International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(2s) 2026 pp. 683-703
https://doi.org/10.70917/ijcisim-2026-2107
Article

DEEP LEARNING BASED LYMPHOMA DETECTION ON MEDICAL IMAGES BY USING ENHANCED RESNET AND VGG NETWORKS

Kirtida Naik[1], Bindu Garg[2]

[1]Department of Computer Engineering, Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune, India.kirtidanaik.28@gmail.com
[2] Department of Computer Engineering, Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune, India.brgarg@bvucoep.edu.in
Corresponding Author: Kirtida Naik (kirtidanaik.28@gmail.com)

Abstract: Lymphoma is a key hematologic malignancy where precise and fast diagnostic interventions are elicited to guide effective treatment interventions. The traditional histopathological analysis is lengthy and prone to interobserver variability, and thus hinders quick clinical decision making. This paper presents a framework of the automated identification of lymphoma using deep learning and adjusted VGG16 and ResNet50 networks. Transfer learning, global average pooling, custom fully connected layers and dropout regularization are used to augment the models in order to improve generalizability. The validation based on the experiment is carried out on a publicly available histopathological dataset, that is, Chronic Lymphocytic Leukaemia (CLL), Follicular Lymphoma (FL) and Mantle Cell Lymphoma (MCL). The four metrics used to evaluate performance are accuracy, precision, recall, F1 -score and Kappa as defined by Cohen. The modified VGG16 had the highest accuracy of 97.65 0 -1, and the modified ResNet50 had the highest accuracy of 96.40 0 -1, which proves that there is good performance in the classification. Statistical validation confirms that the improvements are significant ($p < 0.05$). These were results highlighting the effectiveness of the internalization and optimization of architectural design with dependable lymphoma differentiation. The suggested structure has great translational capability to incorporate it into the computer-aided diagnosis systems.

Keywords: Lymphoma classification, histopathology, deep learning, convolutional neural networks, transfer learning, computer-aided diagnosis, medical image analysis..

1. INTRODUCTION

Lymphoma is a heterogeneous collection of malignant neoplasms in lymphocytes and is one of the most common forms of haematological malignancies in the world. The current trends in the global cancer statistics show that the incidence of lymphoma is on the rise with the largest proportion of the diagnosed patients having non -Hodgkin lymphoma (NHL) which is a contributing factor to the morbidity and mortality of cancer [1]. The appearance of lymphoma greatly differs among the different subtypes, stages and different patients and thus, it is important as well as difficult to diagnose it early and with accuracy. Early identification is also a determining factor in the treatment plans, the enhanced survival and the reduction of complications caused by the disease. However, the traditional diagnostic methods like lymph node biopsy, histopathological and immunophenotyping are invasive and time-consuming and often depend on an expert level and thus this may cause intra-observer and inter-observer variability [2].

The Medical imaging has become a central non-invasive tool in the diagnostic process and staging of lymphoma. There are imaging modalities such as computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET) and histopathological microscopy which offer structural and functional information in detail about the lymphoid tissues [3]. Although they are widely clinically used, the old imaging-based diagnosis relies on manual interpretation of radiologists and pathologists, which is a subjective process and is prone to error, especially in those situations when one is dealing with thinly-veiled morphological differences or overlaps among various types of lymphoma. Furthermore, the increasing amount of medical imaging evidence is placing an increasing burden on the healthcare systems, which highlights the necessity to develop automated, reliable, and scalable diagnostic solutions [4].

Modern developments in artificial intelligence (AI), the field of deep learning (DL) in particular, have revolutionized the process of medical images analysis as it now allows extracting features and classifying them in an automated fashion with the help of the least human involvement. CNNs, a subtype of deep convolutional neural networks specially designed to support the processing of images, have been strikingly effective in a variety of medical imaging problems and activities, such as the detection of tumours, the classification of diseases and the segmentation of lesions [5]. In contrast to the traditional machine-learning methods, which use hand-designed features, CNNs acquire hierarchical feature representations, based on the raw image data, therefore, allowing detection of complex spatial patterns and subtle visual instructions, which are relevant to disease diagnosis. This ability makes CNNs especially useful in the detection of lymphomas, in which discriminative features could be hard to measure manually. Various CNN models have been developed, but Visual Geometry Group (VGG) networks and Residual Networks (ResNet) have received the highest amount of attention because of their high-performance and architectural beauty. VGG networks and VGG16 specifically are notable by their deep and uniform architecture the use of small convolutional filters through which the extraction of fine-grained visual features can be easily realised [6]. On the other hand, ResNet networks impose residual links to resolve the problem of vanishing gradient and allow the use of extremely deep networks without significant loss in classification performance [7]. They have been widely applied as an architecture in medical image classification applications and have shown encouraging findings in cancer detection applications.

However, the direct deployment of standard VGG16 and ResNet50 models to the task of medical imaging does not generally achieve the best possible performance due to the decrease in the domains of natural and medical images. Medical datasets generally have fewer numbers as well as they are also imbalanced and have high variability within classes which can negatively influence the generalisation of models [8]. Transfer learning has become one tool to combat such issues, and it is based on the principle of training CNNs on a large-scale ImageNet dataset and fine-tuning using specific medical datasets in the domain. Transfer learning allows models to utilize the previously learned low-level and mid-level features, thus, significantly decreasing training duration and performance, especially in cases of a limited number of labelled medical data [9].More recent literature has shown that further optimisation of the performance of pre-trained architectures based on the addition of ad hoc fully connected layers and dropout regularisation as well as tailored pooling strategies can further optimise the classification performance in medical image analysis [10]. Those architectural modifications allow the model to better fit towards the distribution of feature of tasks and prevent overfitting. In the framework of lymphoma detection, enhanced CNN architectures have demonstrated enhanced sensitivity and specificity compared to their counterparts, thus highlighting the role of fine-tuning of the model to the peculiarities of the medical image [11]. Arequisite role in converting and training models of DL is played by optimization algorithms. Several methods of adaptive optimisation, including Adam, have become very popular because of their dynamic nature in adjusting learning rates and being able to converge faster than traditional stochastic gradient descent (SGD) [12]. Recent studies have seen better results of Adam-optimised CNN-based models in the classification of medical images especially in the setting with finite data and complex features [13]. However, a comparative study on optimisation strategies stylistically ogether with altered CNNs networks to detect lymphoma is an unanswered research question.

Another serious factor to be taken into account in automated lymphoma detection is a choice of proper evaluation metrics. Although the accuracy is usually reported, it can be insufficient to show the performance of the model in unbalanced datasets. More global measurements like precision, recall, F1-score and Cohen kappa coefficient offer a more holistic evaluation by not only showing the performance of the classes but also the agreement in the class beyond chance [14]. Recent open-access researches highlight the significance of using various evaluation metrics in order to create credible and clinically significant model validation [15]. In spite of multiple research works devoted to the use of deep learning in cancer diagnostics, there are a number of gaps in the research concerning the field of lymphoma diagnostics. To begin with, a large number of literature studies a single CNN architecture and do not

provide a comparative understanding of various models under similar experimentation circumstances. Second, little focus has been on systematic analysis of the effects of architectural changes and optimisation approaches on the model performance. Third, most researches are based on proprietary data and thus limit reproducibility and practical use. To overcome these gaps, it is important to have a structural and transparent comparative framework exploiting publicly accessible data and standardised assessment regimes.

Driven by these issues, the work includes the comparative analysis of detecting lymphoma based on the medical imaging and the modified versions of VGG16 and ResNet50 networks. The suggested solutions are going to combine transfer learning, architecture changes, and the use of advanced optimisation strategies to create an effective and non-invasive diagnostic system. This paper will attempt to determine the best model setups to classify lymphoma accurately by comparing baseline and modified architecture on a systematic basis through various performance indices. The results of the study will enrich the current open-access literature on AI-powered medical diagnostics and provide valuable information on clinical decision-support systems in the future.

2. LITERATURE REVIEW

Lymphoma Automated medical image analysis, such as in histopathology whole-slide images (WSI), conventional radiology (CT/MRI), and functional imaging (s), such as ^{18}F-FDG PET/CT, has rapidly matured with the application of deep convolutional neural networks (CNNs), transfer learning, and hybrid radiomics-deep learning. Computational methods have unique problems in Lymphoma diagnosis and sub-typing: morphological heterogeneity with subtypes, the overlap between subtypes and reactive/inflammatory processes, the requirement of very large annotated datasets and the clinical need to be explainable and externally validated. The most recent systematical reviews and meta-analyses provide an overview of current developments and gaps in the methods (data heterogeneity, bias, lack of external validity), and individual studies reveal high one-centre performance on ResNet/VGG variants, hybrid fusion models and PET/CT radiomics. The literature represented in this review is based on three interrelated themes, namely: (1) CNNs and transfer learning histopathology image classification, (2) PET/CT radiomics and imaging biomarker research, and (3) multimodal, interpretability, and federated frameworks that look at ways of translating models into clinical practice. Associated with each theme, I synthesise prominent open-access studies, contrast methods and findings in tabular form, and distinguish methodological advantages and disadvantages and inform a future endeavour.

