding and Features Optimization for White Blood Cells Detection*. IEEE Access, 2021. 9: p. 151421-151433.
- [10] Anilkumar, K.K., V.J. Manoj, and T.M. Sagi, *Automated detection of leukemia by pretrained deep neural networks and transfer learning: A comparison*. Medical Engineering & Physics, 2021. 98: p. 8-19.
- [11] Rehman, K.U., et al., *A Feature Fusion Attention-based Deep Learning Algorithm for Mammographic Architectural Distortion Classification*. IEEE J Biomed Health Inform, 2025. Pp.
- [12] Kaya, Y., E. Akat, and S. Yıldırım, *Fusion-Brain-Net: A Novel Deep Fusion Model for Brain Tumor Classification*. Brain Behav, 2025. 15(5): p. e70520.
- [13] Wang, Q., et al., *Cross-Attention Based Multi-Resolution Feature Fusion Model for Self-Supervised Cervical OCT Image Classification*. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2023. 20(4): p. 2541-2554.
- [14] Shi, D. and J. Yan, *A Deep learning based network traffic classification algorithm using data augmentation and multi-feature fusion*. 2025.
- [15] Gupta, R., S. Gehlot, and A. Gupta, *C-NMC: B-lineage acute lymphoblastic leukaemia: A blood cancer dataset*. Medical Engineering & Physics, 2022. 103: p. 103793.



Krisadakorn Khongsep He is currently studying for a bachelor's degree in the Faculty of Engineering, Department of Electrical Engineering, University of Phayao. Current research interests: machine learning.



Buntueng Yana (Member, IEEE) received B.Eng. and M.Eng. degrees in Electrical Engineering from Chiang Mai University, Thailand, in 2006 and 2009, respectively, and a Ph.D. in Information Science and Technology from Osaka University, Japan, in 2019. He is currently a lecturer with the Department of Electrical Engineering at the University of Phayao, Thailand, where he began his academic career. His research interests include robotics, automatic control, power generation, and the application of machine learning to medical applications.



Tanawat Palamee is currently pursuing the B.S. degree in Electrical Engineering at the University of Phayao, Thailand. His research interests machine learning, and robotic agriculture.



Sitthichat meekhwan received the bachelor's degree in electrical engineering from the University of Phayao (UP), Phayao, Thailand, in 2019. He is currently studying for a master's degree at the Faculty of Engineering. His current research interests in Machine learning, deep learning, and robotic