



# Hybrid Deep Learning Approach for Non-Hodgkin's Lymphoma using ViT and ResNet

Yohesha SK<sup>1</sup> and Ms. R. Dheepthi<sup>2</sup>

<sup>1,2</sup>Hindustan Institute of Technology and Science  
skyohesha@gmail.com

dheepsraj@gmail.com

**Abstract.** The subtypes of Non-Hodgkin s lymphoma are very important in analyzing the right treatment plans that would be selected and to know the improvement in the outcomes of patients. Nevertheless, addressing the standard histopathological diagnosis is time consuming and subjective. The paper solutions to the above challenges were proposed by providing a new hybrid deep learning process that entailed the combination of the ViT and ResNet-50 models, deep feature extraction, and deep residual learning. The model was being trained and validated on a series of the histopathological photographs, which provided a test accuracy of 96.70 percent and even outperformed the standalone ViT and ResNet structures as well as the other existing systems.

**Keywords:** Non-Hodgkin's Lymphoma, Vision Transformer (ViT), Residual Network(ResNet), Chronic Lymphocytic Leukemia, Follicular Lymphoma, Mantle Cell Lymphoma.

## 1 Introduction

### 1.1 Introduction on Non-Hodgkin's Lymphoma

Please Non-Hodgkin's Lymphoma refers to the heterogeneous collection of cancer types that originates from lymphatic tissues. Among the various types of NHL the FL, MCL, CLL are rare in clinical presentation and requires to be classified accurately. The incorrect classification of these can mislead to incorrect treatment plan, and makes the patient survival risk higher.

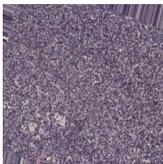


Fig 1. CLL

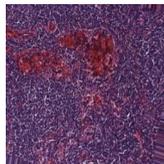


Fig 2. MCL



Fig 3. FL

The conventional diagnosis of the NHL subtypes is carried out based on the histopathological images performed by skilled pathologists. The classical and traditional manner of categorizing its types is a time consuming process that also poses a greater risk of being misclassified especially where there is a case of uncertainty. The deep learning models, in particular, CNNs showed an outstanding achievement at executing the process of extracting the spatial patterns in the histopathological images. Although they do good in drawing the spatial patterns, they have problems in arresting the long-range dependencies and the global contextual relationships. Such a type of drawback is solved by the emerging model of deep learning methods, the Vision Transformers (ViT), which uses the self-attention mechanism, and the ViT attempts to capture the local and global information with a certain possibility of improvement in the field of medical images as a type of classification. Besides ViT, ResNet too fared and demonstrated deep features extraction through application of residual learning to reinforce the gradient flow and the convergence which occur during training.

Although the ViT and the ResNet were more superior in classifying the medical images, the hybrid model that involves both ViT and ResNet to find out the threshold value of classifying the medical images is not discussed. The proposed study deals with the research gap by suggesting a hybrid deep learning method in which the global feature extractor of ViT and the hierarchical learning of ResNet-50 will be combined to conduct the classification of CLL, FL and MCL.

## 1.2 Key Objectives and Contributions of the Study

The following main questions guide the study is to come up with a Camel-based ViT-ResNet architecture to determine the type of NHL subtypes; compare the performance of the proposed Hybrid model with old-fashioned CNN-based models; and improve the diagnostic accuracy and reliability in the use of DL techniques.

The contributions of the research are a hybrid model of ViT-ResNet that can be used to classify NHL subtypes, an evaluation of the performance that proves the results through the use of the existing models, a comparative analysis, which indicates the clinical significance of the model.

## 2 Literature Review

Non Hodgkin Lymphoma (NHL) Diagnosis based on Deep Learning Deep learning techniques have been applied in the classification of Non Hodgkin Lymphoma (NHL) and have gained considerable interest since they have a potential of improving the diagnoses. Medical image classification has been studied using models such as convolutional neural networks (CNNs), Vision Transformers (ViTs) and hybrid systems. In this section, we have a comparative literature review of relevant key literature factors in which we have identified the advantages and limitations of previous methods in regards to our Sydney ViT-ResNet-based approach.