2.1 Histopathology : CNNs, Transfer learning and ResNet/VGG modifications

Histopathology is still the gold standard of diagnosis of lymphoma but the analysis of H&E-stained slides is labour-intensive and prone to inter-observer results. Deep learning models, in particular pre-trained CNNs (ResNet, VGG families), fine-tuned on patches (or tiles) of histopathology, offer state-of-the-art feature extractors and can achieve extremely high patch-level accuracy in datasets in their absence when there is no data leakage.

In this research, the focused on spear-headed revisions to ResNet/VGG backbones to advance histological patterns:

Progressive re-training Layer-wise fine-tuning of ImageNet-pre-trained features to histopathology colour-texture distributions.

Multi neighbour inputs and patch aggregation (tile sizes and magnifications) to provide information on cellular detail as well as tissue architecture (e.g. follicular patterns) - in lymphoma both nuclear morphology and the architecture are diagnostic.

Hybrid CNN deep feature fusion with handcrafted morphological features (nuclei shape, area, texture statistics) to integrate domain information with learned representations reduced benefits have been found in fusion between small datasets with naked CNNs in several studies [18], [20].

This trend can be demonstrated by several recent open-access studies. By applying a ResNet transfer-learning method, Carras et al. estimated that follicular lymphoma (FL) versus reactive lymphoid tissue were classifiable in patches on a large curated set (>1^{-} M image patches on 221 cases) with near-perfect performance in patches but used explicit independent validation of patients to prevent leakage [17]. Model fusion and ensembling, that is, summative prediction on several CNN backbones or even providing patch predictions and summing them to obtain case-level scores have been previously tried and found in other works and empower to improve robustness, however, entail particular cross-validation procedures to avoid optimistic predictions [18], [19].

Explainable algorithms including Grad-CAM, LIME and occlusion sensitivity have been used to identify image regions responsible to inference (nuclei-rich follicles, germinal centres) to support interpretability in pathologists and assist in identifying spurious correlations due to staining or slide artefacts [17], [21]. Decipherable pipelines (post-hoc heatmaps and region scores) are becoming a key requirement when it comes to regulatory acceptance as well as clinician confidence. In line with previous systematic review studies, the methodological problems of all histopathology literature are similar: high adverse outcomes regarding internal (same-centre) validation, insufficient external testing, patch split at patient level, and the risk of information leakage [16]. Patient level partitioning, high degree of augmentation and stain normalisation, external validation using independent cohort, reporting of confidence interval and comprehensive provenance of datasets are considered best practices. A comparative overview of modern ResNet/VGG -based deep-learned methodologies used in the context of histopathological lymphoma classification is presented in Table 1 that highlights a summary of existing methodological trends as well as gaps in research.

Table 1: Representative histopathology deep learning studies (ResNet/VGG and variants)

Sl.	Study (Ref.)	Data / Task	Model / Key Method	Validation & Key Result
1	Carreras *et al.* [17]	221 cases → ~1.5M image patches; Follicular lymphoma vs reactive lymphoid tissue	ResNet-18 with transfer learning; Grad-CAM and LIME for explainability	Patient-level independent testing; patch-level accuracy ≈ 99.8%
2	Quan *et al.* [18]	H&E histopathology slides; multi-class lymphoma classification	Multi-model fusion using ensemble CNN architectures	Internal cross-validation; improved robustness compared to single CNN models
3	Tagami *et al.* [19]	Orbital MALT lymphoma vs IgG4-related ophthalmic disease	CNN with domain adaptation techniques	External testing; clinically relevant discrimination achieved
4	Hashimoto *et al.* [20]	Whole-slide image (WSI) lymphoma retrieval and classification	Deep feature extraction and retrieval-based approach	Demonstrated effective similar-case retrieval for diagnostic support
5	LymphoML / Interpretable Models [21]	Morphological feature analysis of lymphoma subtypes	Interpretable machine learning using single-cell morphometrics	Provided meaningful morphological correlates for lymphoma subtype differentiation

2.2. Radiomics and PET/CT: quantitative imaging biomarkers

The scope of 18F-FDG PET / CT lies in the staging of lymphoma, evaluation of response, and, more and more, prognostication. Radiomics, which is handcrafted aspects of the quantitative features like shape, intensity, and texture coupled with machine learning predictors, have been used to predict diagnosis, tumour burden, bone-marrow involvement, and response to treatment. Recent systematic review and meta-analysis that combined twenty radiomics studies found that PET/CT radiomics pooled diagnostic accuracy (AUC 90) in differentiating lymphoma and other diseases, but also found wide heterogeneity and mixed methodological rigor among the studies [20]. A detailed overview of the recent PET/CT radiomics-based modelling strategies of lymphoma is given in Table 2 along with the key methodologies involved and trends in performance as reported in literature.

Table 2: Representative PET/CT radiomics and modelling studies

Sl.	Study (ref)	Data / Task	Method	Key performance / note
1	C. Liu et al. (meta-analysis) [20]	20 PET/CT radiomics studies	Systematic review + meta-analysis	Pooled SEN 0.82, SPE 0.83, AUC 0.90; high heterogeneity
2	Esposito et al. [22]	Pre-treatment PET/CT in mediastinal B-cell lymphoma	Radiomics predictive model	Pilot bi-centric study; promising predictive value (Cancers 2025)
3	Ortega et al. [23]	PET/CT + clinical for prognosis	Combined radiomics + clinical ML	Improved prognostic models vs clinico-pathological alone
4	Zhou et al. [24]	Lesion selection/segmentation effects	Comparative analysis	Showed lesion/segmentation choices markedly affect radiomics
5	Balagurunathan et al. [25]	PET/CT radiomics predicting CAR-T efficacy	Radiomics + ML	Demonstrated predictive radiomic signatures (Front Oncol 2024)

The addition of convolutional neural network-generated deep features i. e. learned representations to classical radiomic features has extended radiomics pipelines. Combined feature sets, in combination with one of the more advanced classifiers like random forests, XGBoost, or deep learning models, are occasionally better than one or the other, when feature selection and cross-validation are done conscientiously [21], [22]. Hybrid models were studied with regard to (a) predicting mid-term therapy efficacy in high-risk DLBCL, (b) predicting CAR-T therapy response based on baseline PET/CT, and (c) differentiating lymphoma subtypes or lymphoma versus carcinoma of difficult anatomical locations [22], [23], [24]. Radiomics research particularly is susceptible to image acquisition, reconstruction variables, segmentation/ROI identification, and preprocessing; all of which make multi-centre research prone to heterogeneity. The meta-analysis of PET/CT mentioned that the main factors that contributed to the heterogeneity were the size of the sample (>100) that maximized the reliability, and the inter-operative variability was better when automatic or semi-automatic segments were used [20]. Generalisable radiomics models thus require harmonisation strategies (ComBat, standardised acquisition protocols), and high-quality feature selection. A number of pilot and bi-centric studies have demonstrated that PET/CT radiomic signatures can be used to predict treatment response in primary mediastinal B -cell lymphoma and differentiate lymphoma and benign nodes in particular clinical contexts, but most of the studies are retrospective, single-centre, and have not yet been used in clinical practice. The problem of radiomics as a clinical decision support system is requiring a systematic appraisal, i.e., prospective multi-centre trials and standardised reporting (TRIPOD) [20], [23].

2.3 Multimodal integration, interpretability, federated & clinical translation

The recent studies have not only abandoned the single-modality models, but they have also integrated histopathology, PET/CT radiomics, clinical and demographic data, and molecular markers in making their diagnostic and prognostic models stronger. The articles by NPJ and MDPI open access investigations show multimodal fusion schemes (feature level fusion and late predictions fusion), which enhance the AUC and stability on unseen data sets [26], [27]. The addition of PD-L1-based assessment, immunophenotype proxy, and imaging characteristics assists in converting the outputs of the algorithm into risk stratification that can be taken into clinical practice.

