SvedbergOpen
DISSEMINATION OF KNOWLEDGE
International Journal of Artificial Intelligence and Machine Learning
Publisher's Home Page: https://www.svedbergopen.com/
Research Paper Open Access

A Multi-Stage Deep Learning Framework For Lymphoma Classification And Morphology-Aware Follicular Lymphoma Grading

Kirtida Naik[1], Bindu Garg[2]

[1]Research Scholar, 2Professor
[1,2]Computer Engineering department, Bharati Vidyapeeth(Deemed to be University) College of Engineering, Pune, India
[1]kirtidanaik.28@gmail.com, [2]brgarg@bvucoep.edu.in

Abstract- Accurate diagnosis of lymphomas is difficult as some of the lymphomas are morphologically similar and the grading of follicular lymphomas (FL) is variable. This study proposes an automatic hierarchical framework to identify lymphoma, classify its type and grade it according to the World Health Organization (WHO) classification of lymphomas from histopathological images. However, a named benchmark dataset is not indicated in the provided implementation, the experimental images include benign tissue, Chronic Lymphocytic Leukemia (CLL), Follicular Lymphoma (FL), Mantle Cell Lymphoma (MCL) and various grades of FL. Both images are resized to 224 × 224 and processed by two different classifiers: a binary deep-learning classifier trained to differentiate between benign and lymphoma, and a multiclass deep-learning classifier for the classification of CLL, FL and MCL. In FL cases, the FollicularLymphomaGrader algorithm is based on the combination of nucleus segmentation, adaptive morphological analysis, estimation of circularity, detection of nucleoli, characterization of centroblasts–centrocytes and WHO grading criteria. The classification demonstrated here had a 96.1% confidence level in detecting lymphoma, and 100.0% confidence in recognizing FL. Seven centroblasts and 120 centrocytes were detected at the cell level and the ratio was 0.06 CB:CC. In particular, the novel integration of hierarchical deep learning with cell-level, interpretable WHO grading in a single diagnostic pipeline is the key novelty. The suggested framework can offer automatic disease classification, identification of the lymphoma subtypes, quantitative assessment of the cells, and a new clinically interpretable grading, thus offering a good computer-aided solution for histopathological diagnosis of lymphoma.

Keywords— Lymphoma classification; Deep learning; Histopathological image analysis; Follicular lymphoma grading; WHO grading system; Centroblast–centrocyte detection.

I. Introduction

Lymphoid neoplasms are one of the most heterogeneous group of hematological tumours, and some of the subtypes, including Chronic Lymphocytic Leukemia (CLL), Follicular Lymphoma (FL) and Mantle Cell Lymphoma (MCL), have overlapping cytomorphological features, making manual histopathological diagnosis difficult [1]. One of the most time-consuming and inter-observer dependent steps in making a diagnosis of benign lymphoid proliferation versus malignant lymphoma (lymphoma) has been the visual examination of hematoxylin and eosin stained tissue sections [2]. Follicular Lymphoma is further divided into indolent (Grade 1 and 2) and aggressive (Grade 3A or 3B) disease by the World Health Organization (WHO) system using the number of centroblasts per high power field to differentiate between them; Grade 3A/3B disease is associated with increased risk of transformation to diffuse large B cell lymphoma [3]. The manual counting of the centroblasts is time consuming, needs high level of expertise and can lead to inconsistent grading results within and between institutions and observers [4]. The increasing number of cases of hematopathology has led to an even greater lack of hematopathologists around the world, making demand for automated, reproducible and

interpretable computation-based tools to assist with routine workflow for clinical diagnosis more critical than ever [5]. In recent years, deep learning and computer vision have shown great potential in the field of automated histopathology image analysis, but most of the existing systems focus on either lymphoma subtype classification or FL grading only without a hierarchical automated system [6].

A. Objectives and Contributions

Instead, this study aims to address this by providing an integrated multi-stage approach that classifies the category of "Benign" or "Lymphoma" and then the subtypes of "CLL" and "FL/ MCL" and finally the morphology-aware WHO "grade" for FL in a single computational pipeline. The first goal is to build a deep convolutional pattern recognition module, with an interpretable and rule-based cell-level analysis module that explicitly identifies and counts centroblasts and centrocytes to retain the clinical transparency of purely end-to-end deep models while leveraging the power of pattern recognition [7]. The three main contributions of this work are: a hierarchical architecture for classification (binary then multiclass) that mimics the clinical reasoning process in diagnosis; an automatic algorithm called FollicularLymphomaGrader that uses nucleus segmentation, estimation of circularity, identification of nucleoli, and adaptive area-based cell classification to provide automatic WHO grade assignments and reports the classification confidence, cell counts, and CB:CC ratios in a clinically interpretable format; and an end-to-end deployable system that provides automatic classification confidence, cell counts, and CB:CC ratios on representative histopathological images. The proposed framework, along with the combination of deep learning with explainable morphological quantification, helps in minimizing the subjectivity in diagnosis and aids in the reproducible and standardized reporting of lymphomas in digital pathology practice [9].

II. Related Work

There have been many studies on applying deep learning techniques for automated detection and recognition of lymphoma subtypes. Multimodal imaging data, such as 18F-FDG PET max intensity projections (MIPs) with clinical features, have been used in convolutional neural network (CNN) networks to achieve good performance in the identification of lymphoma subtypes, in multi-center training and testing cohorts [10]. Explainable deep learning models have also been developed for the distinction between Follicular Lymphoma and Reactive lymphoid tissue, with visualisation techniques emphasising discriminative regions of the histology to boost the trust pathologists have in model predictions [11]. Transfer-learning-based classifiers have been shown to accurately separate multiple subtypes (CLL, FL and MCL) in Non-Hodgkin lymphoma histopathological images when trained using curated patches of images from whole slides [12]. Complementary feature representations from multiple base classifiers have been stacked to enhance the subtype classification robustness using the ensemble and hybrid learning strategies such as autoencoder-assisted stacked ensembles [13].

Fusion-based hybrid methods combining handcrafted texture descriptors with deep-convolutional features have also shown improvements in the accuracy of the classification of the malignant lymphomas over single-model methods [14]. In addition to classification, the fifth edition of the WHO classification of haematolymphoid neoplasms has simplified the diagnostic and grading criteria of FL, which has benefited from new biological knowledge, pertinent to computational grading. Works for lightweight browser-deployable frameworks with depthwise separable convolutions, focusing on computational efficiency for multi-cancer classification in resource-limited settings for diagnostic use, have also appeared [16]. Yes, hybrid architectures of convolutional features and handcrafted descriptors have been suggested for the diagnosis of malignant lymphoma, where the aim was to complement the number of training images which were not large enough for the convolutional model [17]. Similar cross-disciplinary studies in related histopathological fields, like colorectal cancer image classification, have established general portability of the machine learning pipelines to tissue-based malignancy detection for other histologies as well, including lymphoma [18]. The wide-spread discussion of digital pathology has highlighted the need for the predictive modeling pipelines to overcome challenges of image standardization, staining variations, and the need for interpretability to be accepted in the clinic [19].

