

Brief Report

Not peer-reviewed version

Prediction of Overall Survival Death < 2 Years in Diffuse Large B-Cell Lymphoma Based on Histological Images and Deep Neural Networks

[Joaquim Carreras](#) *

Posted Date: 25 March 2026

doi: 10.20944/preprints202601.0065.v3

Keywords: DLBCL; prognosis; overall survival; deep learning; CNN; XAI; computer vision; AI; reactive lymphoid tissue



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Brief Report

Prediction of Overall Survival Death < 2 Years in Diffuse Large B-Cell Lymphoma Based on Histological Images and Deep Neural Networks

Joaquim Carreras

Department of Pathology, School of Medicine, Tokai University. 143 Shimokasuya, Isehara, Kanagawa, 259-1193, Japan; joaquim.carreras@tokai.ac.jp

Abstract

Diffuse large B-cell lymphoma (DLBCL) is one of the most frequent lymphomas. This proof-of-concept study predicted the prognosis of DLBCL using hematoxylin and eosin (H&E) histological images, computer vision and deep learning. The series included 114 DLBCL cases, split into 2 prognostic groups according to the overall survival, and 44 reactive lymphoid tissue. The curve fitting and slope analysis showed a point of inflection at 2 years, which differentiated patients with aggressive clinical evolution ("Dead < 2 years", $b1 = -0.024$), from the rest with moderate clinical evolution ("Others", $b1 = -0.003$). Twenty different convolutional neural networks (CNNs) were used, and explainable artificial intelligence (XAI) was also applied. The final model based on DarkNet-19 predicted prognosis groups with high performance (test set accuracy = 96.3%). The other performance parameters were precision (94.5%), recall (95.0%), false positive rate (3.1%), specificity (96.9%), and F1 score (94.7%). XAI, including grad-CAM, occlusion sensitivity, and image-LIME, confirmed that the CNN focused on the correct areas. Hybrid partitioning to prevent information leakage with patient-based analysis, image classification between DLBCL and 44 cases of reactive lymphoid tissue, and hyperparameter tuning were also successfully performed. Correlation with the clinicopathological characteristics found that the Dead < 2 years group was correlated with stage III-IV, International Prognostic Index (IPI) High + High/intermediate, progressive disease, non-GCB cell-of-origin, CD10-, BCL2+, and Epstein-Barr virus (EBER)+. Analysis of the microenvironment, immune checkpoint, cell cycle, and germinal center markers showed that Dead < 2 years had higher IL10, PD-L1, and CD163, and lower E2F1 protein expressions. No differences were found for Ki67, CSF1R, CASP8, TNFAIP8, LMO2, MYC, MDM2, CDK6, and TP53 markers at quantitative level. In conclusion, the DLBCL overall survival can be predicted using H&E histological images and deep learning using 2 years point (similar to POD24). This trained CNN can be used as a pretrained model for transfer learning in the future.

Keywords: DLBCL; prognosis; overall survival; deep learning; CNN; XAI; computer vision; AI; reactive lymphoid tissue

1. Introduction

Diffuse large B-cell lymphoma (DLBCL) is one of the most frequent histological subtypes of non-Hodgkin lymphoma (NHL), accounting for approximately 25% of adult NHL cases [1]. Most patients present with a rapid enlarging symptomatic mass with nodal enlargement. In approximately 60% of the cases, DLBCL presents with advanced-stage (III or IV) and high serum lactate dehydrogenase (LDH), and 30% with fever, weight loss, night sweats, and bone marrow infiltration [2–7]. DLBCL is cured in 60% of the cases with current therapy, particularly in patients who achieve complete response with first-line treatment. However, in 40% of the cases, the clinical evolution is unfavorable [5,8,9].

DLBCL is a heterogeneous clinicopathological entity. It is derived from germinal center B cells (centroblasts) or post-germinal activated B cells (immunoblasts) [10,11]. The tumor cells of DLBCL are large (e.g. nuclei twice the size of a small lymphocyte and larger than the nucleus of a macrophage) [10–17]. Centroblasts are large, noncleaved cells with round or oval nuclei, vesicular chromatin, often with multiple peripheral nucleoli, and a narrow rim of basophilic cytoplasm [10–17]. Immunoblasts are usually larger cells with prominent nucleoli and more abundant cytoplasm, often with plasmacytoid characteristics. Some cases are mixed, and other morphological variants exist [10,11,13,14]. The pathogenesis of DLBCL is complex and shows a heterogeneous landscape [1].

DLBCL not otherwise specified (NOS) is defined as having a mature B-cell phenotype and large cell morphology, but having none of the criteria that define specific large B-cell lymphoma subtypes. Using either immunohistochemistry or gene expression studies of DLBCL, the cell-of-origin (COO) status can be determined, including germinal center B-cell (GCB) versus activated B-cell (ABC) subtypes [2,5]. Other large B-cell lymphoma subtypes are primary mediastinal, T-cell/histiocyte-rich, plasmablastic, primary cutaneous leg-type, immune-privileged sites, intravascular, associated with chronic inflammation, IRF4 rearranged, ALK-positive, with 11q aberration, primary effusion lymphoma, and Epstein-Barr virus (EBV)-positive DLBCL [2,5].

Nodal DLBCL can spread to other organs, such as the liver, kidneys, lung, bone marrow, and central nervous system. Extranodal/extramedullary involvement is frequently present in early-stage disease (stage I/II). The most frequent primary extranodal presentation is the gastrointestinal tract, but virtually any tissue could be affected, including the testis, bone, thyroid, salivary glands, tonsils, and skin [1,18–20].

Several prognostic models are used in DLBCL. The International Prognostic Index (IPI) and its variants are the main prognostic variables routinely used. The original IPI included the following factors: age >60 years, high serum LDH, Eastern Cooperative Oncology Group (ECOG) performance status ≥ 2 , clinical stage III–IV, and number of extranodal sites >1 associated with poor prognosis [21]. The COO status determined by immunohistochemistry using CD10, BCL6, and MUM1 or by gene expression profiling correlated with non-GCB/ABC-like with poorer prognosis of the patients [21–27]. *MYC* rearrangement is seen in 5–15% of the cases and can be associated with *BCL2* and *BCL6* rearrangements [28]. Double-hit *MYC* and *BCL2* cases have a worse prognosis [2,5,29–35]. Other techniques, such as deep sequencing, have confirmed the DLBCL heterogeneity and identified driver mutations with different clinical outcomes [2,5,6,17,36–38].

The histological characteristics of the tissue reflect the proteomic, transcriptomic, and genomic pathological background. This study aimed to predict the prognosis of patients with DLBCL based only on the histological evaluation of hematoxylin and eosin staining. The 2 years timepoint, which is similar to the progression of disease 24 (POD24), shows a change in the slope of the survival curve. Therefore, the DLBCL cases were classified according to their overall survival: “aggressive” and “moderate”. Using several models, the prognostic was predicted with high performance, and several explainable artificial intelligence (XAI) techniques were used to visually highlight the regions that are most important for the deep neural network’s classification decision.