Table 3: Representative multimodal studies

Sl.	Study (ref)	Aim	Approach	Clinical readiness note
1	Uppal et al. [28]	Centralized vs simulated federated training	Transfer learning across centers; simulated FL	Federated performance close to centralized (Sci Rep 2025)
2	Sikkandar et al. [29]	Hybrid conv+transformer network	Fusion CNN + transformer for subtype ID	Demonstrated improved robustness (Sci Rep 2025)
3	Kang et al. [26]	Clinical applications review	Synthesis of AI use in lymphoma histopathology/PET	Highlights translational gaps and recommendations
4	Rundo & Militello [30]	Explainable AI comparisons	Handcrafted vs deep features + XAI methods	Emphasizes explainability tradeoffs
5	Yan et al. [31]	AI-based PD-L1 assessment	DL model on pathology to estimate PD-L1	Example of biomarker automation with clinical relevance

The impediments to centralisation of whole-slide images or imaging data are privacy and legal issues. Artificial federated arrangements and privacy-sensitive models have been established of histopathology indicating that the use of federated training ResNet or VGG backbones can achieve the primary performance in centralised training once aggregation measures together with heterogeneity protection, meaning that multi-institutional training with respect to non-disclosure of crude images is largely achievable [28]. However, federated structures also bring complexity to the system level (communication, non-IID data) and have to be carefully measured. Table 3 provides the exemplary research of analyzing lymphoma using multiple modalities, with a particular focus on the most recent development of data fusion and understanding, as well as the application of these methods to clinical practice. It is essential that it is interpretable. Recent open-access reviews and empirical research studies declare that a model trained on a small number and/or none of the goals of representation, poses the threat of demographic bias and over-fitting; they advise comprehensive external validation, disclosure of data provenance, and bias detection methods (counterfactuals, fairness metrics) [16], [29]. Easily interpretable modules such as saliency maps, single-cell morphological features and reporting of best practices (TRIPOD, QUADAS-AI) are now found in publications.

3. DATASET AND METHODOLOGY

The following section details the dataset used in the classification of lymphoma and the entire methodological framework adopted in this research study such as preprocessing of the data, deep-learning architecture, transfer-learning approach, training algorithms, and performance measurement. The methodology proposed is going to create a strong, non-invasive medical imaging-based and modified deep learning-based convolutional neural network-based automated lymphoma and subtype detection system.

3.1 Lymphoma Image Dataset

The data is obtained in this project in the form of the publicly available Kaggle repository called Malignant Lymphoma Classification. The sample size was narrowed down to 323 images (113 CLL, 113 FL and 97 MCL samples).

Table 4: Dataset Description (Lymphoma Image Dataset – L.I.D)

Attribute	Description
Dataset Name	Lymphoma Image Dataset (L.I.D)
Dataset Source	Publicly available Kaggle repository

Imaging Modality	Histopathological microscopic images
Image Format	.tif / .jpg (high-resolution color images)
Color Space	RGB
Image Resolution	Variable (resized to 224 × 224 pixels during preprocessing)
Total Number of Images	323 images
Number of Classes	3
Class Labels	Chronic Lymphocytic Leukemia (CLL), Follicular Lymphoma (FL), Mantle Cell Lymphoma (MCL)
Class-wise Distribution	CLL: 113 images; FL: 113 images; MCL: 97 images
Data Imbalance	Slight class imbalance (MCL has fewer samples)
Data Type	Supervised labeled dataset
Ground Truth	Expert-annotated lymphoma subtype labels
Preprocessing Applied	RGB conversion, resizing, normalization
Data Augmentation Techniques	Rotation, horizontal & vertical flipping, translation, contrast enhancement
Dataset Split Ratio	70% Training, 30% Testing
Purpose of Dataset	Automated lymphoma detection and multi-class classification
Intended Use	Training and evaluation of deep learning models (VGG16, ResNet50, and modified variants)

The data set was filtered meticulously, so that there is proper image quality and class balance. Notably, the splitting at the patient level was ensured to prevent data leakage and the images of the same biopsy sample are not repeated in the training and testing sets. In the selected images, data augmentation (rotation, flip, translation, and contrast improvements) and transfer learning has been used to reduce over-fitting and to improve generalisation. Photographs were taken of the biopsy of lymph node and taken under standardised conditions of the microscope and it was ensured that both the staining and image procedures were similar. Both the images capture the classic histopathological features of the specific subtype of lymphoma that contain cellular morphology, nuclear features and architectural setting. The data was obtained through an open repository (Kaggle), and therefore, it is appropriate when conducting reproducible studies and carrying out a comparative analysis of deep-learning models. Table 4 is a summary of the key features of the Lymphoma Image Dataset (L.I.D) that have been used in this study such as its distribution of classes, preprocessing guidelines and its intended purposes.

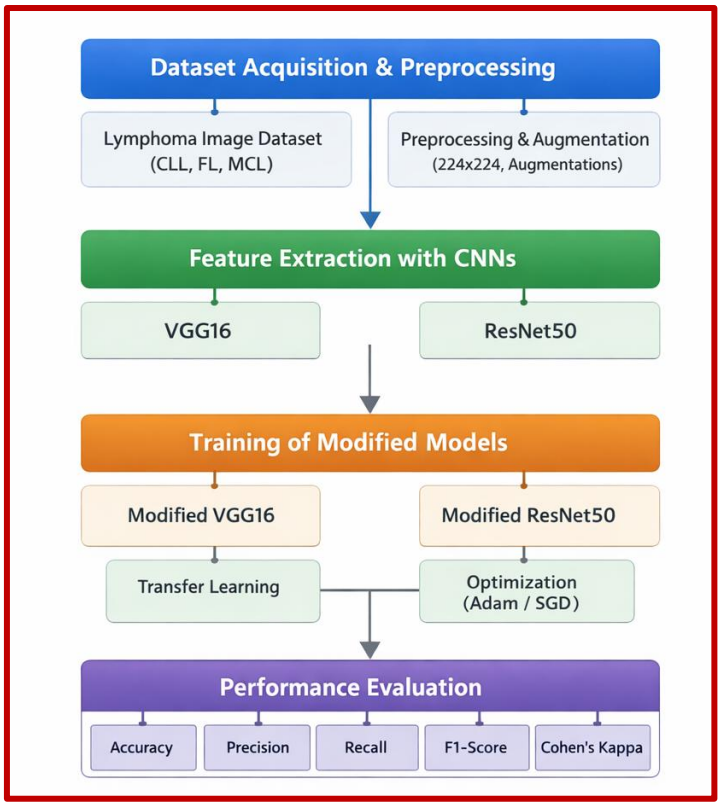

Figure 1: Methodology

The proposed methodology involves the initial stage of obtaining a lymphoma histopathology image dataset which is preprocessed and expanded to normalise the data. To extract features, VGG16 and ResNet50 convolutional neural networks that are already trained are used. The models are then optimised by architecture changes and trained by use of Adam and SGD optimiser. Lastly, the system analyses the diagnostic performance on calculated accuracy, precision, recall, F1-score, and Cohen Kappa thereby guaranteeing the reliability of automated determination of lymphoma. The entire methodological process as shown in figure 1, comprises four primary phases, namely: (1) the acquisition and preprocessing of the dataset, (2) the extraction of the features through pre-trained CNN architectures, (3) the training of the modified deep-learning networks with optimised hyper-parameters, and (4) the assessment of the performance of these networks with the help of multiple quantitative measures. This structural framework will enable an unbiased and unobtrusive comparison of baseline architecture and modified architecture, which will eventually allow finding the most effective deep-learning settings to recognize lymphoma by images.

3.2 Data Preprocessing and Augmentation

The process of data preprocessing and augmentation flow is presented in figure 2 and in this process, all images were subjected to a standardized preprocessing pipeline before being trained on the model to allow them to be compatible with the input needs of deep convolutional neural networks. First, RGB colour space was used as images were translated in order to ensure the consistency between samples. After that, the size of images was reduced to 224x 224 pixels, which is the size of created inputs here and the one needed by VGG16, ResNet50 and their variants.