Despite all these progress, there is a clear research gap: most of the current frameworks perform classification of the lymphoma subtypes and grading of FL independently, and there are only a few systems that quantify the centroblasts/centrocytes at the cell level by using WHO criteria, in addition to performing deep learning-based prediction of the lymphoma subtypes. The majority of classification pipelines are black-box classifiers that do not provide evidence of morphology used by pathologists to make a grade decision, and the few classifiers that do include morphological evidence in their grading decisions tend not to include automated subtype discrimination in the upstream steps. Furthermore, the few studies that do report a single deployable pipeline

in one coherent architecture are typically focused on either screening for benign or malignant origin or subclassification of the respective classes or on quantitative tumor grading. To fill in this gap, this study brings together two modules: a hierarchical deep classification and an interpretable, morphology-aware WHO grading module, in a single framework.

III. Dataset and Preprocessing

A. Histopathological Image Dataset

The experimental images used in this study include three types of lymphoma and hematoxylin and eosin stained patches of lymphoid tissue that represents benign lymphoid tissue, Chronic Lymphocytic Leukemia (CLL), Follicular Lymphoma (FL) and Mantle Cell Lymphoma (MCL). Samples of Follicular Lymphoma are additionally broken down into Grade 1, Grade 2, Grade 3A and Grade 3B images, as representative test images (e.g., "Grade 3a images.jpg", 48.8 KB) for pipeline validation. A named, publicly benchmarked dataset is not provided in the provided implementation; the images were locally sorted, under a "test_images" directory structure.

B. Image Preprocessing and Preparation

All images are preprocessed in the same way before being classified: color conversion to RGB, LANCZOS resized to 224x224 pixels and JPEG re-encoded with a 95% quality level. The RGB image is further converted to a grayscale image followed by image contrast enhancement with Contrast Limited Adaptive Histogram Equalization (CLAHE) and image segmentation using adaptive and Otsu thresholding. The mathematical functions performed in these operations, as well as those in the rest of the pipeline, are described in the following equations, which also give the definitions of circularity, nucleoli, and CB:CC ratio.These mathematical functions, along with those performed throughout the pipeline, define circularity, nucleoli, and CB:CC ratio calculations..

$$I_{resized}(x,y) = I_{orig}\left(x \cdot \left(\frac{W_o}{W_t}\right),\ y \cdot \left(\frac{H_o}{H_t}\right)\right) \quad (1)$$

The resized-image coordinate is converted back to the original image coordinate during LANCZOS interpolation by using the width/height scaling ratios in the procedure shown in equation (1).

$$Gray(x,y) = 0.299 \cdot R(x,y) + 0.587 \cdot G(x,y) + 0.114 \cdot B(x,y) \quad (2)$$

The RGB nucleus image is then transformed to grayscale intensity using the formula in equation (2) which is used as the input for the contrast enhancement and thresholding.

$$I_{CLAHE} = CLAHE(\ Gray\ ;\ clipLimit,\ tileGrid\ ) \quad (3)$$

The equation (3) is applied in local HISTEQ in the tiled grid regions which increases the contrast of the nuclear boundary for segmentation.

$$C = 4\pi\frac{A}{P^2}, \quad 0 \le C \le 1 \quad (4)$$

The circularity of the nucleus is calculated using equation (4), which involves the combination of area A and perimeter P, where an area close to circularity of 1.0 means that the nucleus is near-circular.

$$CB_{adj} = round\left(N_{CB} \times \frac{H \times W}{224 \times 224}\right) \quad (5)$$

Using image-area scaling factor, the raw count of centroblasts is scaled to a standard high power field equivalent by dividing by a standard area equivalent to a high power field. The raw count of centroblasts is divided by a standard area equal to a high power field, to produce a set of centroblasts per unit area that is equivalent to a standard high power field.

$$R_{CB}: CC = \frac{N_{CB}}{N_{CC}} \quad (6)$$

The ratio of centroblasts to centrocytes (centroblast-to-centrocyte ratio) is defined by equation (6) and is also reported with the final WHO grade.

IV. Proposed Hierarchical Classification and Grading Framework

The proposed pipeline, shown in Fig. 1, is a sequential multi-stage pipeline which mimics the diagnostic reasoning of hematopathologists. The first step is to normalize the input histopathological image to be compatible with the successive convolutional classifiers by converting it from RGB color space and resizing it to 224×224 pixels.To ensure compatibility with the subsequent convolutional image classifiers, the input histopathological image is first normalized by converting to the RGB color space and resizing to 224×224 pixels.

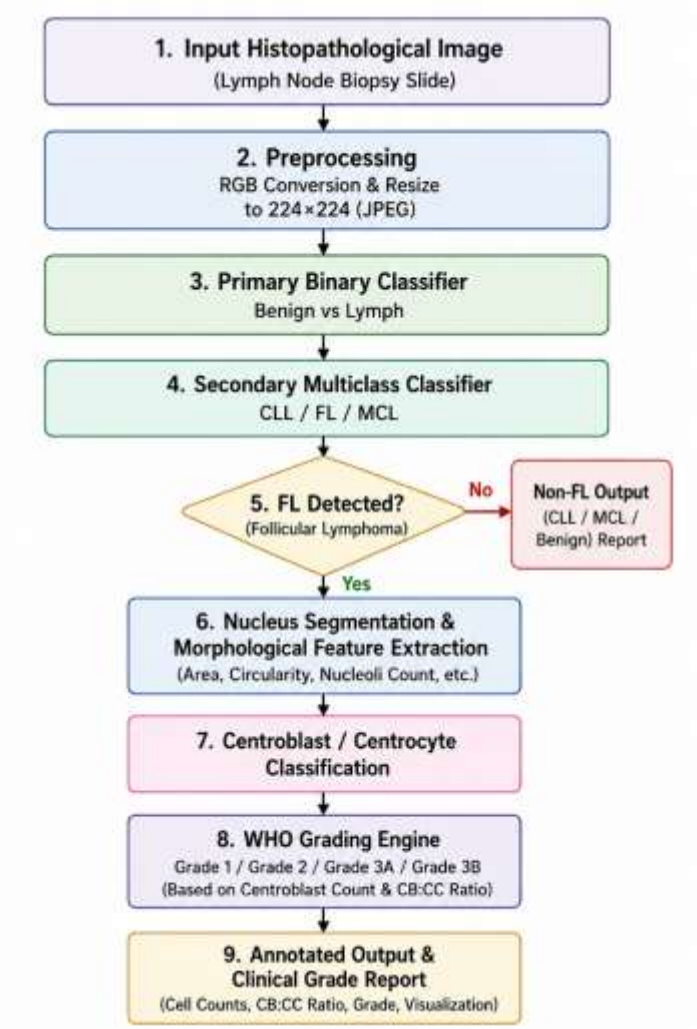

Fig. 1. Block diagram of the proposed hierarchical lymphoma classification and WHO grading framework