2. Materials and Methods

2.1. Patients and Samples

A series of 114 patients with a histological diagnosis of conventional diffuse large B-cell lymphoma (DLBCL) from 2007 to 2011 were selected from the Department of Pathology, Tokai University, School of Medicine. The cases were diagnosed according to the current classifications of hematolymphoid neoplasia [2,5,10], which included the evaluation of hematoxylin and eosin (H&E), immunophenotype, and molecular techniques when required [2,5,6].

This study was conducted following the guidelines of the Declaration of Helsinki of the World Medical Association and ethical principles for medical research involving human participants. The Institutional Review Board of Tokai University approved the study (IRB20-156).

Immune microenvironment data were retrieved from our previous publications [39–41], and immunohistochemistry was reanalyzed. For CD163, IL10, and PD-L1, new immunohistochemistry (IHC) was performed as the number of cases in this study increased from previous publications. The IHC methodology, including primary antibody details, is shown in Appendix B.

2.2. Overall Survival Curve Analysis

Analysis of the overall survival plot identified 2 regions with different prognoses based on the 2 years point. The rational is that at the 2 years the slope of the overall survival curve changes (point of inflection), changing from aggressive ($b1 = -0.024$; $y = 0.98 * \exp(-0.024 * x)$; exponential equation, $R^2 = 0.985$, $p < 0.001$) to moderate clinical behavior ($b1 = -0.003$; $y = 0.69 * \exp(-0.003 * x)$; exponential equation, $R^2 = 0.868$, $p < 0.001$) (Table 1). The “aggressive” group was mainly characterized by death event before 2 years ($p < 0.001$). A logistic regression was performed to ascertain the effect of group variable on the likelihood that participants died. The aggressive group was 15.4 times more likely to exhibit a death event than the moderate group ($p < 0.001$) (Tables 1 and 2) (Figure 1).

Table 1. Difference in death event between the two prognostic groups.

Model summary				Parameter estimates			
Equatio	R	F	P	Consta	B1	B2	B3
Linear	0.825	528.5	<	0.831	-		
Quadrat	0.893	461.2	<	0.874	-	2.565E	
Cubic	0.940	572.4	<	0.918	-	0.000	-
Compou	0.888	888.3	<	0.833	0.99		
Growth	0.888	888.3	<	-0.183	-		
Exponen	0.888	888.3	<	0.833	-		

Table 2. Difference in death event between the two prognostic groups.

	Alive	Dead	P	Exp	95% C.I. for
“Aggressive”	52/68	16/68	< 0.001	15.4	5.9 – 39.8
“Moderate”	8/46	38/46	< 0.001	0.31	

C.I., confidence internal.

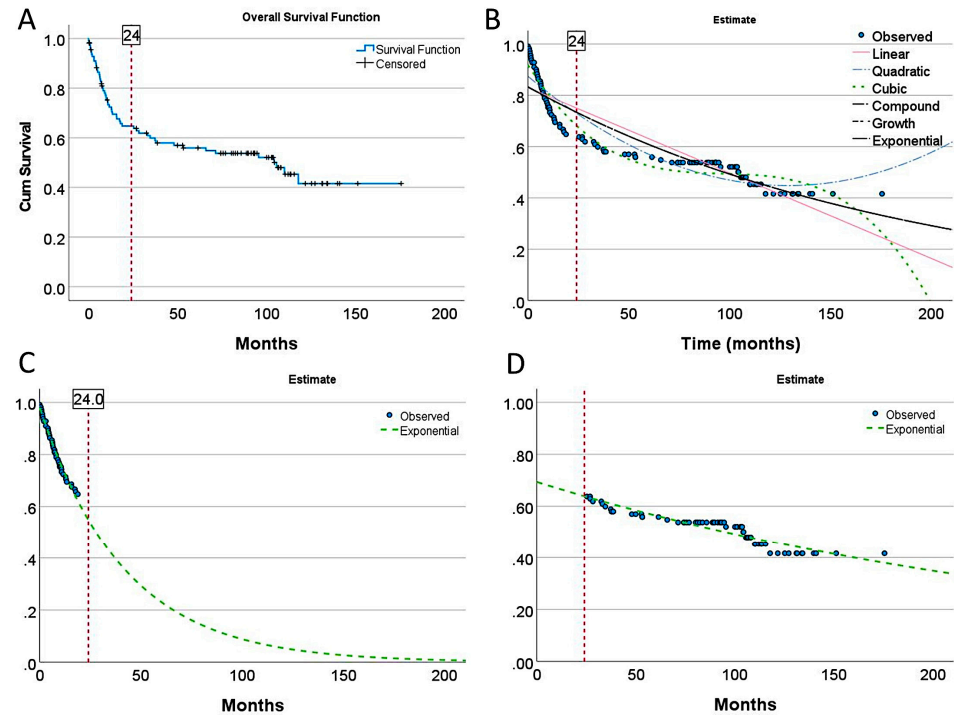


Figure 1. Overall survival groups. (A) The cases were grouped according to their survival based on the slope of the Kaplan-Meier survival curve and point of inflection at the 2 years point: “aggressive” (mainly cases who died within the first 24 months (Dead < 2 years); $b1 = -0.024$), and more “moderate” clinical evolution ($b1 = -0.003$). (B) A detailed analysis was performed using curve estimation. (C) An exponential curve estimation allowed quantifying the different slopes of each group.

2.3. Survival Groups

Based on the overall survival curve and the 2 years inflection point 2 survival groups were defined: the aggressive clinical behavior group characterized by a death event before the 2 years (“Dead < 2 years”) and the “Others” group. The “Others” group included all cases with a follow-up of > 2 years both censored (alive) or with death event. The “Others” group initially also included 8 cases of alive patients that had a follow-up of less than 2 years. Notably, in a subsequent analysis these 8 cases were excluded from the study.

2.4. Deep Learning Image Classification

Whole-tissue sections were stained with H&E and digitized using a slide scanner (NanoZoomer S360, C13220-01, Hamamatsu Photonics K.K.). The whole neoplastic areas were identified, digitally extracted at 20× magnification and 150 dpi, and split into image patches of 224×224×3 resolution. After splitting, image patches of size different from 224×224 and with less than 80% of tissue were discarded.

First, the series was split into a training set (70%) and a validation set (10%) to help prevent overfitting during training and a test set (20%). The training set patches were shuffled before training. No augmentation options were used during the training, including random reflection axis, rotation (degrees), rescaling, horizontal translation (pixels) or vertical translation (pixels).

The training options were as follows: sgd solver, 0.001 initial learning rate, constant learning rate schedule, 128 minibatch size, 5 max epochs, and 50 validation frequency. In the NasNet-Large CNN, due to hardware limitations to run the analysis, the minibatch size was set at 16.

Image patches were classified using transfer learning and 20 different types of CNN, including AlexNet, DarkNet-19, DarkNet-53, DenseNet-201, EfficientNet-b0, GoogleLeNet, GoogleLeNet-places365, Inception-ResNet-v2, Inception-v3, MobileNet-v2, NasNet-Mobile, NasNet-Large, ResNet-18, ResNet-50, ResNet-101, Shufflenet, SqueezeNet, VGG-16, VGG-19, and Xception.