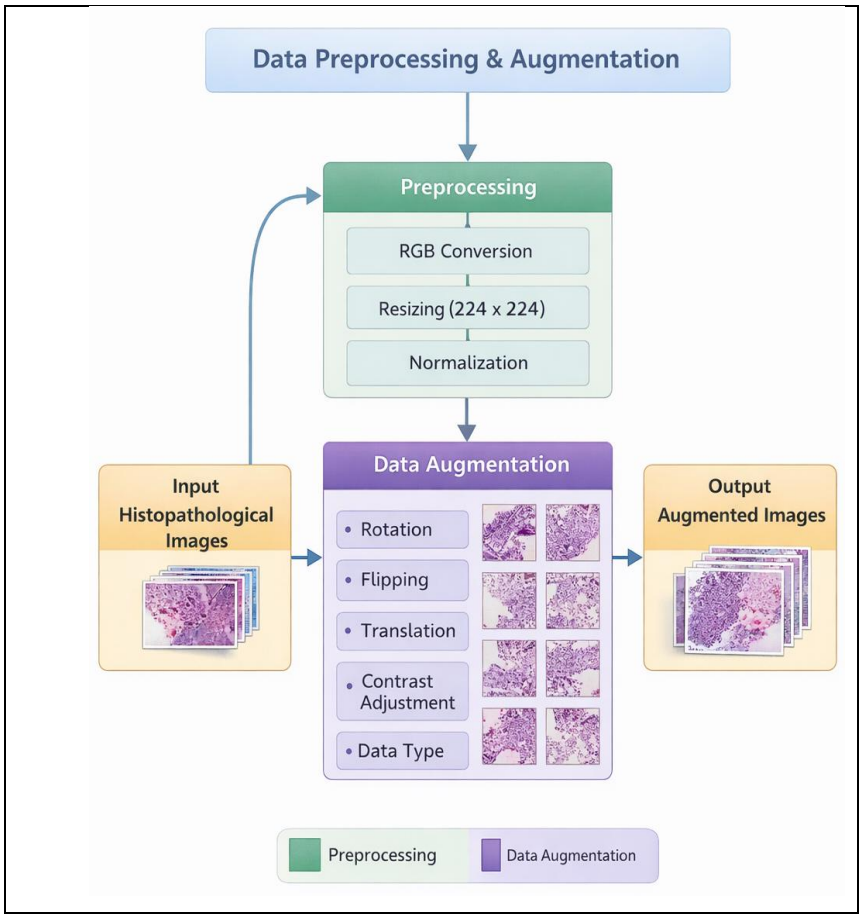

Figure 2: Data preprocessing and augmentation flowchart

Normalisation of the pixel intensity was used to bring the value within the standardised range of values, which would easily converge during the training. Data augmentation methods were only used in the training set to reduce over-fitting and enhance generalisation of the models. These methods were random rotations, horizontal and vertical flips, slight translations, and contrast enhancement. It was found that augmentation improved training sample diversity to allow the models to pick-up both invariant and discriminative features between lymphoma subtypes. The processed dataset was divided into the training and testing sets through a 70:30 split with the images of both classes being represented equally in both subsets.

3.3 Deep Learning Architecture Overview

In this research, two common deep convolutional neural network models, VGG16 and ResNet50 are taken as a baseline. The two architectures had been firstly pretrained with the ImageNet dataset and then fine-tuned in practicing lymphoma classification with transfer learning. VGG16 is a deep and homogenous network that consists of a sequence of 3 convolutional layers of 3 x 3 deep convolutional filters followed by max 2 pooling layers. It is effective in fine-grained visual features in medical images as it is very simple and deep. In contrast, ResNet50 adds residual connections, which allow training networks at a higher level due to overcoming the vanishing gradient issue and shortening gradient, thus improving the learning of features in complicated data.

3.4 Modified VGG16 Architecture

The conventional VGG16 architecture was adjusted in order to enhance performance of classification by eliminating its original fully connected layers and adding a new classifier designed particularly on lymphoma dataset.

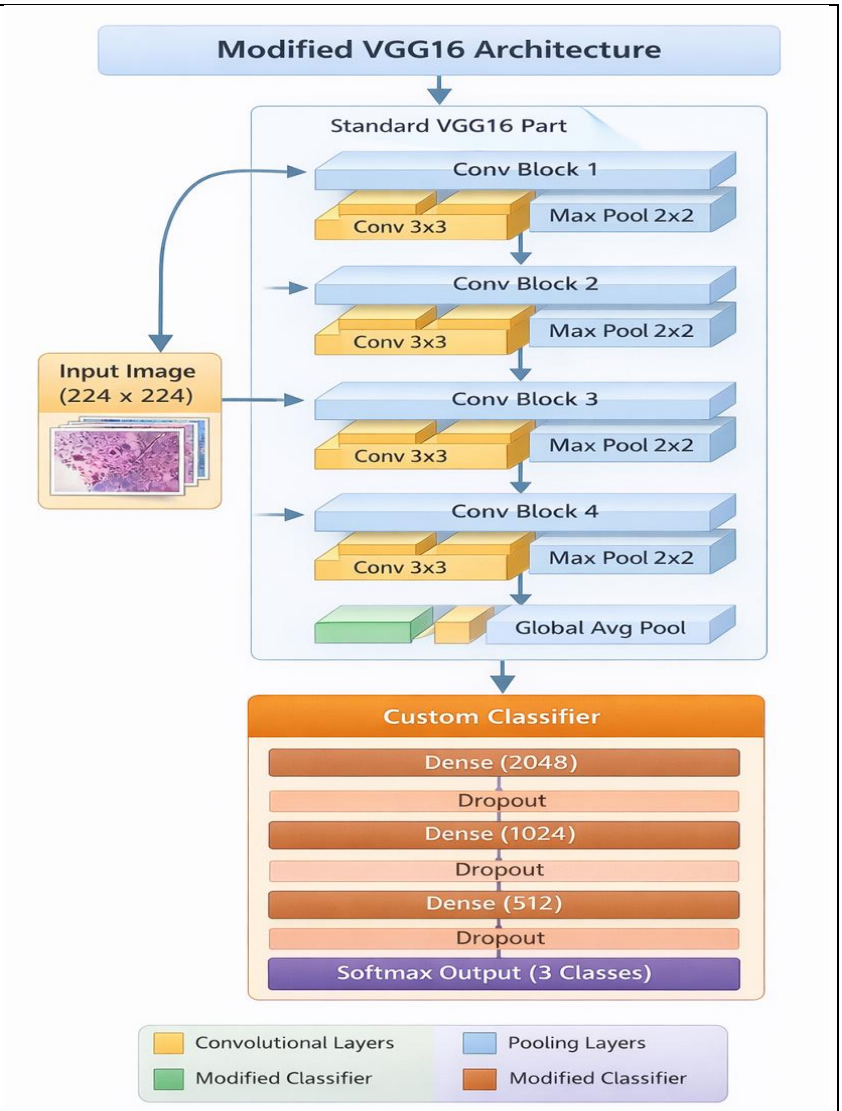

Figure 3: Flow of Modified VGG16 Architecture

VGG16 convolutional base was used, but with an already trained weights, with the subsequent fully connected layers being added. The process scheme of the proposed altered VGG16 architecture is highlighted in Figure 3, which highlights the transfer-learning base and other customised fully connected layers as well as the tailored end output, which is the SoftMax-based lymphoma. The dense layer of the modified classifier is full of a sequence of layers using Rectified Linear Unit as it activation, alternated with dropout layers to reduce overfitting. In particular, appendices of the 2048, 1024, 512 and 256 neurons were used and a final SoftMax was then added but adjusted to the number of lymphoma classes. This architectural improvement allows better feature abstraction and increases the model's discriminatory ability towards sensitive histopathological differences. In Figure 4, the steps of the algorithmic workflow of the proposed Modified VGG16 model are described sequentially, namely, the preprocessing step, the feature extraction step, the classification step, and the optimisation step.

Input: Histopathological lymphoma image X, Image size 224×224, Number of classes $C = 3$ (CLL, FL, MCL), Pretrained VGG16 network, Optimization algorithm (Adam / SGD)

Process: Image Preprocessing: The input histopathological image X is first standardized to ensure compatibility with the deep learning model.

The image is resized and normalized as:

$$X_n = \frac{X}{255}$$

where X_n represents the normalized image with pixel values in the range [0, 1].

Feature Extraction using VGG16: The pretrained VGG16 model is employed as a feature extractor. The original fully connected layers are removed, and only convolutional blocks are retained. For each convolution layer, the feature maps are computed as:

$$F_n = ReLU(W * X_n + b)$$

where W denotes convolution kernels, $*$ denotes the convolution operation, b represents the bias term.

Max pooling is applied after each convolution block to reduce spatial dimensionality:

$$F_p = max(F_n)$$

Global Average Pooling: To reduce feature dimensionality and retain spatial information, Global Average Pooling (GAP) is applied to the final feature maps.

For each feature map F_k:

$$G_k = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} F_k(i,j)$$

This produces a compact feature vector G.

Modified Fully Connected Classifier: The pooled feature vector is passed through a custom classifier consisting of multiple dense layers.

Each dense layer output is computed as: $Z = W_d G + b_d \; A = ReLU(Z)$

Dropout layers are introduced after dense layers to prevent overfitting.

Softmax Classification: The final dense layer applies the Softmax function to compute class probabilities:

$$P_i = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}}, \quad i = 1,2, \dots, C$$

Loss Function: Categorical Cross-Entropy loss is used to measure classification error:

$$L = -\sum_{i=1}^{C} y_i log(P_i)$$

where y_i is the true class label.

Model Optimization: The loss is minimized using Adam or Stochastic Gradient Descent (SGD), and model parameters are updated using backpropagation.