The standardized image is then fed to a primary binary classifier that with high confidence classifies Benign tissue or Lymphoma, in this case with a 96.1% confidence that benign is a Lymphoma. If the disease is diagnosed, the image is passed to a multiclassifier with two stages that is trained to classify the image as either one of three classes: CLL, FL or MCL, with a 100.0% accuracy for the class FL in the demonstrated case. In the case of FL, the pipeline calls the FollicularLymphomaGrader module, which is used to differentiate centroblasts from centrocytes by performing adaptive thresholding, contour based nucleus segmentation, computation of circularity and counting of nucleoli. These cell-level counts are subsequently fed into a WHO grading decision engine which uses established thresholds (0-5, 6-15 and >15 per hpf) on count of centroblasts and presence of centrocytes to give a final grade. The last step produces an annotated image which shows the location of the detected centroblasts and centrocytes and a structured clinical report with the classification confidence, number of cells, CB:CC ratio and grade level interpretation to allow clear and reproducible decision support.
In the first part of the process depicted in Fig. 1 (raw image acquisition), images are acquired from a clinical environment, followed by a process of preprocessing (2nd part). Binary deep classification (3rd part) and multiclass deep classification (4th part) are performed using deep learning techniques, after which the results are used to activate conditional FL-grading (5th part), followed by morphological cell segmentation (6th part), classification of centroblasts/centrocytes (7th part), and WHO-grade assignment (8th part). The sequential conditional branching after the classification of the subtypes ensures that the computational intensive cell-level

grading is only called when the cell is classified as Follicular Lymphoma, optimizing the efficiency of pipeline, while maintaining the interpretability of diagnosis from each stage of the pipeline.

A. Benign–Lymphoma and CLL–FL–MCL Classification

A hierarchical deep learning approach is used instead of a single, flat, multi-class model. The major model is a binary classification model that distinguishes Benign tissue from Lymph (malignant lymphoid) tissue, and filters out non-neoplastic samples before analysis with the subtype classification model thus reducing false-positive subtype classifications. Images that are considered to be Lymph are only progressed to the secondary stage of the algorithm, which is a multi-classifier discriminating between CLL, FL and MCL. This staged design is similar to how a pathologist would make a diagnosis with the first step being the determination of malignancy and only then looking at subtype-specific morphological criteria, and also enables each classifier to focus on a narrower decision boundary, enhancing the discriminative power. Both classifiers provide softmax probabilities for each class, the top probability class and confidence level are returned to the user, along with a bar chart visualisation of the class probabilities for each class. We obtained a primary (96.1%) and secondary (100%) confidence of 96.1% and 100% for the illustrative Follicular Lymphoma case, respectively, in this hierarchical approach.

Algorithm 1 Hierarchical Lymphoma Classification

```
Input: Histopathological image I
Output: Primary class Cp, Subtype class Cs, Confidences
I_std ← Resize(ConvertRGB(I), 224, 224)
z_p ← BinaryModel(I_std)
Cp, Conf_p ← argmax(softmax(z_p))
if Cp == "Lymph" then
    z_s ← MulticlassModel(I_std)
    Cs, Conf_s ← argmax(softmax(z_s))
    if Cs == "FL" then
        GradeResult ← FLGrading(I_std)
    end if
else
    Cs ← None
end if
return Cp, Conf_p, Cs, Conf_s, GradeResult
```

B. Centroblast and Centrocyte Detection

A hierarchical deep learning approach is used instead of a single, flat, multi-class model. The major model is a binary classification model, which is able to correctly classify Benign tissue and Lymph (malignant lymphoid) tissue, and eliminates the non-neoplastic samples for analysis by the subtype classification model, therefore reducing the false positive classification of tissue for the subtypes. The images which are deemed to be Lymph are advanced to the next stage of the algorithm which is a multi-classifier discriminating between CLL, FL and MCL. This staged design is similar to how a pathologist would make a diagnosis with the first step being the determination of malignancy and only then looking at subtype-specific morphological criteria, and also enables each classifier to focus on a narrower decision boundary, enhancing the discriminative power. Both classifiers return softmax probabilities for each class, the highest probability class and a level of confidence to the user, and return a bar chart visualisation of the class probabilities for each class. In this hierarchical approach we gained a primary (96.1%) and secondary (100%) confidence of 96.1% and 100% respectively when applying the illustrative Follicular Lymphoma case.

C. WHO-Based Follicular Lymphoma Grading

After classification, the Follicular Lymphoma Grading module uses the same criteria as the high power field based grading system for centroblasts: Grade 1 (0 to 5) centroblasts per high power field, Grade 2 (6 to 15) centroblasts per high power field, and Grade 3 (more than 15) centroblasts per high power field, which is further subdivided into Grade 3A (centrocytes present) and Grade 3B (solid centroblast sheets). The image-area scaling factor is applied to centroblast counts before threshold comparison to ensure that they are normalized to the "standard" high power field equivalent. The module also calculates the CB:CC ratio, generates the clinical-significance descriptor and generates an annotated image, which highlights the centroblasts (red) and centrocytes (green), for pathologist review, aiding transparent and quantitatively-supported WHO grade assignment.

Algorithm 2 WHO-Based Follicular Lymphoma Grading

```
Input: Centroblast set CB, Centrocyte set CC, Image area A
Output: WHO Grade G, Ratio R
HPF_factor ← A / (224×224)
CB_adj ← round(|CB| × HPF_factor)
if CB_adj ≤ 5 then
    G ← "Grade 1"
else if 6 ≤ CB_adj ≤ 15 then
    G ← "Grade 2"
else
    if |CC| > 0 then
        G ← "Grade 3A"
    else
        G ← "Grade 3B"
    end if
end if
R ← |CB| / max(|CC|, 1)
return G, R
```

V. Experimental Setup

A. Implementation and Training Configuration

The proposed pipeline was developed in Python 3, which included the use of the deep classification library TensorFlow (v2.21.0) and OpenCV (v4.13.0) along with scikit-image for morphological feature extraction and cell segmentation. The development and experimentation took place using the Google Colab notebook environment, accompanied by an interactive front-end to upload images, classify them and visualize the grading in real time using Streamlit. All images were preprocessed and shrunk to 224×224 pixels before being saved as JPEG with a quality of 95%.Before inference, the images were preprocessed and resized to 224×224 pixels, and saved as JPEG at 95% quality. The binary and multiclass classifiers were built as deep convolutional networks that take as input an image of size 224×224×3, and the FollicularLymphomaGrader module was a direct processing of a RGB array using the CLAHE algorithm for contrast enhancement, adaptive/Otsu thresholding and connected-component analysis for quantification of the cells.