The following hardware and software were used: NanoZoomer S360, #C13220-01 (Hamamatsu Photonics K.K.); 12-Core AMD Ryzen 9 5900X CPU, 4900 MHz; 49075 MB DDR4-3200 SDRAM; NVIDIA GeForce RTX 4080 SUPER GPU; MATLAB R2023b Update 10 (23.2.0.2859533), 64-bit (win64); PhotoScape v3.7; IBM SPSS Statistics version 27 (Release 27.0.1.0, 64-bit edition); NDP.view 2.9.29 (RUO) 2022/01/14 (Hamamatsu Photonics K.K.).

Figures 2 to 7 depict the methodological design.

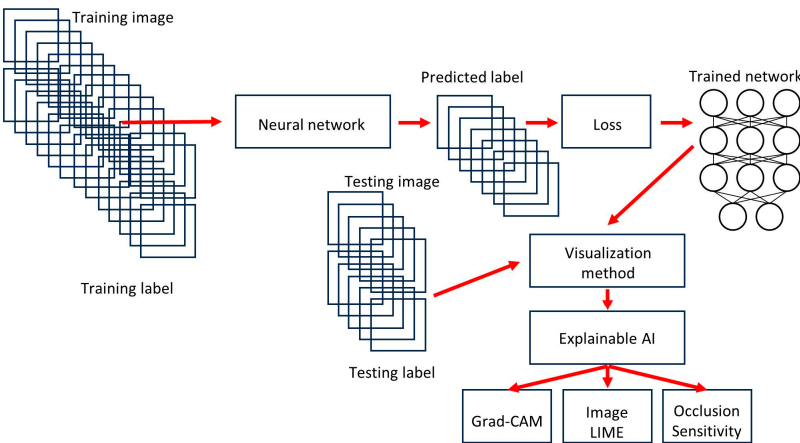


Figure 2. Methodology. The series was randomly split into training (70%) and validation (10%) sets for training, and test (20%) set using new data. Explainable artificial intelligence (XAI) was performed using Grad-CAM, image LIME, and occlusion sensitivity. First, since the aim was to identify which models were more suitable (best time and performance ratio) and to create a pretrained convolutional neural network (CNN) to be used in the future as transfer learning in a larger series of cases of lymphoma, the initial analysis was patched-based. All image patches of $224 \times 224 \times 3$ were pooled into three different folders: training set for training the network, validation set for testing the performance during training, and test set used after training to assess how well the network performed in new data. All patches were mutually exclusive between folders; no repeated patches were found. During the deep learning workflow, common types of transformations, such as geometric transformations, cropping, and adding noise was not performed. Secondly, all analyses were repeated using a hybrid partitioning and test set at the patient level to avoid information leakage.

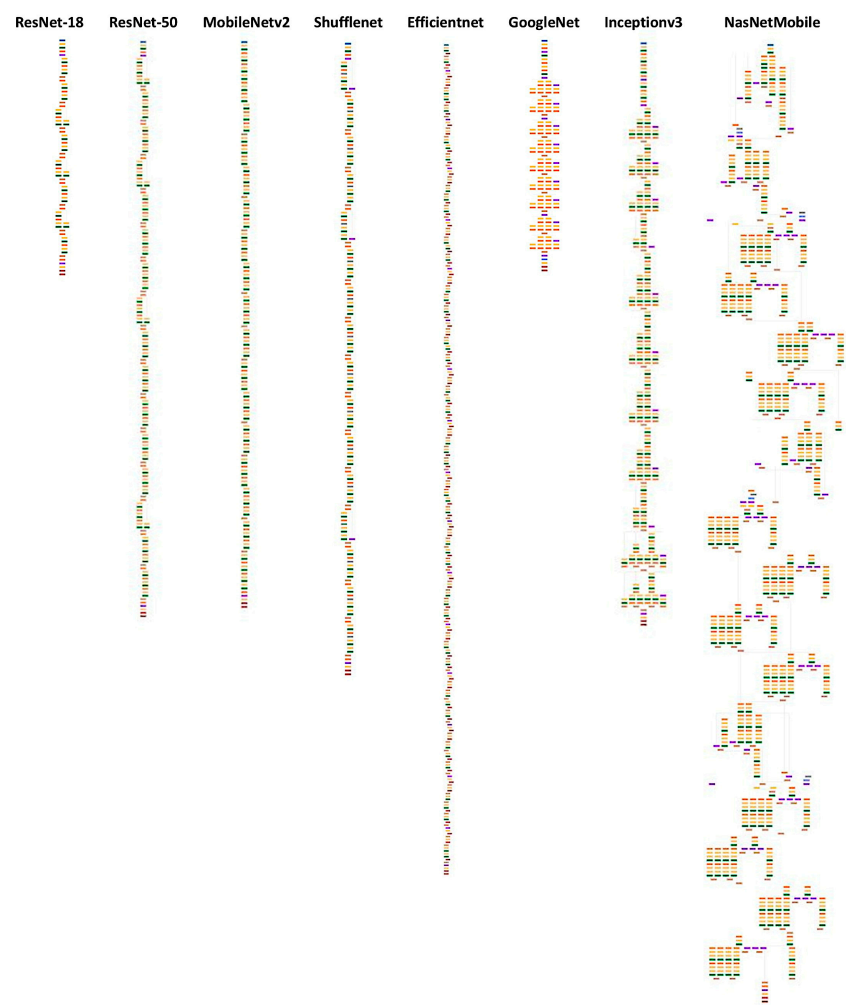


Figure 3. Architectures of the main different types of convolutional neural networks used in this study.

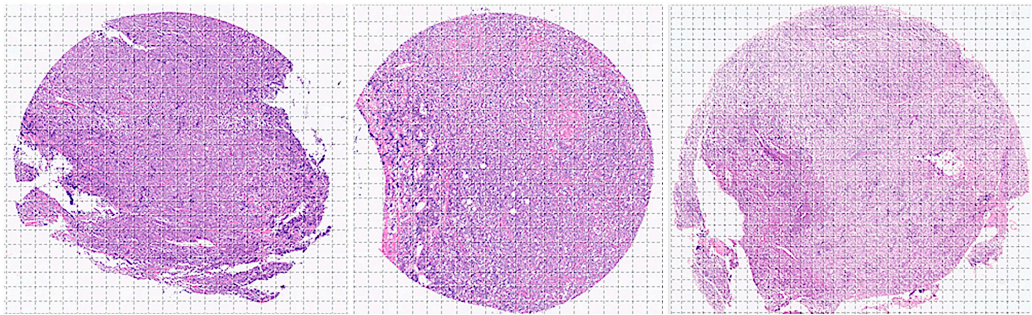


Figure 4. Image splitting. Histological images stained with hematoxylin and eosin (H&E) were first digitalized at 200×magnification and 150 dpi. Later, the images were split at 224×224×3 resolution, which is suitable for convolutional neural network processing. This image shows the splitting of 3 cases of DLBCL obtained from the H&E of a tissue microarray.