## 2.1 Lymphoma Subtype using DenseNet121 [1]

**Summary.** In this work, the DenseNet121 architecture is utilised to categorise subtypes of lymphoma based on histopathological imagery. It is based on the dense connection model, which enhances the flow of gradient and reuse of features and hence optimizes the classification performance with tuning of hyperparameters.

*Pros.* i. DenseNet121 builds on the concept of feature reuse - and is more learning-efficient.

ii. Performs well in the classification of medical images.

*Cons (In comparison to our Model):* i. Does not have any attention mechanisms of Vision Transformers, thus has somewhat limited coverage over long range dependencies.

ii. It does not use the residual learning of ResNet that can increase the deep feature learning and gradient transfer.

iii. Strategies employed to tune the hyperparameter might not be applicable across a wide range of data.

## 2.2 Classification of Leukemia and Lymphoma using Faster R-CNN [2]

**Summary.** The [2] addresses the problem of presenting Lymphomas and Leukemia using hybrid technique of VGG-16 and Regional Proposal Network. The model makes use of Faster R-CNN along with the VGG-16, which adheres to the methodology of Object detection that includes Region of Interest (ROI) and CNN. The model achieved a level of accuracy of 87.575 percent.

Ensemble methodology of ViT and ResNet that implements classification methodology of Image Classification using CNN and Transformer in an ensemble technique is proposed to be adopted. Since the proposed system has integrated the usage of ViT, the model will be computationally complex.

## 2.3 Medical Imaging with Vision Transformers

**Summary.** This paper addresses the application of the Vision Transformers (ViTs) with medical image classification, focusing on their self-attention mechanism, giving the possibility to extract higher aspects of spatial, contextual features. The findings point towards the fact that ViTs outperform traditional CNNs in various medical images tasks.

*Pros:* i. ViTs have been able to extract global dependencies in images so as to enhance the accuracy of classification.

ii. Addresses the disadvantage of the traditional CNNs with regard to local feature extraction.

*Cons (In comparison with our Model):*i. ViTs are quite demanding in terms of dataset size in order to generalize properly, and this might not be at all times possible in medical imaging.

ii. Does not have the abilities of ResNet, to combine feature extraction hierarchically to improve deep learning performance.ii. It does not use the residual learning of ResNet that can increase the deep feature learning and gradient transfer.

iii. Strategies employed to tune the hyperparameter might not be applicable across a wide range of data.

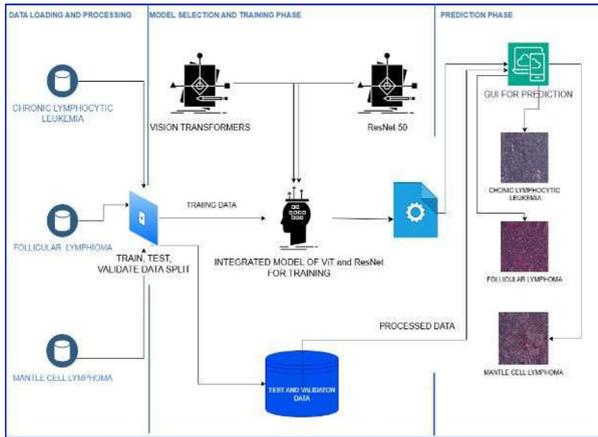
#### 2.4 Medical Diagnosis with Hybrid CNN-ViT Models

The general study adds to these already existing methods and uses both the Vision Transformers and ResNet to classify the Non-Hodgkin Lymphoma. Our design in comparison to pure ViT-based approaches also uses the hierarchical feature learning of ResNet, which leads to increased robustness of feature extraction. More so, our approach, unlike hybrid CNN-ViT models, is augmented with residual learning, which enhances absolute gradient stability and presentation of deep features. Our model would integrate all of these advanced architectures in order to increase the predictions accuracy and generalization in NHL classification.

### 3 Proposed System

The research problem is narrowed down to come up with a powerful deep learning model to classify Non- Hodgkin Lymphoma (NHL) subtypes. The model encompasses hybrid architecture which involves incorporating both Convolutional Neural Network and Transformer to improve feature extraction as well as accuracy of classification. The three subtypes of the NHL that are hardly encountered in the medical field that have been researched include Follicular Lymphoma (FL), Chronic Lymphocytic Leukemia (CLL) and Mantle Cell Lymphoma (MCL). The model has a 3000 image set of histopathological images. The dataset is divided into 3 sets on Training, Validation and Testing in the percentage ratio of 7:2:1 respectively.