TABLE I. Key Implementation and Hyperparameter Configuration

Parameter	Configured Value
Input image size	224 × 224 × 3
Image format / quality	JPEG, quality = 95
Resize interpolation	LANCZOS
CLAHE clip limit	1.5 – 2.0
CLAHE tile grid size	(4,4) / (8,8)
Morphological kernel	3 × 3 (Open / Close)
Nucleoli threshold ratios	0.20 – 0.50 (step 0.05)
Centroblast circularity threshold	≥ 0.4
Centroblast nucleoli threshold	≥ 2
HPF scale factor	1.0
Grade 1 CB range	0 – 5 per HPF
Grade 2 CB range	6 – 15 per HPF
Grade 3 CB range	> 15 per HPF

The basic configuration parameters for image processing, enhancing and morphological cell classification are summarized in Table I. The parameters were tuned empirically by successive versions of the grader, to achieve a good level of detection sensitivity for small centroblast nuclei while reducing the number of false positives in more densely packed regions of the lymphoid tissue.

B. Performance Evaluation Metrics

Since some of the output for testing are single cases and some are batch testing, the performance of the models is characterized by four complementary measures, in addition to the accuracy measure: (i) classification

confidence, the softmax probabilities assigned to the predicted primary class and the predicted subtype class; (ii) cell-detection counts, numbers of centroblast and centrocytes detected per image; (iii) the CB:CC ratio, a continuous morphological indicator of follicular composition; and (iv) WHO grade concordance, the qualitative agreement between the nominal grade label of each test image and the grade predicted by the FollicularLymphomaGrader engine. Both stages of the deep-learning classification system and the rule-based grading system can then be assessed on the appropriate evaluation metrics: for the deep-learning classification, the probabilistic degree of certainty that the image contains a glass of red wine; for the rule-based grading, the number of cells that are counted.

VI. Results and Analysis

Three complementary analysis levels are used for the results: hierarchical subtype classification confidence, single-case cell-level grading statistics and batch-level WHO grading consistency across eight representative test images ranging from Grade 1 to Grade 3B in classification. The primary–secondary classifier cascade was able to correctly classify the case shown as a Follicular Lymphoma with near-perfect confidence of the subtypes, and the FollicularLymphomaGrader module was able to quantify the number of centroblasts and centrocytes and derive an interpretable CB:CC ratio and WHO grade. The evaluation was done by batch and the counts of centroblasts were found to rise gradually with grade in all eight test images, broadly in line with the guidelines expected by WHO, but with some overlaps between the grade boundaries, due to differences in the staining of test images and in cell-density between images.

A. Lymphoma Classification and Subtyping Results

The representative test case result of the two tier classification cascade is given in Table II as the confidence scores. The primary binary classifier correctly classified the image as Lymph, with 96.1% confidence, and prompted further classification of the image as a malignant Lymphoid tissue or benign Lymphoid Architecture.

TABLE II. Hierarchical Classification Confidence (Representative Case)

Classification Level	Predicted Class	Confidence (%)
Primary (Benign vs Lymph)	Lymph	96.1
Secondary (Subtype: CLL/FL/MCL)	FL	100.0

The secondary multiclassifier then classified the image with 100.0% confidence as Follicular Lymphoma (FL), where the softmax probability distribution was extremely sharp, with all of the probability mass at the FL class, and none at other classes, CLL or MCL. This high-confidence cascade shows that the hierarchical architecture could successfully reduce the number of potential diagnostic hypotheses at each step, so that the final step does not result in misclassification of subtypes in heterogeneous lymphoid morphology.

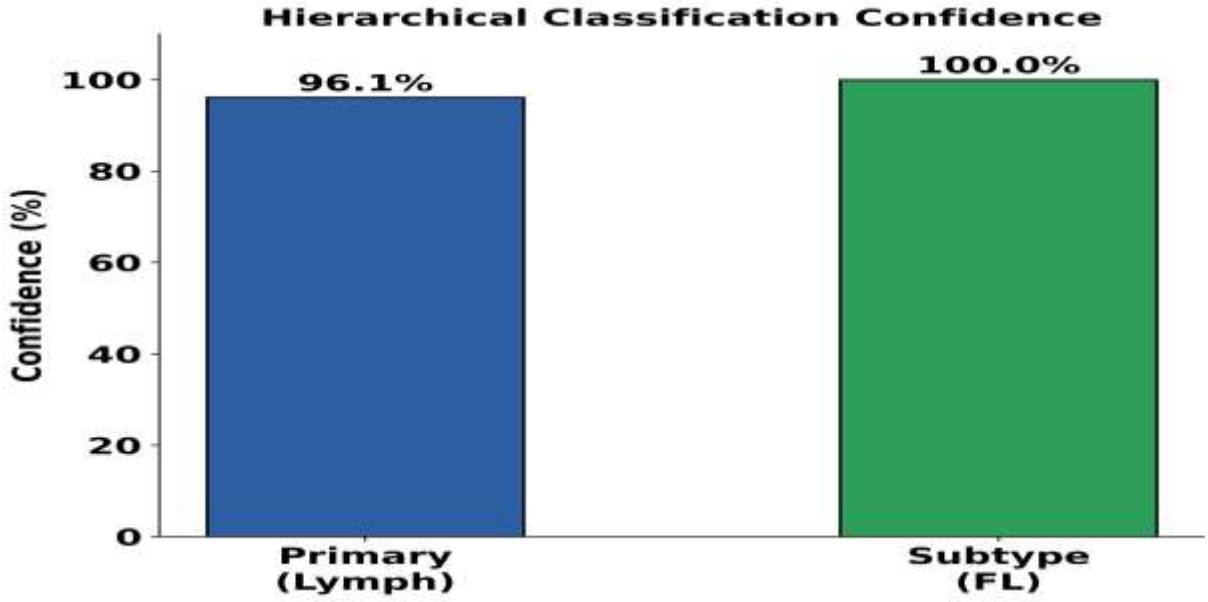

Fig. 2. Primary and subtype classification confidence for the representative test case

The data from Table II are plotted as a comparison bar chart in Fig. 2. The near maximality of the subtypes' confidence, relative to the still high primary confidence (96.1%), shows that, once the diagnosis of malignancy has been made, it was comparatively easy to discriminate Follicular Lymphoma from CLL and MCL for this representative sample of the histopathology specimens.