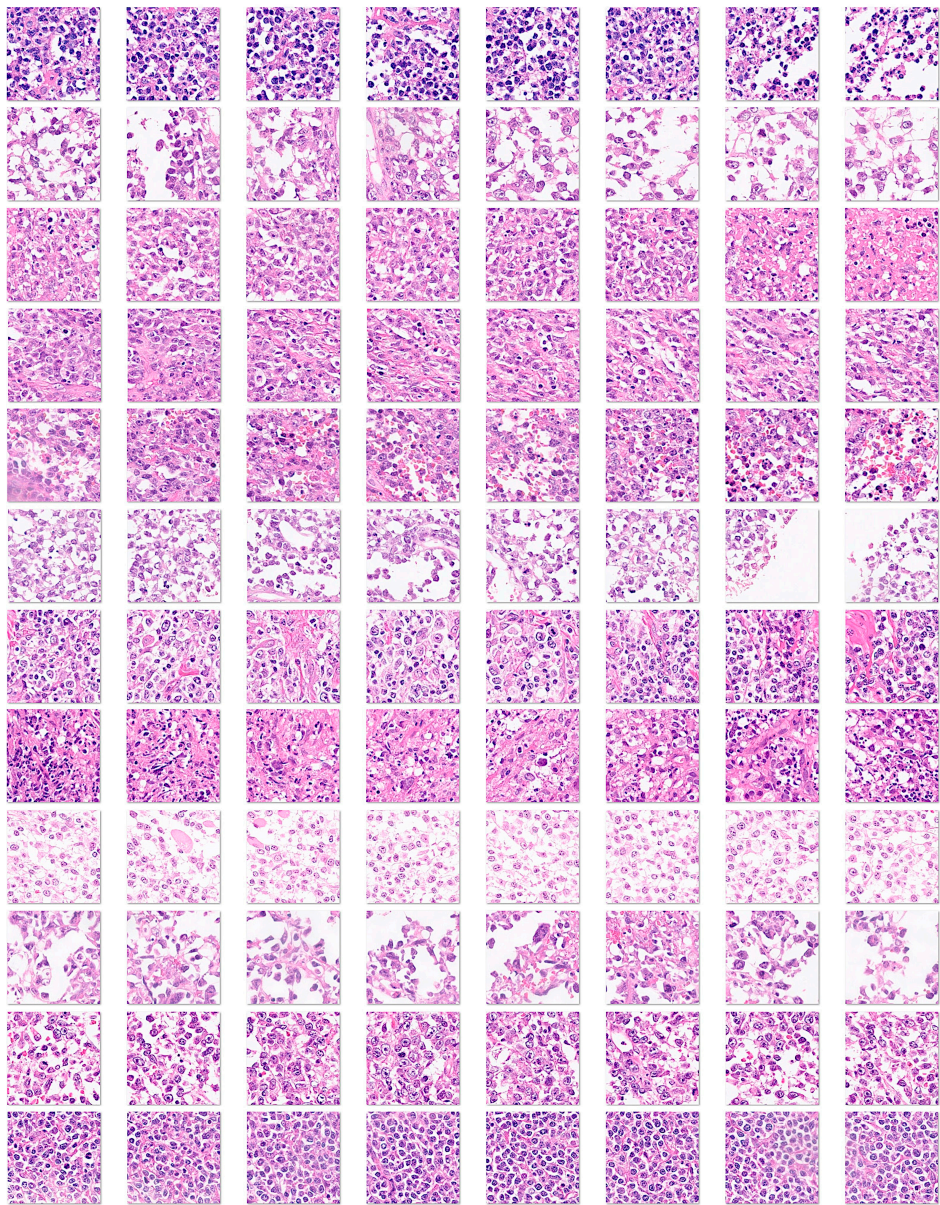


Figure 5. Image splitting of the aggressive (Dead < 2 years) group. These DLBCL cases were characterized by death events before the first 2 years of overall survival. Histological images stained with hematoxylin and

eosin (H&E) were first digitalized at 200×magnification and 150 dpi. Later, the images were split at 224×224×3 resolution, which is suitable for convolutional neural network processing. This image shows the splitting of 12 cases of DLBCL, which is a heterogeneous clinicopathologic entity. DLBCL is derived from germinal center B cells (centroblasts) or post-germinal activated B cells (immunoblasts).