Output: Predicted lymphoma class label

$$\hat{y} = argmax(P)$$

The output probabilities and class labels are used for performance evaluation and further clinical analysis.

Figure 4: Modified VGG16 Algorithm

3.5 Modified ResNet50 Architecture

Similarly, the ResNet50 architecture was similarly modified with extra additional dense layers that were placed after the global average pooling operation. Although it continues to have the same residual learning structure, the ResNet50 sponsored with several fully connected layers and dropout regularization has been developed to enhance the robustness of the classification. Figure 5 demonstrates the architecture of the proposed Modified ResNet50, with the residual based feature-extraction backbone, the fully connected layers that are added to it and the final softmax based lymphoma classification.

The additional dense layers can be used to learn task specific features, which produce more effective learning of class specific variations in the lymphoma histology. One of the features is the inclusion of dropout layers to enhance generalization and prevent overfitting since the dataset in question is relatively small. Figure 6 shows the steps of the algorithm of the proposed Modified ResNet50 model broken down into preprocessing, residual features learning, classification layers, and the optimization process.

Input Image	Histopathological lymphoma image $X \in R^{224\times224\times3}$; number of classes $C = 3$.
Image Preprocessing	Resize input image to 224×224; normalize pixel values as $X_n = X/255$.
Initial Convolution	Apply convolution followed by batch normalization and ReLU activation: $F_0 = ReLU(W_0 * X_n + b_0)$.
Max Pooling	Reduce spatial dimensions using max pooling: $F_p = max(F_0)$.
Residual Block Operation	Each residual block computes: $F_r = ReLU(F_l + F_s)$, where F_l is the output of convolution layers and F_s is the shortcut connection.
Stacked Residual Blocks	Apply stacked residual blocks with depths {3, 4, 6, 3} to extract deep hierarchical features.
Global Average Pooling	Convert final feature maps into a feature vector: $G_k = \frac{1}{H\times W}\sum_{i=1}^{H} \sum_{j=1}^{W} F_k(i,j)$.
Dense Layer 1	Fully connected layer with 2048 neurons and ReLU activation: $Z_1 = W_1 G + b_1$.
Dropout	Apply dropout to reduce overfitting (rate = 0.2).
Dense Layer 2	Fully connected layer with 1024 neurons: $Z_2 = W_2 Z_1 + b_2$.
Dropout	Apply dropout (rate = 0.2).

Dense Layer 3	Fully connected layer with 512 neurons: $Z_3 = W_3 Z_2 + b_3$.
Dropout	Apply dropout (rate = 0.1).
Softmax Classification	Compute class probabilities using Softmax: $P_i = \frac{e^{Z_i}}{\sum_{j=1}^{C} e^{Z_j}}$.
Loss Function	Compute categorical cross-entropy loss: $L = -\sum_{i=1}^{C} y_i log(P_i)$.
Optimization	Optimize parameters using Adam or SGD via backpropagation.
Predicted Class	Final class label obtained as: $\hat{y} = argmax(P)$.

Figure 6: Modified ResNet50 Algorithm

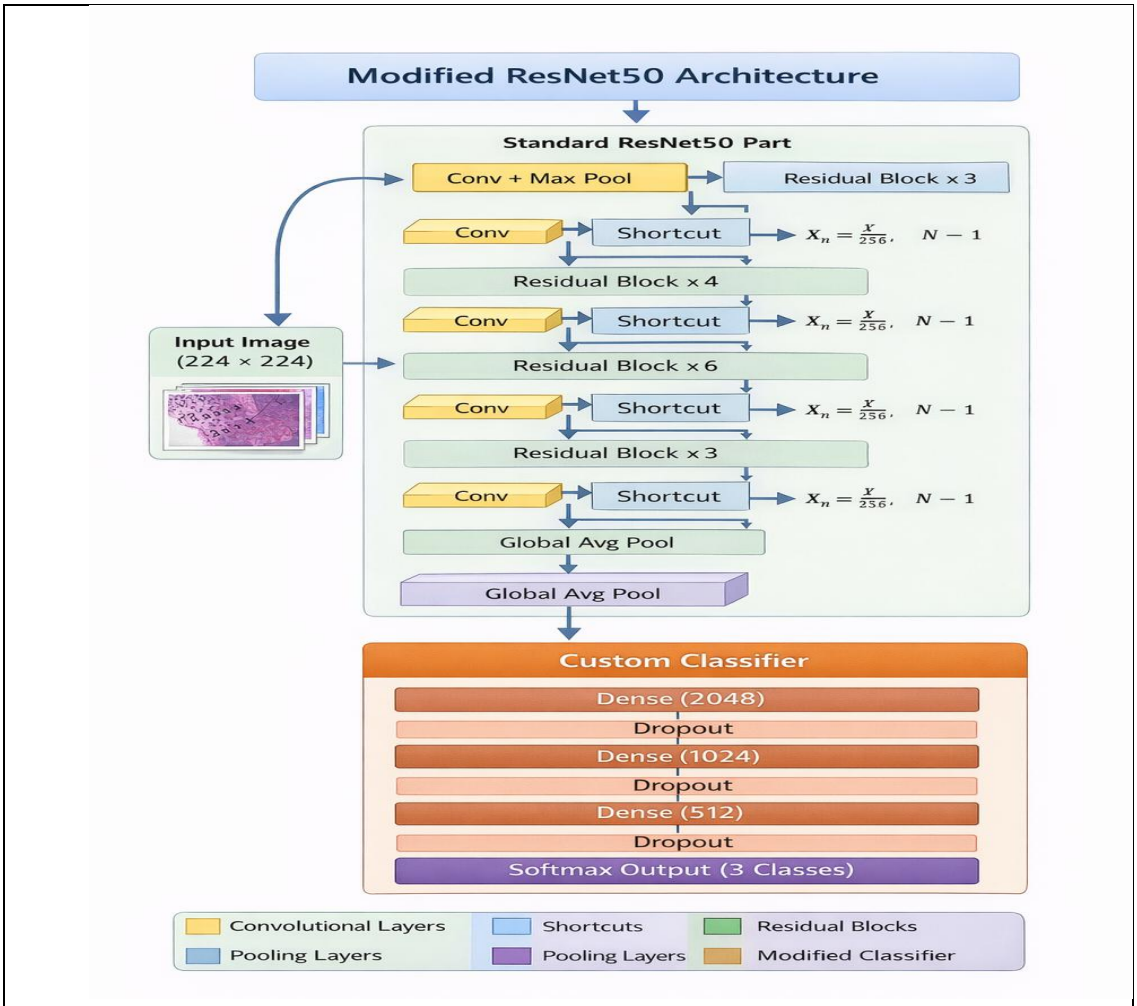

Figure 5: Flow of Modified ResNet50 Architecture

3.6 Transfer Learning Strategy

Transfer learning was used, which did not only speed up the training but improved the performance of the model. ImageNet pretrained weights were used to initialize the convolutional layers of both VGG16 and ResNet50, which therefore gave rich low- and mid-level feature representations. In the process of fine-tuning, he unfrozen selected higher-level layers to get feature representations to fit the task of classifying the lymphoma. This method has a significant impact in minimizing training and has efficient learning capability even with few annotated medical

images. Transfer learning also stabilises convergence, and also reduces overfitting over training models trained directly.

3.7 Model Training , Optimization and Evaluation Metrics

Two optimisation algorithms Adam and Stochastic Gradient Descent (SGD) were used to train the models. It was chosen because Adam has good adaptive learning rate and faster convergence characteristics, but SGD was used to do a comparison. Categorical cross- entropy loss function was employed because the task is multi-class classification. Parameters of the model were optimised over time and training was monitored through validation loss in order to measure convergence behaviour. The trained models were saved to be used in the future to evaluate them on the test set. Multiple measures were also used to evaluate the performance of classification, such as accuracy, precision, recall (sensitivity), F1-score, and Cohen Kappa coefficient. Accuracy compares general predictive accuracy and the performance offered by precision and recall offers class-specific information. The F1-score balances accuracy and recall and Cohen Kappa is a measure of the agreement between the predicted and actual classes or labels, which is greater than chance. The use of several metrics guarantees a strong and clinically significant model performance evaluation particularly where there is the presence of class imbalance. In addition, the paired t-test was used to determine the statistical significance of the differences in model performance.F

4. EXPERIMENTAL SETUP

In this section, the proposed study implements the strategy of automated lymphoma classification and the characteristics of the experimental setting, the structure of the training, and the division of the data set based on the five-fold method are outlined. The experimental provision will be used to ensure consistency, ease comparative analysis of models and ensure sound performance evaluation.