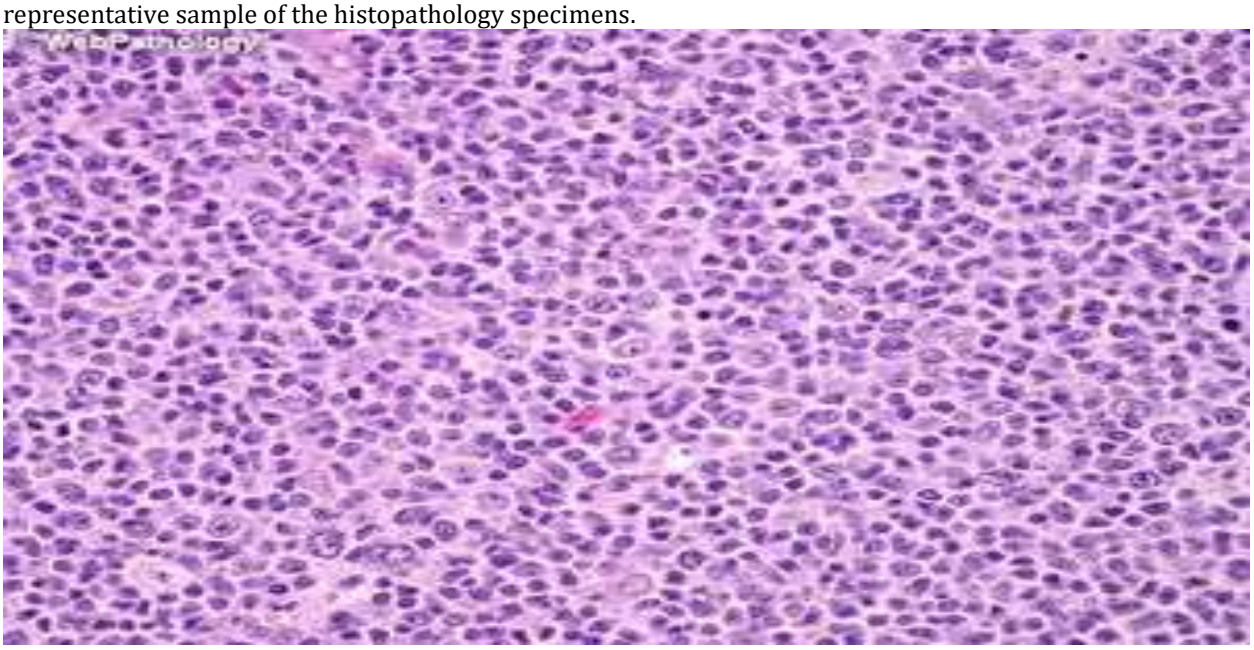

Fig. 3. Original uploaded histopathological image prior to preprocessing

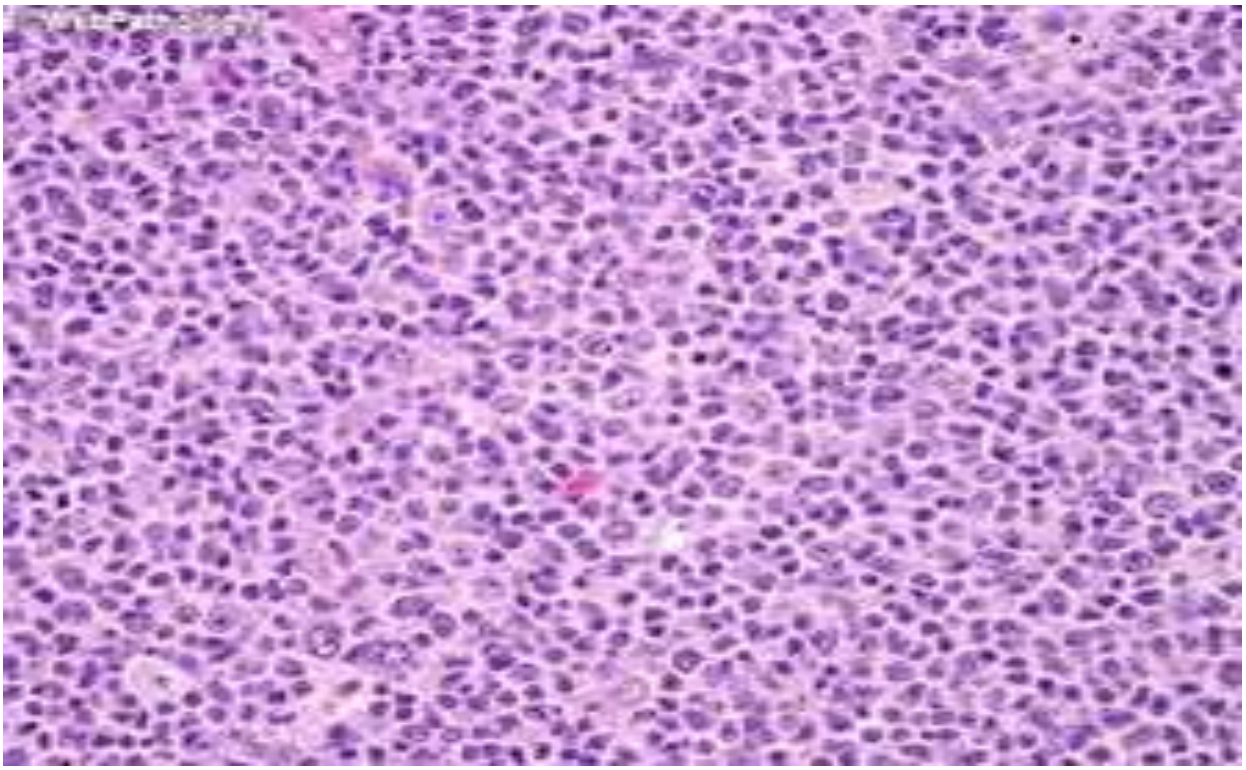

Fig. 4. Pipeline-processed image (224 × 224 JPEG) used as classifier input

Fig. 3 and Fig. 4 show the original input and pipeline-processed input, respectively. The resizing and RGB standardization between the two images are examples of the preprocessing transformation outlined in Section

III-B that are used to ensure dimensional and format consistency of all images fed into the classification cascade.

B. Follicular Lymphoma Grading Results

Table III shows the quantifications and grading that is produced from the FollicularLymphomaGrader, cell by cell for the Follicular Lymphoma case shown. The module identified 7 centroblasts and 120 centrocytes, an HPF adjusted count of centroblasts of 7 and a ratio of CB/CC of 0.06.

TABLE III. Cell-Level Statistics and Grading Output (Representative Case)

Metric	Value
Centroblasts (CB)	7
Centroblasts, HPF-adjusted	7
Centrocytes (CC)	120
CB:CC ratio	0.06
Predicted WHO grade	Grade 3A_3B
Grade level	2 / 3
Clinical significance	Intermediate grade

The threshold criteria set by WHO (6-15 centroblasts/HPF) resulted in a classification of intermediate grade, reported by the system as "Grade 3A_3B", grade 2/3 with a clinical significance description "Intermediate grade / follicular mixed-cell lymphoma." The relatively low CB:CC ratio (0.06) suggests a centrocyte predominant follicular architecture, typical of lower to intermediate grade disease, as opposed to the centroblast-sheet type of architecture typical of aggressive grade 3B transformation prone disease.