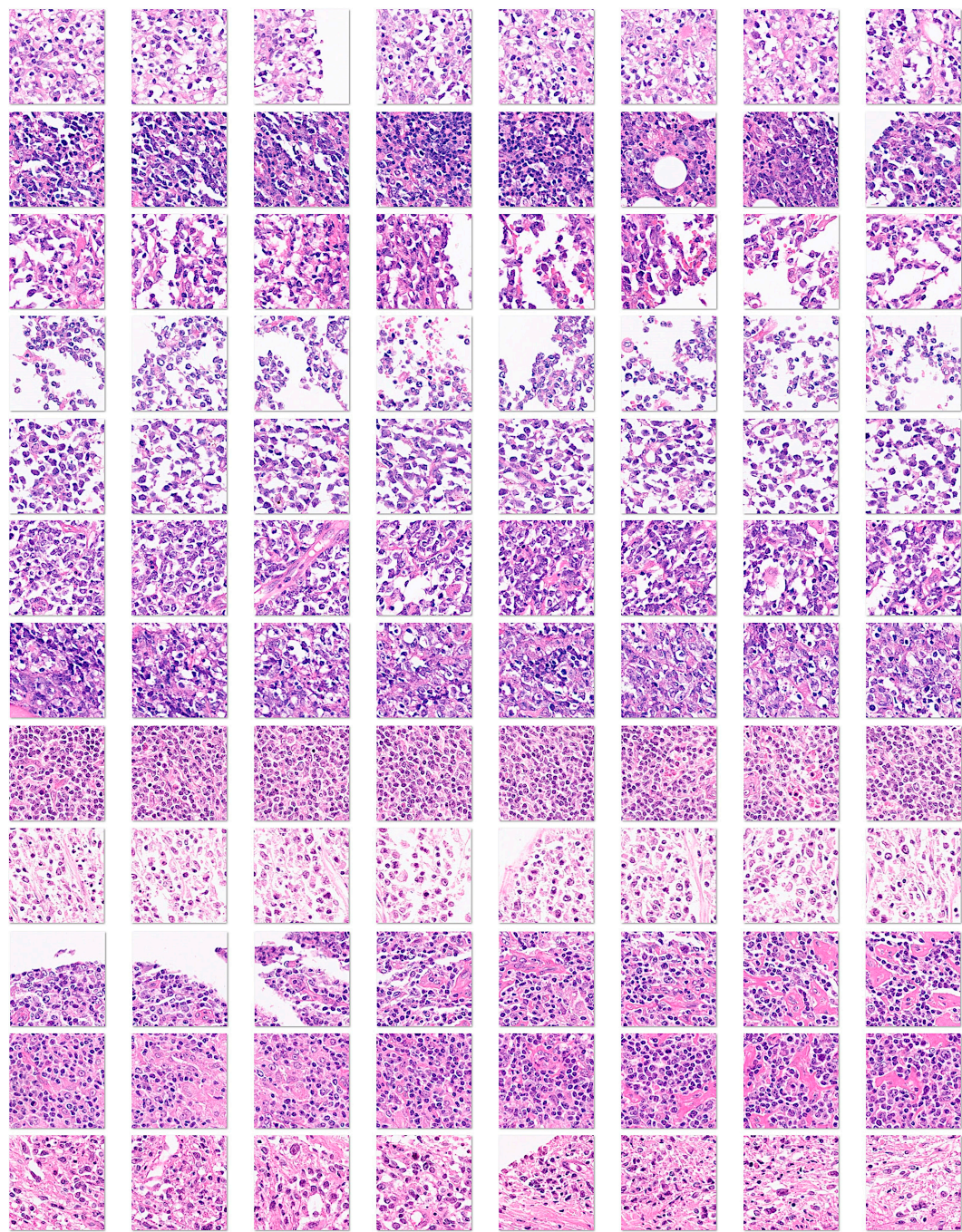


Figure 6. Image splitting of the Others group). This image shows examples of DLBCL cases that did not have a death event before the 2 years of overall survival follow-up (i.e. less aggressive cases, “others”). Histological images stained with hematoxylin and eosin (H&E) were first digitalized at 200×magnification and 150 dpi. Later, the images were split at 224×224×3 resolution, which is suitable for convolutional neural network processing. This image shows the splitting of 12 cases of DLBCL. DLBCL is a heterogeneous clinicopathologic entity. DLBCL is derived from germinal center B cells (centroblasts) or postgerminal activated B cells (immunoblasts).

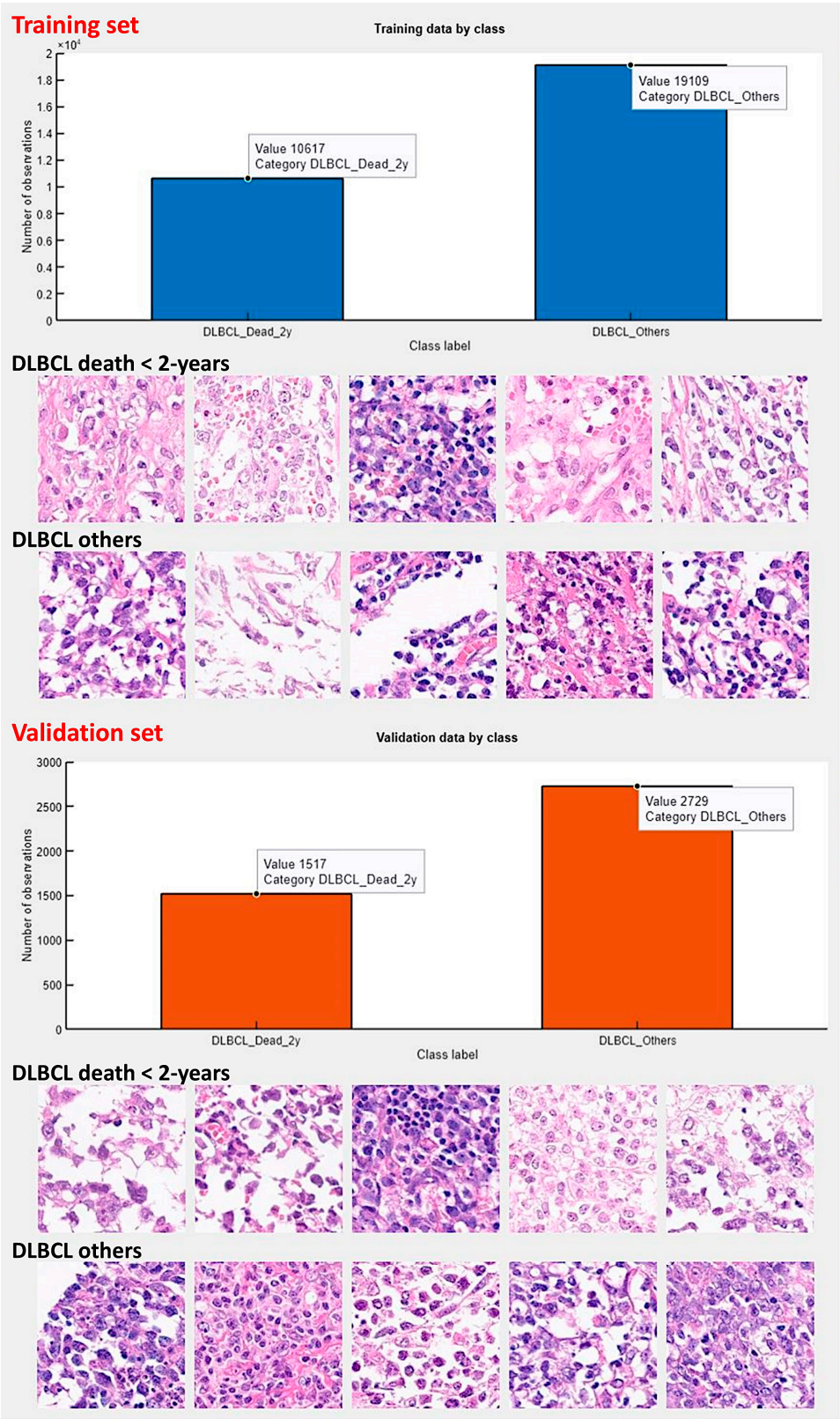


Figure 7. Training and validation datasets. The series was split into a training set (70%), validation set (10%) to help prevent overfitting, and test set (20%). No augmentation options were used during the training. This figure shows the number of image patches and random example images of the training and validation sets.

3. Results

3.1. Dataset 1. CNN Performance Analyses and Image Classification

In the training/validation set, DarkNet-19 had the best validation accuracy (95.71%), and the accuracy/training time ratio was moderate (0.114). The most efficient architecture was ResNet-18, which had a validation accuracy of 92.11% and a ratio of 0.426 (Table 3 and Figure 8).

Table 3. Training/validation set CNN performance.

CNN model	Layers	Latency	Connections	Image Input	Training time	Validation	Efficiency	Relative
NasNet-Large	84	1	146	331×3	1429	96.42	0.00	564.2
DarkNet-19	20	6	63	256×2	14	95.71	0.11	5.5
DarkNet-53	41	1	206	256×2	120	95.6	0.01	47.6
DenseNet-201	20	7	805	224×2	255	93.9	0.00	100.7
ResNet-101	44	3	379	224×2	114	93.31	0.01	45.1
Inception-v3	23	3	349	299×2	53	92.25	0.02	21.0
ResNet-50	25	1	192	224×2	14	92.25	0.10	5.5
ResNet-18	11	7	78	224×2	3 min	92.11	0.42	1.4
VGG-16	13	4	40	224×2	155	92.11	0.00	61.3
MobileNet-v2	3.	1	163	224×2	12	90.86	0.11	5.1
Inception-VGG-19	55	8	921	299×2	509	89.83	0.00	201.0
EfficientNet-	14	4	46	224×2	181	89.54	0.00	71.6
GoogleLeNet-	5.	2	363	224×2	54	89.07	0.07	21.3
GoogleLeNet	5.	1	170	224×2	5 min	88.37	0.26	2.2
Shufflenet	6.	1	170	224×2	5 min	88.34	0.27	2.1
NasNet-	1.	1	187	224×2	5 min	88.22	0.25	2.3
Xception	5.	9	107	224×2	30	87.73	0.04	12.0
AlexNet	22	1	181	299×2	522	87.28	0.00	206.4
SqueezeNet	60	2	24	227×2	2 min	83.94	0.55	1.0
	1.	6	75	227×2	2 min	79.35	0.48	1.1

Efficiency: Validation accuracy/time (%/sec). Relative time: faster CNN (AlexNet) as reference.