The dataset of each image has been transformed using a set of transformations such as image resizing, normalization and conversion to tensors on trying to make it compatible with the deep learning models.



**Fig 4.** ViT and ResNet Architecture.

The implemented hybrid system combines the two state of the art architecture: ResNet-50: It is a pretrained convolutional neural network (CNN) that is used as a feature extractor. I removed the fully connected layer of ResNet-50 and this enabled it to perform the task of feature extracting on the input images.

ViT: The ViT is also a pretrained model that the extracted feature is processed in. The last classification layer of ViT is substituted with a personal layer developed to perform three-class-classification.

Training process The output of the model is trained with Cross-entropy loss, which is an error measurement of classification, the process of training model with AdamW optimizer is adopted to enhance convergence, model is trained on 32 batch size which balances the acceleration of model training and the computation requirement.

## 4 Module Description

The system designed has several modules based on which different functions of classification of subtypes of Non-Hodgkin Lymphoma are performed. The modules that the proposed system has are:

The task of the dataset processing module is to render the histopathological images ready to perform the training operation of the model as the module loads the number of 3000 images with their classification as FL, CLL and MCL are divided with the proportion of 7:2:1 in the particular subset. The preprocessing processes which included resizing, normalization and conversion to tensor, were also done by the module. Resizing: The images included in the dataset may be of various size, in order to make it comparable an image is resized to a unique size. Normalization: It takes care that the pixel level value lies in a certain range, in order to enhance convergence of the image. Tensor Conversion: The images will be required to be converted to tensors to use in deep learning models. Module Extracting Features: The module of feature extraction

takes the high-level features of the image through a pre trained ResNet-50. This involves the removal of the fully connected layer of the ResNet-50 thus enabling it to perform only feature extraction.

ViT: Transformer Based Feature Processing Module:

The module of feature processing based on transformers is formed using the Vision Transformer method that processes feature representations of ResNet-50 to give the long-range dependencies that employs the self-attention mechanism to improve feature comprehension. This substitutes the default classification layer used by ViT with three class classification. Classification Module: The classification module gives the custom fully connected layers to classify into three classes, which gives one of the three NHL variant (FL, CLL, MCL). The module makes model attain the best accuracy based on the use of powerful deep learning algorithms.

Training Module: The training module is provided with the 70 percent of the dataset with the cross-entropy loss that measures classification errors, as well as uses the AdamW optimizer that updates the weights efficiently and increases the convergence.

$$L = - \sum_C y_i \log(y_i)$$

The model has been trained using the batch size of 32 and across different epochs to benefit generalization.

Evaluation and Testing Module: The module of evaluation and testing is adopted to gauge the functioning of the trained ResNet-ViT model, by comparing the metrics of accuracy, precision, recall and F1-score to determine the efficiency of the classification. Validation of the dataset is done on the ratio of the 20 % of the dataset where the rest of the 10 % of the dataset is used to test and see which can detect better between standalone CNN and Transformer model.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

The modules are the data preprocessing module, feature extraction module, Transformer based feature processing module, Classification module, Training and Evaluation and Testing module, which will help in creation of the Hybrid Deep learning model based on ResNet and ViT.

## 5 Results

Confusion Matrix Analysis:

The confusion matrix and the testing accuracy of the model including only ViT is resulted as represented in the Fig 5 and Fig 6 respectively:

The ViT model is tested initially as the standalone model using the 10% of the entire dataset considered as the test dataset, which resulted in 82% accuracy, following which the standalone model of the Residual Network ResNet resulted 78% of accuracy upon implementing multiple epochs and the classification report as follows:

Model loaded successfully!  
 Test Accuracy: 0.8218

Fig 5 Testing Accuracy of Vision Transformer

Test Accuracy: 0.7888

Fig 6 Testing Accuracy of ResNet

An optimal accuracy was presented by the model ViT and ResNet to achieve the creation of the Hybrid model that will have a better accuracy and precision.