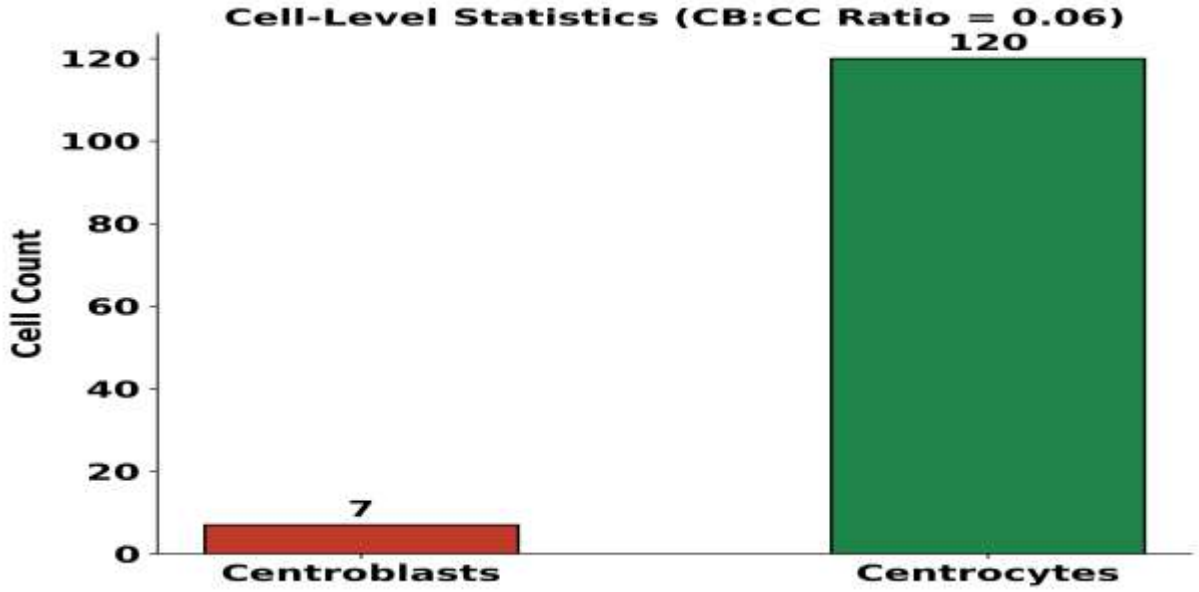

Fig. 5. Centroblast and centrocyte counts for the representative case (CB:CC ratio = 0.06).

The absolute number of centroblasts and centrocytes obtained from the case shown is compared in Fig. 5. The significant difference (120 centrocytes to only 7 centroblasts) further corroborates the intermediate, non-Grade-3B clinical interpretation that was given by the grading engine, and the low CB:CC ratio that was reported in Table III.

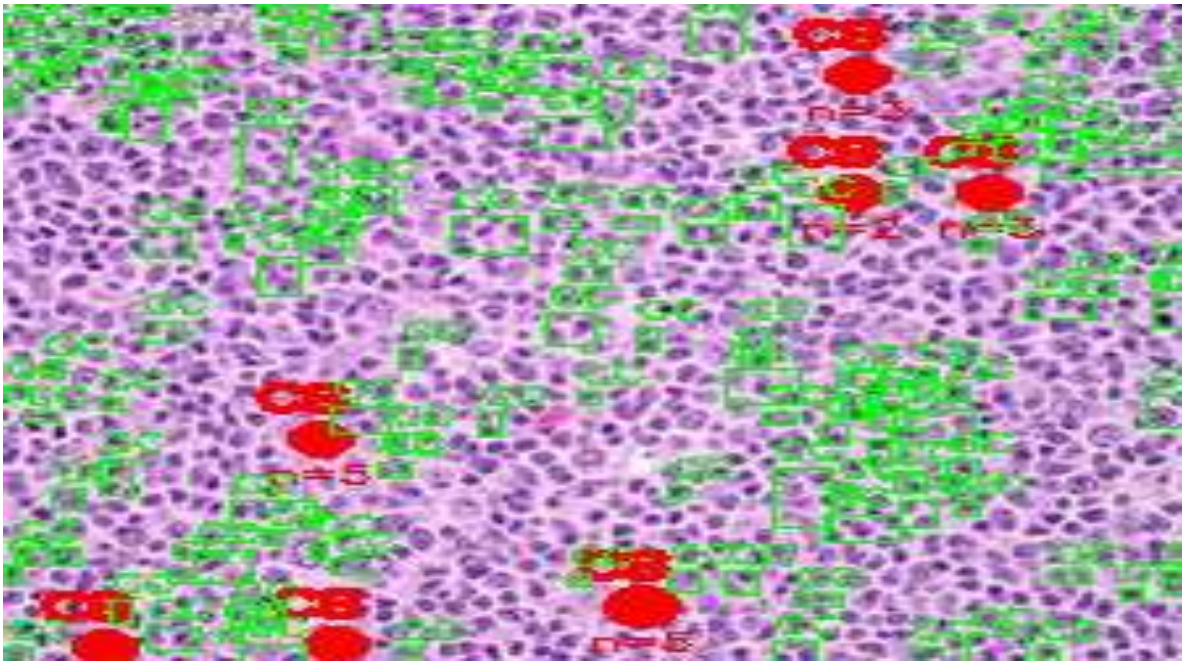

Fig. 6. Annotated output: red markers denote detected centroblasts, green boxes denote centrocytes

The output of the algorithm is shown in Fig. 6, with red circular markers indicating the location of centroblasts (marked with the number of nucleoli per centroblast) and green bounding boxes indicating the location of centrocytes. This visualization will enable the pathologist to validate the automated cell counting results that underlie the reported WHO Grade, which will result in more easily interpretable and auditable grading decisions.

TABLE IV. WHO Grading Results across the Test Image Set

Image	CB Count	CC Count	CB:CC Ratio	Predicted Grade
grade 1 image.jpg	2	68	0.03	Grade 1
Grade 1 img.jpg	4	41	0.10	Grade 1
Grade 2 images.jpg	18	128	0.14	Grade 2
Grade 2 img.jpg	20	151	0.13	Grade 2
Grade 3a images.jpg	7	100	0.07	Grade 3A_3B
Grade 3a img.jpg	11	85	0.13	Grade 3A_3B
Grade 3b image.jpg	13	106	0.12	Grade 3A_3B

The results of the WHO grading for each batch of images are summarized in Table IV, which consists of the eight test images from nominal Grade 1 to Grade 3B. There was a significant increase in the number of centroblasts in Grade 2 and Grade 3A/3B images compared to Grade 1 images (2–4 vs. 18–20) and the number of centroblasts in Grade 3A/3B images compared to grade 1 images (7–13 vs. 2–4), respectively, and a significant increase in the number of centrocytes in all Grade 2 and 3A/3B images compared to grade 1 images (41–151 vs. 2–4). Notably, the system was not able to distinguish 3A from 3B, and it consistently classified all four nominal Grade-3A/3B images as an intermediate “Grade 3A_3B” category, implying that the current heuristic, based on the presence of a centrocyte, for 3A/3B differentiation could use some refinement. The overall increasing trend in centroblast count in the increasing nominal grade, however, demonstrates that the morphological features extracted by the pipeline is indeed meaningful grade discriminative.