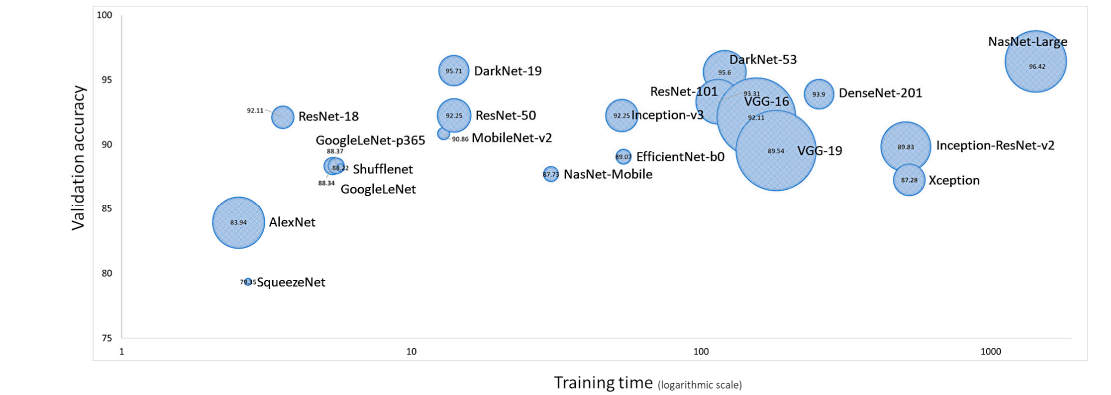


Figure 8. Neural network training performances. The most important characteristics are the neural network accuracy (y-axis), speed (x-axis), and size (circle). Choosing a neural network is a tradeoff between these characteristics. NasNet-Large had the best validation accuracy (96.52%), followed by DarkNet-19 (95.71%). However, in relation with AlexNet that was the fastest, NasNet-Large took 564.2 more time to compute (i.e. relative time).

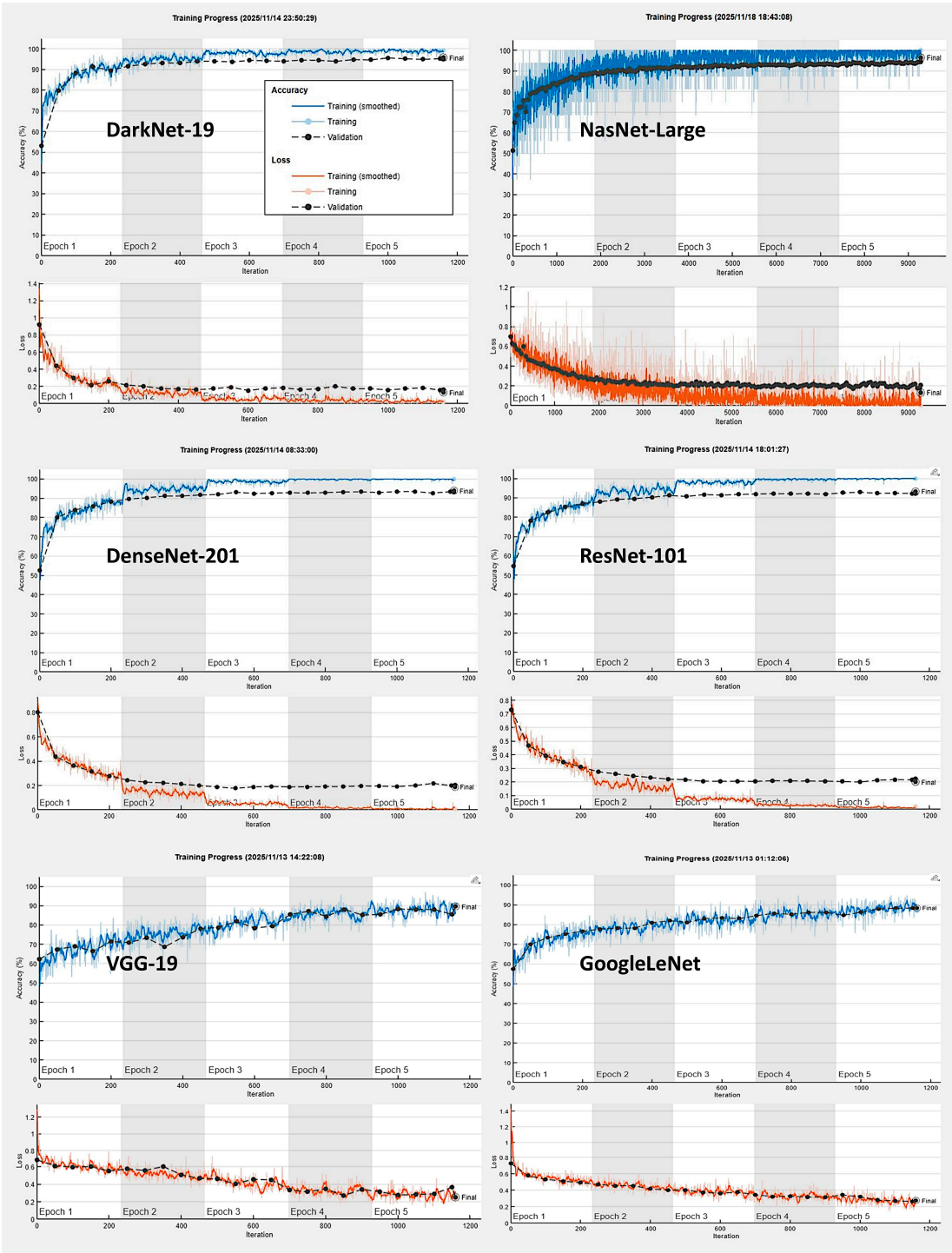


Figure 9. CNN training plots.

In the validation set, DarkNet-19 had the best accuracy (96.26%), followed by NasNet-Large (96.21%), DarkNet-53 (95.47%), and DenseNet-201 (93.67%). The confusion charts of the models with higher performance are shown in Figures 10 and 11. All the performance parameters of the models in the test set are shown in Table 4 and Figure 12.