Hybrid model of the classification of the Non-Hodgkin Lymphoma subtypes fixed a training of 98.52 % and the evaluation of 97.67 %. When the trained model was tested and the test dataset of 303 images were processed. The accuracy to be tested comes out with 96.70 per cent accuracy that is higher than the standalone as well as the existing model. To make sure the model predicts the subtypes of NHL with a higher accuracy. The classification report of the Hybrid ResNet-50 - ViT model is:

```

Model loaded successfully!
Predictions and labels saved!
Classification Report:

```

	precision	recall	f1-score	support
CLL	1.0000	0.9810	0.9479	101
FL	0.9712	1.0000	0.9854	101
MCL	0.9352	1.0000	0.9665	101
accuracy			0.9670	303
macro avg	0.9688	0.9670	0.9666	303
weighted avg	0.9688	0.9670	0.9666	303

Fig 7. Classification report of ResNet-ViT model

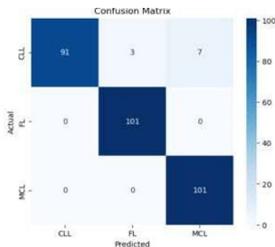


Fig 8. Graphical representing of the Confusion Matrix

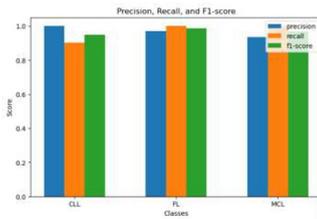
Note that Fig 9 shows a clear depiction that the histopathological images of FL and MCL has been well classified by the hybrid model but that of the was classified incorrectly. The CLL based images were misclassified by model as FL and MCL with values 3 and 7 respectively.

The bar graph that shows the CLL FL, MCL types precision, recall and f1 score:

The accuracy of the FL and MCL is relatively lower, since the model was precise in classifying the types unlike, the CLL that failed to classidify some of the 101 images.

The precision of the model outworks that of the hybrid architecture of VGG-16 + RPN (Regional Proposal Network) [2] that stood at 87.575 %.

The essential contrast of the propped hybrid system against the current one of faster R-CNN-client VGG-16.



**Fig 9.** Bar Graph representing the subtypes of NHL and their performance metrics.

**Table 1.** Comparison between Proposed System and Existing System

Feature	Proposed System	Existing System
Model Architecture	Ensemble of ViT and ResNet 60	Faster R-CNN with VGG-16
Focus of Study	Only Lymphomas	Lymphomas and Leukemia
Methodology	Image Classification using CNN and Transformer in ensemble method	Only detection with ROI and CNN layers.
Accuracy	96.5	87.575

## 6 Conclusion

The proposed hybrid deep learning hybridises the strength of ViT and ResNet-50 in classifying the subtypes of NHL through histopathological images. The model proposed is highly superior to the conventional CNN, standalone CNN and the transformer models in overcoming the limitation and challenges that the prior models are facing through the utilization of the global contextual learning and the deep residual by using the feature extraction process. The experimental test finds effectiveness of the model by meeting testing accuracy of 96.70 percent. The given hybrid model aids in supporting the pathologists to achieve effective and precise diagnosis of NHL.

## 7 Future Works

Future work in the research involves more improvements in the extraction of the feature because of the 10 of the CLL based histopathological images that has been erroneously classified. Findings will involve data augmentation strategies and interpretability of the model to be increased in order to make them more clinically applicable. Training of the model with more histopathological images and computer aided diagnosis and the efficiency of the proposed model is also a task to be completed in the future.