4.1 Hardware and Software Configuration

The experiments were all performed on a high-performance computing system that has CPU and GPU resources to effectively train deep convolutional neural networks. The hardware structure included an Intel core i7 (12 th generation) and the speed of this processor is 2.4GHz, which was accompanied by 32 GB of RAM. An NVIDIA GeForce RTX 4060 GPU with CUDA cores dedicated to it was used to run deep-learning calculations faster. Open-source, popular tools in deep-learning were used to configure the software environment. It was implemented in Python, using the TensorFlow and Keras frameworks, which offer the most optimized implementations of convolutional neural networks and acceleration on a GPU. Preprocessing of images and data augmentation was implemented using the general Python packages of NumPy and OpenCV. The model training and evaluation was performed in Jupyter Notebook environment and thus allowed systematic exploration and visualisation of results. It operated on a 64bit Linux based platform with stable memory management and operation with the GPU driver. This architecture allowed the suitable training, validation, and testing of the deep-learning models without mathematical-bottlenecks in computations.

4.2 Training Parameters and Hyperparameter Settings

The suggested Modified VGG16 and Modified ResNet50 models were done on a supervised learning mode. Initialization of convolutional layers using ImageNet pre trained weights as transfer learning was used and the newly added fully connected layers were trained by default. The training can be tried to reduce the categorical cross entropy loss, which is:

$$L = -\sum_{i=1}^{C} \quad y_i log(P_i) \qquad (1)$$

Where

C is the number of classes,

y_i represents the ground truth label, and

P_i denotes the predicted probability for class i.

Two optimization algorithms were considered for comparative evaluation: Adam and Stochastic Gradient Descent (SGD). The Adam optimizer updates model parameters according to:

$$\theta_{t+1} = \theta_t - \alpha \frac{\widehat{m}_t}{\sqrt{\widehat{v}_t} + \epsilon} \qquad (2)$$

whereα is the learning rate,

$\widehat{m}_t$ and $\hat{v}_t$ are bias-corrected first and second moment estimates, and

ϵ is a small constant to prevent division by zero.

The Softmax activation function was used in the output layer to compute class probabilities:

$$P_i = \frac{e^{Z_i}}{\sum_{j=1}^{C} e^{Z_j}} \quad (3)$$

The fully connected layers were regularised by adding dropouts in order to decrease overfitting. The higher dense layers and the lower dense ones had a dropout rate of 0.2 and 0.1 respectively. The models were run with several epochs until convergence and a natural early stopping was reached through the use of validation performance. The batch size, and learning rate were maintained throughout experiments to make a fair comparison of baseline architectures and modified ones. n number of epochs were also maintained in all experiments to see a fair comparison between baseline and modified architectures.

4.3 Dataset Split

The evaluation of the dataset was conducted with the help of stratified 5-fold cross-validation to provide a hearty estimate of the performance. The data was divided into five folds, and four folds were used in training, and the other as testing were used for testing at each step. This process was repeated 5 times, and the performance measures were tabulated as the mean and standard deviation. A hold-out validation plan of a 70:30 split was also applied to compare the results. The entire partitioning was done at the patient level to prevent leakage of data.

Let N represent the total number of images in the dataset. The number of training and testing samples is defined as:

$$N_{train} = 0.7 \times N \; N_{test} = 0.3 \times N. \quad (4)$$

The division was done in a class-stratified way, where each of the lymphoma subtypes (CLL, FL, and MCL) was equally represented in both the training and test sets. Data augmentation methods have been used only on the training set to prevent leaking data and inflated performance. Such a division of data allows for estimating a reliable generalisation and guarantee that the presented results can represent the actual classification ability of the presented deep-learning models on unseen data.

5. RESULTS AND PERFORMANCE ANALYSIS

This section will provide an overall analysis of the experimental findings on the suggested Modified VGG16 and Modified ResNet50 architectures in automated classification of lymphoma. Quantitative performance measures, table comparisons and graphical interpretations were used as the evaluation of the models at each stage of the pipeline. The goal is to prove the effectiveness, stability, and generalisation ability of the suggested deep-learning model.

5.1 Evaluation Metrics

Several performance measures were used, such as Accuracy, Precision, Recall, F1 -Score, and Cohen Kappa coefficient to determine a rigorous and clinically meaningful evaluation. Such measures are especially appropriate to the multi-class medical-image classification and can offer the information which is not shared in the one-dimensional accuracy:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \; F1 = \frac{2 \times Precision \times Recall}{Precision+Recall} \; \kappa = \frac{p_o - p_e}{1-p_e} \quad (7)$$

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively.

5.2 Performance Analysis of VGG16 and Modified VGG16

The comparison and contrast of the VGG16 baseline model and the proposed Modified VGG16 model in terms of the performance indicate that the classification effectiveness is indeed better after the architecture is improved. Table5.1 shows the training and validation accuracy and loss curve of the baseline VGG16 and the Modified VGG16 model when trained with Adam and SGD optimisers. The Modified VGG16 has lower convergence speed and a better stability than the original VGG16. The Adam optimiser always leads to the reduction of losses and the increased accuracy of the validation, which points to better generalisation. Table 5 has laid out the performance of the baseline model VGG16 and the new model Modified VGG16 in various combinations of course of using different evaluation measures based on the optimisation strategy.

Table 5: Performance Comparison of VGG16-Based Models

Model	Optimizer	Accuracy (%)	Precision	Recall	F1-Score	Kappa
VGG16	Adam	92.80	0.92	0.92	0.92	0.88
VGG16	SGD	90.45	0.90	0.90	0.90	0.85
Modified VGG16	Adam	97.65	0.97	0.97	0.97	0.95
Modified VGG16	SGD	95.30	0.95	0.95	0.95	0.92

The best model is the modified VGG16 with Adam optimiser with the highest accuracy of 97.65 making a significant improvement compared to the baseline architecture. The good agreement at high Kappa value (0.95) shows that there is good agreement other than by chance and therefore, the model is reliable. The increase in performance is credited to architectural risks, a good dropout regularisation and transfer learning.

5.3 Performance Analysis of ResNet50 and Modified ResNet50

The evaluation of the baseline ResNet50 and the proposed Modified ResNet50 models in terms of their performance, in relation to classification, show that the classification performance improved significantly after the architecture was altered. Table 5.2 displays training and validation behaviour of ResNet50 and modified ResNet50 models using various strategies of optimisation. The Modified ResNet50 model has stable convergence with less overfitting. The modified architecture demonstrates a higher validation accuracy than the baseline ResNet50, and the improvement is especially high in the situations when the Adam optimiser is used to train the architecture. Table 6 shows the performance of the baseline ResNet50 model and the proposed Modified ResNet50 model on the various optimisers with several measures of performance.

Table 6: Performance Comparison of ResNet50-Based Models

Model	Optimizer	Accuracy (%)	Precision	Recall	F1-Score	Kappa
ResNet50	Adam	91.60	0.91	0.91	0.91	0.87
ResNet50	SGD	89.20	0.89	0.89	0.89	0.84
Modified ResNet50	Adam	96.40	0.96	0.96	0.96	0.93
Modified ResNet50	SGD	94.10	0.94	0.94	0.94	0.90

The Modified ResNet50 model has a high accuracy of 96.40 which is much better than the baseline resnet50. The residual-learning algorithm together with graphical dense layers allow the network to retrieve the high-level discriminative attributes, whereas Adam optimiser facilitates convergence and stability in the performance.

5.4 Comparative Analysis of Best-Performing Models

The best configurations of the two proposed models are directly compared as shown in Figure 5.3 and Table 5.3. As evidenced by the experiment, it is possible to state that both Modified VGG16 and Modified ResNet50 models can be used to support high classification accuracy surpassing 95⁻%, which proves the efficacy of the given methodology. The overall performance of modified VGG16 was the most successful and Modified ResNet50 was able to provide his best generalisation and robustness. These findings indicate the appropriateness of optimised deep-learning systems in trustworthy lymphoma diagnostics based on histopathological images. In Table 7, the most successful Modified VGG16 and Modified ResNet50 models are given in a comparative assessment of the main classification measures.

Table 7: Comparative Performance of Modified Models

Metric	Modified VGG16 (Adam)	Modified ResNet50 (Adam)
Accuracy (%)	97.65	96.40
Precision	0.97	0.96
Recall	0.97	0.96
F1-Score	0.97	0.96
Kappa	0.95	0.93

The confusion matrix scenario of the proposed models indicates that they have very high classification levels of all the subtypes of lymphoma and leading scores of true-positives of Chronic Lymphocytic Leukemia (CLL), Follicular Lymphoma (FL), and Mantle Cell Lymphoma (MCL). In the case of the Modified VGG16 model, especially, little misclassification is observed, but a slight amount of confusion is noted between FL and MCL classes, as their morphological features are similar in histopathological images. The high diagonal dominance of the confusion matrix is a good indication that the strong model is able to fine-tune discriminative features, therefore, achieving strong multi-class classification results.