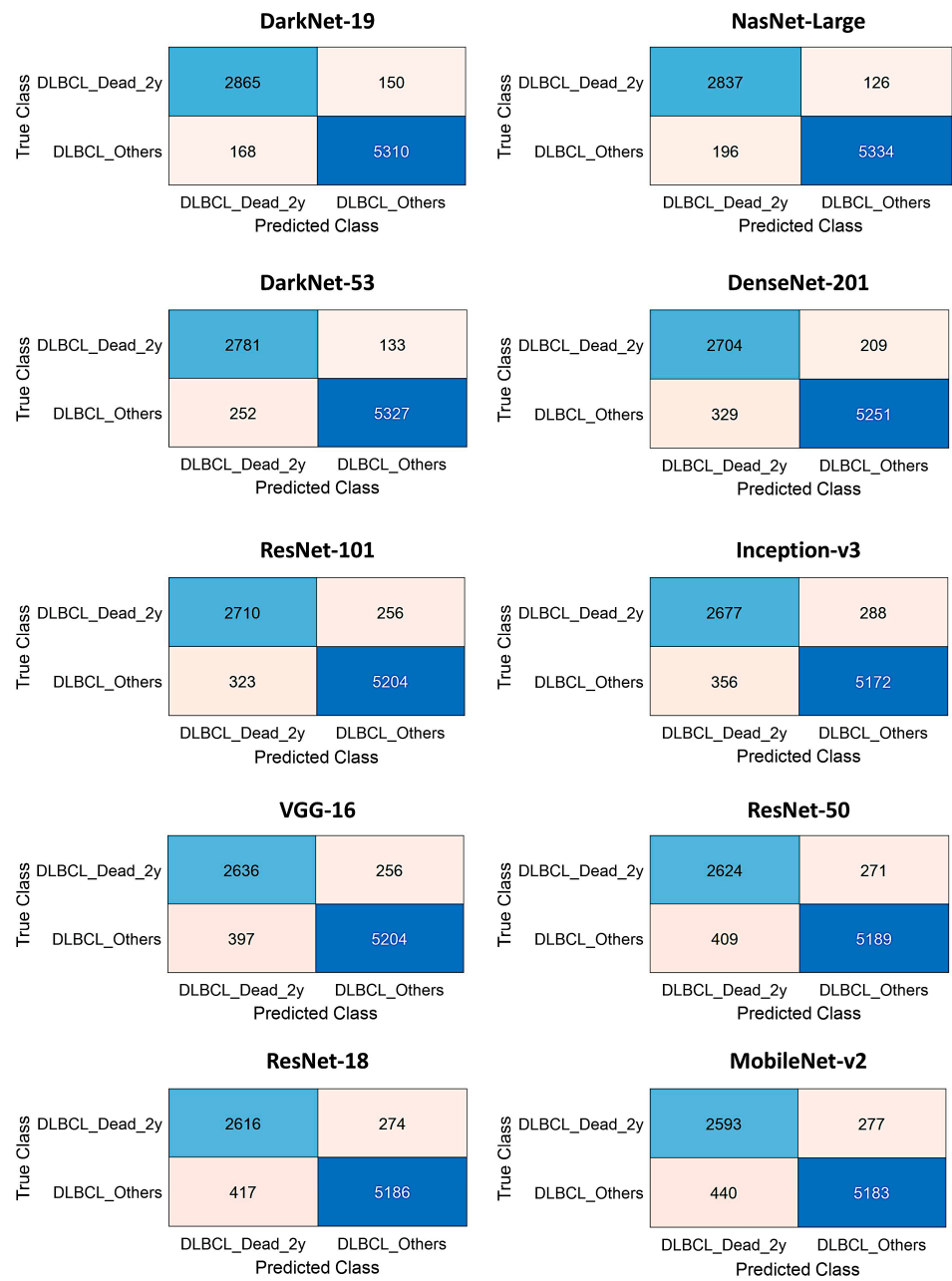


Figure 10. Test set confusion charts.

Table 4. The test set CNN performance parameters.

CNN	Accura	Precisi	Reca	False	Specifi	F1
DarkNet-19	96.26	94.46	95.02	3.07	96.93	94.74
NasNet-	96.21	93.54	95.75	3.54	96.46	94.63
DarkNet-53	95.47	91.69	95.44	4.52	95.48	93.53
DenseNet-	93.67	89.15	92.83	5.9	94.1	90.95
ResNet-101	93.18	89.35	91.37	5.84	94.16	90.35
Inception-v3	92.42	88.26	90.29	6.44	93.56	89.26
VGG-16	92.31	86.91	91.15	7.09	92.91	88.98
ResNet-50	91.99	86.52	90.64	7.31	92.69	88.53
ResNet-18	91.86	86.25	90.52	7.44	92.56	88.33
MobileNet-	91.56	85.49	90.35	7.83	92.17	87.85
Inception-	90.77	84.8	88.84	8.24	91.76	86.77
VGG-19	88.73	83.94	84.42	8.89	91.11	84.18

GoogleLeNet	88.71	86.52	82.67	7.69	92.31	84.55
EfficientNet	88.67	79.72	87.45	10.74	89.26	83.41
GoogleLeNet	88.6	77.78	88.92	11.54	88.46	82.98
Shufflenet	88.6	83.18	84.64	9.25	90.75	83.9
NasNet-	87.72	77.68	86.55	11.73	88.27	81.88
Xception	86.95	78.23	84.11	11.64	88.36	81.07
AlexNet	84.09	75.87	78.8	13.13	86.87	77.31
SqueezeNet	79.16	49.39	86.44	22.71	77.29	62.86

Recall equals sensitivity and true positive rate. False positive rate equals 1-specificity.