## References

1. Mrs. Deepthi S, Dr. Malepati Chandra Sekhar, "Non Hodgkin Lymphoma Classification using Improved Predator optimization Based Densenet121 Model", *J.Electrical Systems* 20-3(2024): 2937-2948.
2. S. A. Zaouk, T. Matar, M. A. Ali, A. Kassem and L. Hamawy, "Classification of Leukemia and Lymphoma using Faster R-CNN," 2022 International Conference on Microelectronics (ICM), Casablanca, Morocco, 2022.
3. Özgür, E., Saygılı, A. "A new approach for automatic classification of non-Hodgkin lymphoma using deep learning and classical learning methods on histopathological images". *Neural Comput & Applic* **36**, 20537–20560 (2024).
4. H. Khelil, A. El Moumene Zerari and L. Djerou, "Accurate diagnosis of non-Hodgkin lymphoma on whole-slide images using deep learning," 2022 IEEE 9th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT).
5. V. C. Pezoulas, T. P. Exarchos, N. S. Tachos, A. Goules, A. G. Tzioufas and D. I. Fotiadis, "Boosting the performance of MALT lymphoma classification in patients with primary Sjögren's Syndrome through data augmentation: a case study," 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Sydney, Australia, 2023.
6. Z. Zheng, H. Zhang, X. Li, S. Liu and Y. Teng, "ResNet-Based Model for Cancer Detection," 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE), Guangzhou, China, 2021.
7. M. G. Dahmani, M. Tarhouni and S. Zidi, "Vision Transformers (ViT) for Enhanced Skin Cancer Classification," 2024 IEEE International Conference on Artificial Intelligence & Green Energy (ICAIGE), Yasmine Hammamet, Tunisia, 2024
8. V. Raji, S. Maheswari, S. S. Dharinya, P. Chinniah, K. K. S. and K. T. Raja, "An Effective Detection of Brain Tumor Detection Using ResNet-50 Model," 2024 International Conference on Advancement in Renewable Energy and Intelligent Systems (AREIS), Thrissur, India, 2024.
9. M. Habijan and I. Galić, "Ensemble Transfer Learning for Lymphoma Classification," 2024 31st International Conference on Systems, Signals and Image Processing (IWSSIP), Graz, Austria, 2024.
10. A. A. R. Riyanto et al., "Lymphoma sub-type classification using DenseNet 169-based Transfer Learning Architecture," 2024 International Seminar on Application for Technology of Information and Communication (iSemantic), Semarang, Indonesia, 2024.
11. Archita, C. Prabha, A. Nath and R. Singh, "Enhancing Lymphoma Diagnosis: Transfer Learning with DenseNet201 for Subtype Classification," 2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAIC), Salem, India, 2024,
12. Geord Steinbuss, Mark Kriegsmann, Christine Zgorzelski, Alexander Brobeil, Benjamin, Goeppert, Sascha Dietrich, Gunhild Mechttersheimer, Katharina Kriegsmann, "Deep Learning for the Classification of Non-Hodgkin Lymphoma on Histopathological Images", *Article: PubMed Cancers(Basel)* 2021 May.
13. M. Imran, B. Haq, E. Elbasi, A. E. Topcu and W. Shao, "Transformer-Based Hierarchical Model for Non-Small Cell Lung Cancer Detection and Classification," in *IEEE Access*, vol. 12, pp. 145920-145933, 2024.
14. Sriwastawa, A., Arul Jothi, J.A. Vision transformer and its variants for image classification in digital breast cancer histopathology: a comparative study. *Multimed Tools Appl* **83**, 39731–39753 (2024).

15. H. Sahu, R. Kashyap and B. K. Dewangan, "Hybrid Deep learning based Semi-supervised Model for Medical Imaging," *2022 OPJU International Technology Conference on Emerging Technologies for Sustainable Development (OTCON)*, Raigarh, Chhattisgarh, India, 2023.
16. M. Hayat, N. Ahmad, A. Nasir and Z. Ahmad Tariq, "Hybrid Deep Learning EfficientNetV2 and Vision Transformer (EffNetV2-ViT) Model for Breast Cancer Histopathological Image Classification," in *IEEE Access*, vol. 12, pp. 184119-184131, 2024.
17. E. M. Asem Othman, "Using Pretrained Vision Transformer for Breast Cancer Binary Classification," *2023 3rd International Conference on Emerging Smart Technologies and Applications (eSmarTA)*, Taiz, Yemen, 2023.
18. M. Mohana, R. Rithika and V. Sivasakthi, "Revolutionizing Lymphoma Diagnosis with Deep Learning and Natural Language Generation," *2024 2nd International Conference on Computer, Communication and Control (IC4)*, Indore, India, 2024.
19. J. Omar Bappi, M. A. T. Rony, M. Shariful Islam, S. Alshathri and W. El-Shafai, "A Novel Deep Learning Approach for Accurate Cancer Type and Subtype Identification," in *IEEE Access*, vol. 12, pp. 94116-94134, 2024, doi: 10.1109/ACCESS.2024.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