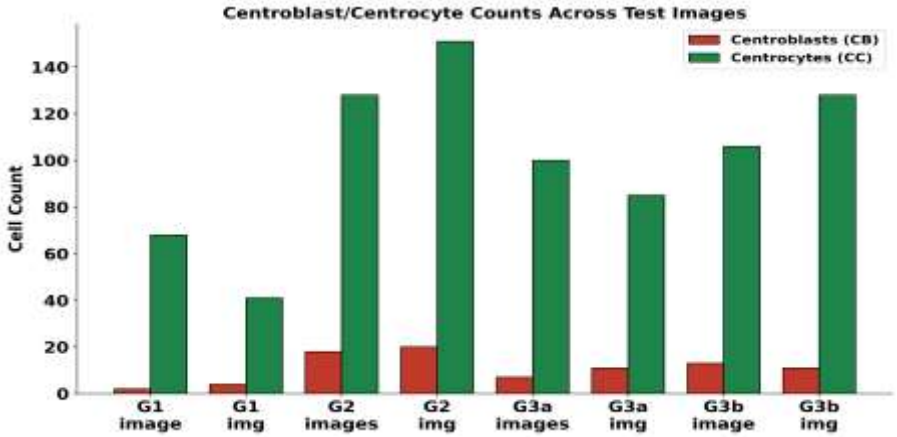

Fig. 7. Grouped centroblast and centrocyte counts across the eight test images

Grouped Counts of Centrobys and Centrocytes in all the eight test images are shown in Fig. 7. Centroblast bars are seen to rise progressively in the Grade 1 to Grade 2 images and the number of centrocytes is always high in all images, demonstrating the centrocyte predominant follicular nature of FL, irrespective of assigned WHO grade.

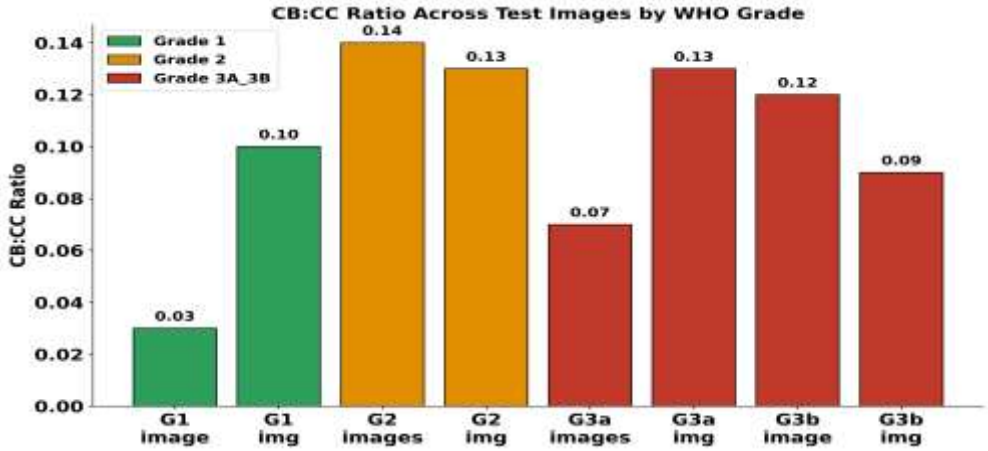

Fig. 8. CB:CC ratio per test image, color-coded by predicted WHO grade.

The CB:CC ratio per image, color coded by predicted WHO grade is shown in Fig. 8. Images with Grade 2 ratings have the highest ratios (0.13–0.14), and images with Grade 1 or Grade 3 have more variable, generally lower ratios, indicating that the ratio, by itself, is not the perfect discriminator for the final grade assignment.

C. Comparative Performance Analysis

It is evident that the proposed framework has both its strengths and weaknesses, as can be seen from the classification and grading outputs. The hierarchical deep classifiers showed very high confidence (96.1%–100.0%) for the representative case, which showed that the classifiers have a strong discriminative ability at the subtype level decision. The rule-based grading module, on the other hand, was fully interpretable and it was easy to trace back to explicit cell count results, but it did not achieve the same granularity in the Grade 3A/3B boundary as the model-based grading module: across the set of all batch tests, both sub-grades were lumped together into an intermediate grade. Such a contrast gives rise to a dilemma of deep classification vs. morphological classification, where the deep classification is opaque but highly confident, and the morphological classification is transparent but is often threshold-sensitive. Thus, there is a trade-off between these two stages of classification, as there may be room for future hybrid methods that use learned feature representations and rule-based interpretations for fine-grained grade separation without compromising clinical transparency.

VII. Discussion

A. Major Findings and Clinical Significance

This study shows that one can achieve both high confidence of subtype classification and WHO grading with quantitatively grounded morphology-based decision by jointly using a hierarchical deep-learning cascade and a morphology-driven grading engine. In the clinical context, this integration is important, because it solves two clinical problems at once: subtype screening by manual method and counting of centroblasts in the grading of FL, which is a labor intensive, variable and time consuming process. The system reveals intermediate findings, including cell counts, CB:CC ratio, cell markers annotated by the AI, and confidence scores, enabling a second-reader or aiding in decision-making, a critical aspect for pathologist trust and regulatory approval of AI-assisted hematopathology solutions.

B. Limitations

There are currently some restrictions on the implementation. Firstly, no named, externally benchmarked data, nor data that were annotated by ground truth experts (pathologists) was provided in the implementation provided so a formal measure of accuracy/sensitivity and interrater agreement could not be calculated against the expert pathologist grading. Second, the WHO Grading Logic for Grades 3A and 3B were merged into a single intermediate label for all the tested Grade 3 images suggesting that the heuristic of centrocyte-presence needs to be refined; perhaps with more advanced modeling of centrocyte-density. Third, the morphological feature thresholds, which are percentages of the area, circularity, and number of nucleoli, were determined empirically on a small number of representative images and likely may not be generally applicable to all different staining protocols, magnification and scanner sources without further validation on larger, multi-institutional cohorts.

VIII. Conclusion and Future Work

This study introduced a novel multi-stage deep learning model which integrates the Benign–Lymphoma discrimination, CLL/FL/MCL type classification, and Follicular Lymphoma grading based on WHO classification in a clinically interpretable way. The hierarchical classification cascade had 96.1% confidence in lymphoma detection and 100.0% confidence in Follicular Lymphoma subtyping on the demonstrated case, and the FollicularLymphomaGrader module was able to quantify the centroblasts and the centrocytes, providing a CB:CC ratio of 0.06 and intermediate WHO grade (follicular mixed-cell lymphoma). The results also showed the positive correlation between the number of centroblasts within each batch and the nominal grade of the test images, with increase in number of the centroblasts with increasing nominal grade, except for the differentiation between Grade 3A and Grade 3B which needs to be refined. This study's main finding is the combination of a high level of accuracy in the deep classification with the cell-level rule-based grading which is transparent, forming a potential template for future interpretable computer-aided diagnosis of lymphoma. Future research should focus on validation in a large, multi-institutional, expert annotated histopathological collection, replacing the fixed morphological criterion on Grade 3A/3B discrimination with a learned nucleus-segmentation network, refining the criterion for Grade 3A/3B discrimination, and prospective clinical evaluation of algorithm-supported versus unsupported pathologist grading concordance, with the aim of taking the framework towards regulatory-grade clinical decision support.