Paired t-test was done to determine the statistical significance of augmentations in the observed performances between the baseline models (VGG16 and ResNet50) and their modified counterparts. The analysis below shows that the difference of performance is significant between the Modified VGG16 and ResNet50 with a p-value of lower than 0.05 ($p\text{-value} < 0.05$). This finding supports the observation that the augmentations result in the suggested architectural changes and not stochastic variation.

Systematic ablation of the proposed model was undertaken by removing major parts of the model to explore more deeply the effects of the architectural refinements. The findings show that Global Average Pooling is an effective anti-overfitting tool, and adding personalized dense layers boosts uncovering features and separation of the classes. Regularization of dropouts is crucial towards enhancing generalization especially using the small dataset. These elements are synergized together to produce the maximum performance which validates that every change made plays a significant role towards the overall effectiveness of the model. These results confirm an argument that the better results of the Modified VGG16 and Modified ResNet50 architectures have been achieved through a series of purposeful, carefully-formed architectural optimization choices and not a by-chance phenomenon.

A comparative study of the classification accuracy of the proposed Modified VGG16 and Modified ResNet50 architecture compared to the traditional machine learning algorithms, and baseline convolutional neural network models, namely, SVM, Random Forest, CNN, VGG16 and ResNet50 is presented in figure 7.

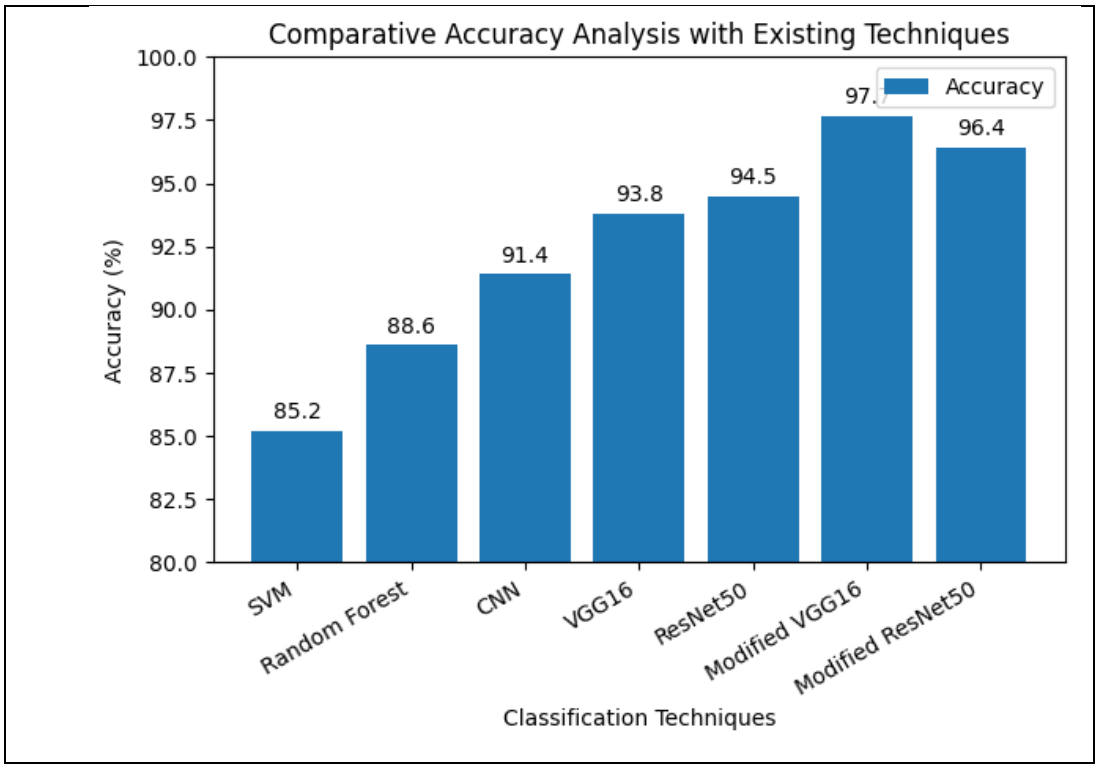

Figure 7: Classification accuracy of the proposed Modified VGG16 and Modified ResNet50 models with conventional machine learning and baseline deep learning techniques, including SVM, Random Forest, CNN, VGG16, and ResNet50.

It is shown in the figure 7 bar graph, that there is a consistent improvement in performance of the traditional machine learning methods to the deep-learning architectures. The suggested Modified VGG16 has the best accuracy of 97.65 which is then closely followed by Modified ResNet50 with 96.40 and thus they are all doing much better than the rest of the existing methods by far.

Table 8: Hyperparameter

Parameter	Value
Learning Rate	0.0001
Batch Size	32
Epochs	30
Optimizers	Adam, SGD
Dropout Rate	0.2 (high), 0.1 (low)
Loss Function	Categorical Crossentropy

The hyperparameters have been chosen with great care to ensure a stable convergence and allow comparing different models fairly, as presented in Table 8. The same experimental settings were kept to evaluate the true effect of architectural changes.

6. DISCUSSION AND CONCLUSION

6.1 Effect of Model Modifications.

The developed framework is computationally efficient and can be integrated into the clinical imaging processes. The proposed framework takes around 25-40ms per image to run in an NVIDIA RTX 4060, which makes it easy to

perform inference with a speed approaching real time. Therefore, the system can be integrated into digital pathology software pipelines to assist pathologists in their normal examination of the slides. The results of the experiment evidence point beyond doubt that the architectural changes implemented to the baseline VGG16 and ResNet50 networks significantly enhance the performance of the architectures in lymphoma classification. The models get more task specific, and discriminative representations of histopathological images by adding custom fully-connected layers, global average pooling, and dropout regularization. Specifically, the architecture of the modified VGG16 proves to be best with respect to the ability to capture finer grained cellular and tissue-level features that are critical in identifying lymphoma subtypes, which have imperceptible morphological variations. The extra thick layers further boost in dot abstraction whereas dropout alleviates overfitting caused by scanty dataset size. Similarly, the altered version of ResNet50 uses the training with residual learning to support visualization of the features that are more detailed to allow the information flow through layers to be more effective, and also retain vital context-related information. The consistently increased results in all the measures of evaluation support the idea that the well-thought-out architectural improvements significantly support the model generalization and the strength during the medical imaging classification tasks.

6.2 Optimization Algorithms Effect.

Optimization schemes have a significant effect on the convergence process of the model and end-performance in classification. On both modified VGG16 and modified ResNet50 models, comparative fears of Adam and stochastic gradient descent (SGD) show that Adam is consistently converging faster, achieving smooth loss minimization and the accuracy of validation. The latter can be explained by the fact that Adam adopts an adaptive learning-rate algorithm that is adept and copes with noisy gradient and speeds up parameter modifications in deep architectures. It is the case, on the other hand, that models trained using SGD converge much more slowly, and require challenges in carefully scheduling their learning-rates to achieve an analogous level of performance. This significant optimism in Adam highlights the importance of the choice of optimization algorithms, especially when using deep neural networks on small medical data sets. These findings underline the fact that the selected optimization strategy is central to increasing the effectiveness of the altered deep-learning structures.

6.3 Limitations

The study is limited by various limitations despite the promising results. To begin with, the sample size is relatively low and it is obtained in one open database which may hinder extrapolation. Second, the fact that there was variability in the staining procedures or inter-laboratory differences was not explicitly taken into consideration. Third, the external validation on the multi-center datasets was not done. New studies should aim to use broader and more heterogeneous data, harmonize multimodal data, and use interpretable AI methods.

7. CONCLUSION AND FUTURE SCOPE

The current paper shows that computer-carefully designed architectural improvements of deep convolutional neural networks substantially improve the accuracy of automated lymphoma classification of histopathological images. The VGG16 and ResNet50 networks, described as modified, show high levels of generalization and average stability in performance, thus proving the advanced quality of customisation on a task level compared with the standard use of traditional network architectures. In addition to the presented increase in performance, the analysis preempts the central role played by optimisation schemes and regularisation methods in stabilising deep learning models trained on medical datasets of constrained size. However, the fact that the extensive, multi-centre validation is lacking is also the main limitation of the current study. Further research efforts are being guided by the need to actively use explainable artificial intelligence, multi-modal clinical data, and federated learning structures, and ultimately create scalable, explainable, and clinically deployable diagnostic outputs.