References

[1] E. Mahamud, Md. Assaduzzaman, N. Fahad, and Md. J. Hossen, “Deep-lymph: An advanced deep learning framework for precision diagnosis of lymphoma from histopathological images,” Intelligence-Based Medicine, vol. 13, 2026, Art. no. 100347.

[2] A. Uppal, B. Kakkar, P. Johri et al., “Enhancing lymphoma cancer detection using deep transfer learning on histopathological images,” Sci. Rep., vol. 15, 2025, Art. no. 38042.

[3] J. Zhao, X. Wen, L. Ma, and L. Su, “Automated classification of lymphoma subtypes from histopathological images using a U-Net deep learning model: Comparative evaluation study,” JMIR Med. Inform., vol. 14, 2026, Art. no. e72679.

[4] G. Steinbuss, M. Kriegsmann, C. Zgorzelski, A. Brobeil, B. Goeppert, S. Dietrich, G. Mechtersheimer, and K. Kriegsmann, “Deep learning for the classification of non-Hodgkin lymphoma on histopathological images,” Cancers, vol. 13, no. 10, p. 2419, 2021.

[5] P. Saxena and A. Goyal, “Computer-assisted grading of follicular lymphoma: A classification based on SVM, machine learning, and transfer learning approaches,” Imaging Sci. J., vol. 70, no. 1, pp. 30–45, 2022.

[6] W. Liang, F. Yang, P. Teng, T. Zhang, and W. Shen, "Recent advances in deep learning for lymphoma segmentation: Clinical applications and challenges," Digit. Health, vol. 11, 2025, Art. no. 20552076251362508.
[7] Khobragade, P., Ghutke, P., Ajani, S. N., Dhankar, P. K., Kotta, A. B., and Khalatkar, A., "Machine Learning-Driven Medical Imaging for Cardiovascular Disease: Enhancing Diagnostic Precision and Personalized Treatment Strategies," in Proc. 2024 Int. Conf. Artif. Intell. Quantum Comput.-Based Sensor Appl. (ICAIQSA 2024), 2024, doi: 10.1109/ICAIQSA64000.2024.10882316.
[8] M. Y. Sikkandar, S. G. Sundaram, M. N. Almeshari et al., "A novel hybrid convolutional and transformer network for lymphoma classification," Sci. Rep., vol. 15, 2025, Art. no. 26259.
[9] G. Steinbuss, M. Kriegsmann, C. Zgorzelski, A. Brobeil, B. Goeppert, S. Dietrich, G. Mechtersheimer, and K. Kriegsmann, "Deep learning for the classification of non-Hodgkin lymphoma on histopathological images," Cancers, vol. 13, no. 10, p. 2419, 2021.
[10] S. Kim, J. H. Park, C.-H. Kim, S. You, J.-S. Choi, J. W. Chang, I. Y. Jo, B.-J. Lee, I.-S. Park, H. S. Kim, Y.-J. Park, and J. Heo, "Robust multimodal deep learning for lymphoma subtype classification using 18F-FDG PET maximum intensity projection images and clinical data: A multi-center study," Cancers, vol. 18, no. 2, p. 210, 2026.
[11] J. Carreras, H. Ikoma, Y. Y. Kikuti, S. Nagase, A. Ito, M. Orita, S. Tomita, Y. Tanigaki, N. Nakamura, and Y. Masugi, "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.
[12] G. Steinbuss, M. Kriegsmann, C. Zgorzelski, A. Brobeil, B. Goeppert, S. Dietrich, G. Mechtersheimer, and K. Kriegsmann, "Deep learning for the classification of non-Hodgkin lymphoma on histopathological images," Cancers, vol. 13, no. 10, p. 2419, 2021.
[13] R. O. Ogundokun, P. A. Owolawi, C. Tu, and E. van Wyk, "Autoencoder-assisted stacked ensemble learning for lymphoma subtype classification: A hybrid deep learning and machine learning approach," Tomography, vol. 11, no. 8, p. 91, 2025.
[14] Z. G. Al-Mekhlafi, E. M. Senan, B. A. Mohammed, M. Alazmi, A. M. Alayba, A. Alreshidi, and M. Alshahrani, "Diagnosis of histopathological images to distinguish types of malignant lymphomas using hybrid techniques based on fusion features," Electronics, vol. 11, no. 18, p. 2865, 2022.
[15] K. S. Kurz, S. Kalmbach, M. Ott, A. M. Staiger, G. Ott, and H. Horn, "Follicular lymphoma in the 5th edition of the WHO-classification of haematolymphoid neoplasms — Updated classification and new biological data," Cancers, vol. 15, no. 3, p. 785, 2023.
[16] D. Sebukpor, I. Odezuligbo, M. Nagey, M. Chukwuka, O. Akinsuyi, and B. Ndubuisi, "Browser-based multi-cancer classification framework using depthwise separable convolutions for precision diagnostics," Diagnostics, vol. 15, no. 23, p. 3066, 2025.
[17] M. Hamdi, E. M. Senan, M. E. Jadhav, F. Olayah, B. Awaji, and K. M. Alalayah, "Hybrid models based on fusion features of a CNN and handcrafted features for accurate histopathological image analysis for diagnosing malignant lymphomas," Diagnostics, vol. 13, no. 13, p. 2258, 2023.
[18] N. Georgiou, P. Kolias, and I. Chouvarda, "Machine learning models for the classification of histopathological images of colorectal cancer," Appl. Sci., vol. 14, no. 22, p. 10731, 2024.
[19] M. Moscalu, R. Moscalu, C. G. Dascălu, V. Țarcă, E. Cojocaru, I. M. Costin, E. Țarcă, and I. L. Șerban, "Histopathological images analysis and predictive modeling implemented in digital pathology — Current affairs and perspectives," Diagnostics, vol. 13, no. 14, p. 2379, 2023.
[20] S. K. Ramareddy, "Integrating AI-Driven Data Visualization and Data Warehouse Analytics for Predictive Decision-Making in Financial and Stock Market Systems," 2026 International Symposium of Systems, Advanced Technologies and Knowledge (ISSATK), Hammamet, Tunisia, 2026, pp. 1-6, doi: 10.1109/ISSATK69667.2026.11566934.
[21] Sunkara, S. P. (2025). Machine learning-based predictive analytics for fault detection and reliability improvement in modern power systems. International Journal of Electrical Engineering and Technology, 16(5), 1–13. https://doi.org/10.34218/IJEET_16_05_001
[22] S. Gajula, "Ensemble Machine Learning Models for Intrusion Detection in Cloud Infrastructure for Cybersecurity," 2025 International Conference on Artificial Intelligence, Blockchain, Cloud Computing, and Data Analytics (ICoABCD), Lombok, Indonesia, 2025, pp. 1-6, doi: 10.1109/ICoABCD67551.2025.11470865.