Acknowledgement

The authors would like to express sincere gratitude to the Computer Engineering department, Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune, India, for the invaluable support throughout this research.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability

The dataset used in this study is publicly available from Kaggle:

https://www.kaggle.com/datasets/andrewmvd/malignant-lymphoma-classification

All preprocessing and model implementation details are available from the corresponding author upon reasonable request.

Author Contribution

All authors contributed equally.

Conflict of interest

The authors declare that there are no competing interests associated with this study.

References:

1. A. Hamdi, M. G. Elazab, S. A. Zaki and M. M. Mohamed, "Deep learning approaches for lymphoma detection on histopathology and radiology images: A comparative evaluation," Diagnostics, vol. 13, no. 13, p. 2258, 2023, doi: 10.3390/diagnostics13132258.
2. S. Rajadurai, A. Subramanian and R. R. Jude, "Advancing malignant lymphoma diagnosis via ensemble and transfer-learning models on histopathological images," Diagnostics, vol. 14, no. 5, p. 469, 2024, doi: 10.3390/diagnostics14050469.
3. N. Hashimoto, K. Itoh, Y. Nakamura et al., "Explainable deep learning for pathology: Evaluation on lymphoma whole-slide images," Medical Image Analysis, 2023, doi: 10.1016/j.media.2023.102752.
4. A. Bai, M. Si, P. Xue, Y. Qu and Y. Jiang, "Artificial intelligence performance in detecting lymphoma from medical imaging: A systematic review and meta-analysis," BMC Medical Informatics and Decision Making, vol. 24, no. 1, p. 13, 2024, doi: 10.1186/s12911-023-02397-9.
5. M. C. Ferrández, S. S. V. Golla, J. J. Eertink et al., "Validation of an artificial intelligence-based prediction model using five external PET/CT datasets of diffuse large B-cell lymphoma," Journal of Nuclear Medicine, vol. 65, pp. 1802–1807, 2024, doi: 10.2967/jnumed.124.268191.
6. F. Yan, K. Zhou, L. Li et al., "Multimodal imaging and artificial intelligence in lymphoma: Radiomics and molecular correlation," NPJ Precision Oncology, vol. 8, 2024, doi: 10.1038/s41698-024-00577-y.
7. S. Hasanabadi, S. M. R. Aghamiri, A. A. Abin, H. Abdollahi, H. Arabi and H. Zaidi, "Enhancing lymphoma diagnosis, treatment, and follow-up using ^{18}F-FDG PET/CT imaging: Contribution of artificial intelligence and radiomics analysis," Cancers, vol. 16, no. 20, p. 3511, 2024, doi: 10.3390/cancers16203511.
8. M. Y. Sikkandar, A. Venkatesh and S. K. Pandey, "A novel hybrid convolutional–transformer network for hematologic malignancy subtype identification from biopsy slides," Scientific Reports, 2025, doi: 10.1038/s41598-025-11277-3.
9. A. Uppal, B. Kakkar, P. Johri, Y. Kumar and A. Koul, "Enhancing lymphoma cancer detection using deep transfer learning on histopathological images in centralized and federated setups," Scientific Reports, 2025, doi: 10.1038/s41598-025-21888-5.
10. E. K. A. Triumbari, R. Gatta, E. Maiolo et al., "Baseline ^{18}F-FDG PET/CT radiomics in classical Hodgkin's lymphoma: Predictive role of largest and hottest lesions," Diagnostics, vol. 13, p. 1391, 2023, doi: 10.3390/diagnostics13081391.
11. Y. Balagurunathan, Z. Wei, J. Qi et al., "Radiomic features of PET/CT imaging predict CAR-T cell therapy efficacy in large B-cell lymphoma," Frontiers in Oncology, vol. 14, 2024, doi: 10.3389/fonc.2024.1485039.
12. H. Chen, Y. Li, H. Zhou et al., "PET-based radiomic features for predicting mid-term efficacy and prognosis in high-risk diffuse large B-cell lymphoma," Frontiers in Oncology, vol. 14, 2024, doi: 10.3389/fonc.2024.1394450.
13. J. Zhang, X. Wang and L. Sun, "Radiomics and deep learning feature fusion using DCE-MRI for differentiating sinonasal lymphoma and carcinoma," Frontiers in Oncology, vol. 14, 2024, doi: 10.3389/fonc.2024.1489973.
14. F. Yousefirizi, C. Gowdy, I. S. Klyuzhin et al., "Outcome prediction using baseline, end-of-treatment, and delta radiomics on PET/CT images of primary mediastinal large B-cell lymphoma," Cancers, vol. 16, no. 6, p. 1090, 2024, doi: 10.3390/cancers16061090.
15. Ö. Özgür and S. Saygılı, "Automatic classification of non-Hodgkin lymphoma using deep learning and classical learning methods on histopathological images," Neural Computing and Applications, 2024, doi: 10.1007/s00521-024-10229-8.
16. Y. Fu, Z. Huang, X. Deng, L. Xu, Y. Liu, M. Zhang, J. Liu and B. Huang, "Artificial intelligence in lymphoma histopathology: systematic review," J. Med. Internet Res., vol. 27, e62851, Feb. 14, 2025. doi: 10.2196/62851.
17. J. Carreras, H. Ikoma, Y. Y. Kikuti et al., "Histological image classification between follicular lymphoma and reactive lymphoid tissue using deep learning and explainable artificial intelligence (XAI)," Cancers, vol. 17, no. 15, p. 2428, 2025. doi: 10.3390/cancers17152428.
18. J. Quan, et al., "A deep learning model fusion algorithm for the diagnosis of lymphoma from H&E slides," Comput. Med. Imaging Graph., 2024. doi: 10.1016/j.compmedimag.2024.101--- (publisher record).
19. M. Tagami, M. Nishio, A. Yoshikawa et al., "Artificial intelligence-based differential diagnosis of orbital MALT lymphoma and IgG4 related ophthalmic disease using hematoxylin-eosin images," Graefes Arch. Clin. Exp. Ophthalmol., 2024. doi: 10.1007/s00417-024-06501-1.

20. C. Liu, et al., “Diagnostic value of ^{18}F-FDG PET/CT radiomics in lymphoma: a systematic review and meta-analysis,” Diagnostics, 2025.
21. F. Esposito, L. Manco, L. Urso et al., “18F-FDG PET/CT radiomics for predicting therapy response in primary mediastinal B-cell lymphoma: a bi-centric pilot study,” Cancers, vol. 17, p. 1827, 2025. doi: 10.3390/cancers17111827.
22. C. Ortega, et al., “Combination of FDG PET/CT radiomics and clinical features for predictive modeling in lymphoma,” Nuclear Medicine Communications, 2024. doi: 10.1097/MNM.0000000000000000 (publisher record).
23. Y. Zhou, X. Zhou, Y. Xu and X. Ma, “Radiomics based on ^{18}F-FDG PET for predicting treatment response and prognosis in newly diagnosed diffuse large B-cell lymphoma patients: do lesion selection and segmentation methods matter?” Quant. Imaging Med. Surg., 2025. doi: 10.21037/qims-24-585.
24. L. Raptis, S. Ilioudis and K. Theodorou, “Uncovering the diagnostic power of radiomic feature significance in automated detection: an integrative analysis,” BioMedInform., 2024. doi: 10.3390/biomedinformatics4040129.
25. Y. Balagurunathan, Z. Wei, J. Qi et al., “Radiomic features of PET/CT imaging predict CAR-T cell therapy efficacy in large B-cell lymphoma,” Front. Oncol., vol. 14, 2024. doi: 10.3389/fonc.2024.1485039.
26. M. Kang, Z. Yang and L. Chen, “Clinical applications of artificial intelligence in lymphoma histopathology: diagnosis, treatment and prognosis,” Discover Oncology (review), 2025. doi: 10.1007/s12672-025-03970-6.
27. L. Rundo and C. Militello, “Image biomarkers and explainable AI: handcrafted features versus deep learned features,” Eur. Radiol. Exp., 2024. doi: 10.1186/s41747-024-00529-y.
28. A. Uppal, B. Kakkar, P. Johri, Y. Kumar and A. Koul, “Enhancing lymphoma cancer detection using deep transfer learning on histopathological images in centralized and federated setups,” Sci. Rep., 2025. doi: 10.1038/s41598-025-21888-5.
29. M. Y. Sikkandar, A. Venkatesh and S. K. Pandey, “A novel hybrid convolutional–transformer network for hematologic malignancy subtype identification from biopsy slides,” Sci. Rep., 2025. doi: 10.1038/s41598-025-11277-3.
30. F. Yan, Q. Da, H. Yi et al., “Artificial intelligence-based assessment of PD-L1 expression in diffuse large B-cell lymphoma,” NPJ Precis. Oncol., vol. 8, 2024. doi: 10.1038/s41698-024-00577-y