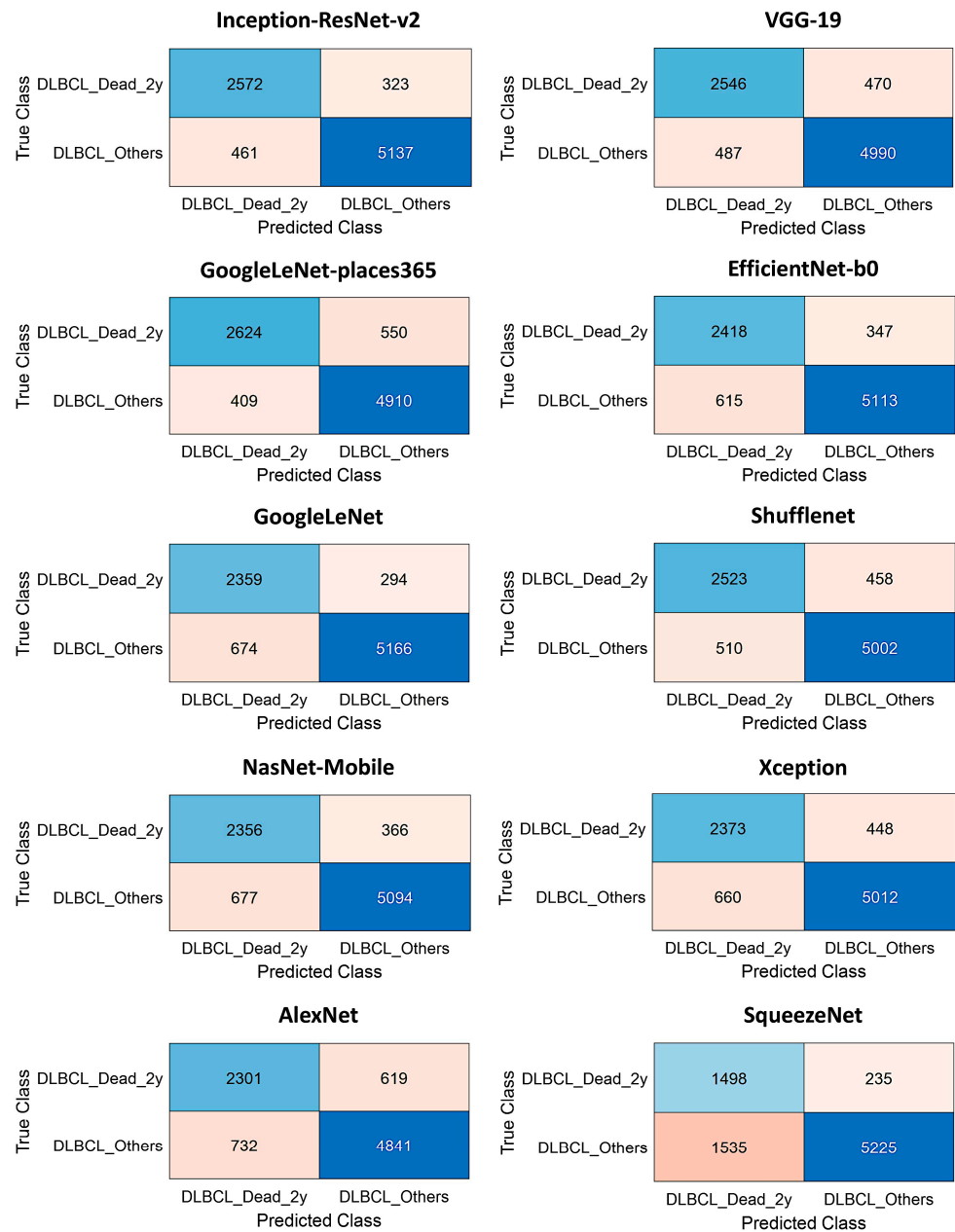


Figure 11. The test set confusion charts, continuation.

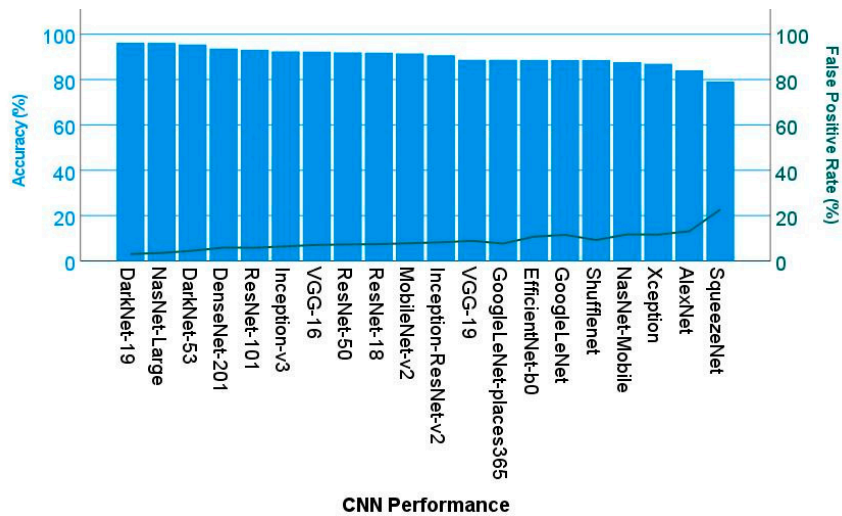


Figure 12. Test set CNN performance. This figure shows the performance parameters of accuracy, and false positive rate of different CNN.

Explainable artificial intelligence (XAI) was used to identify the areas of the images that the network DarkNet-19 used for classification. Figures 13 and 14 show the Grad-CAM, occlusion sensitivity, and image Lime analysis.

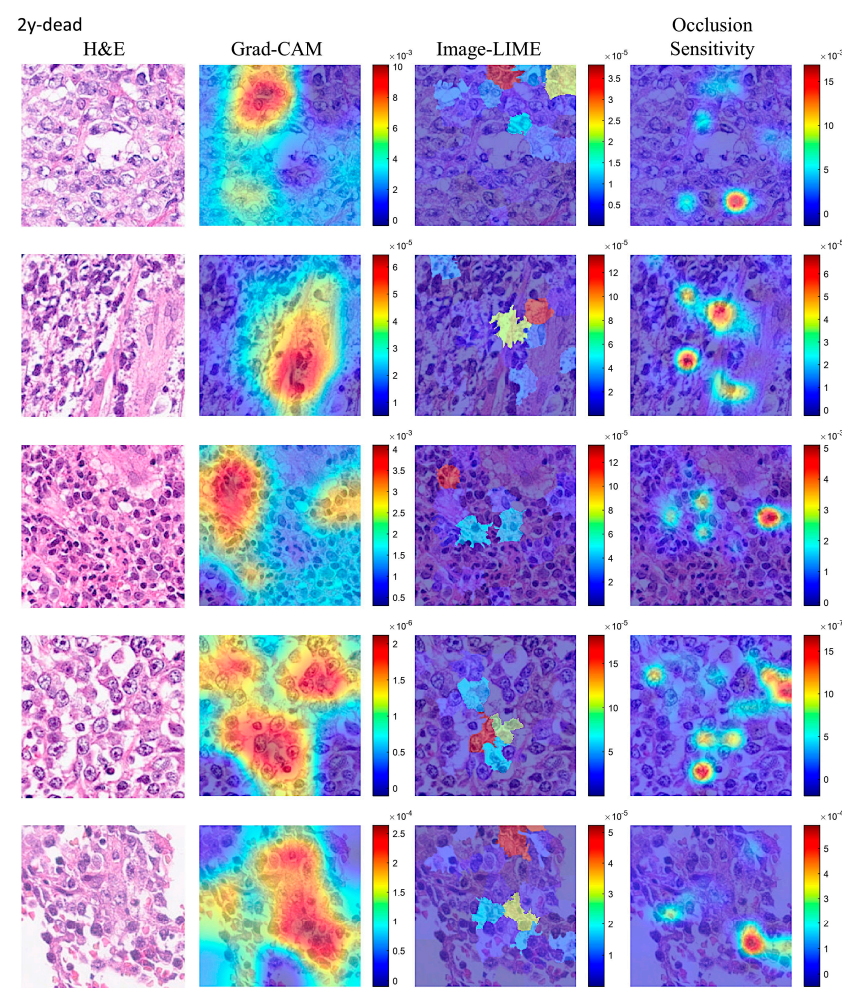


Figure 13. Explainable artificial intelligence (XAI) of the Dead < 2 years group. This figure shows 4 different image patches analyzed with XAI techniques, including grad-CAM, image LIME, and occlusion sensitivity. XAI techniques were used to identify the areas of an image that the network DarkNet-19 used for classification. XAI showed that CNNs focused on the epithelial component of neoplasia.

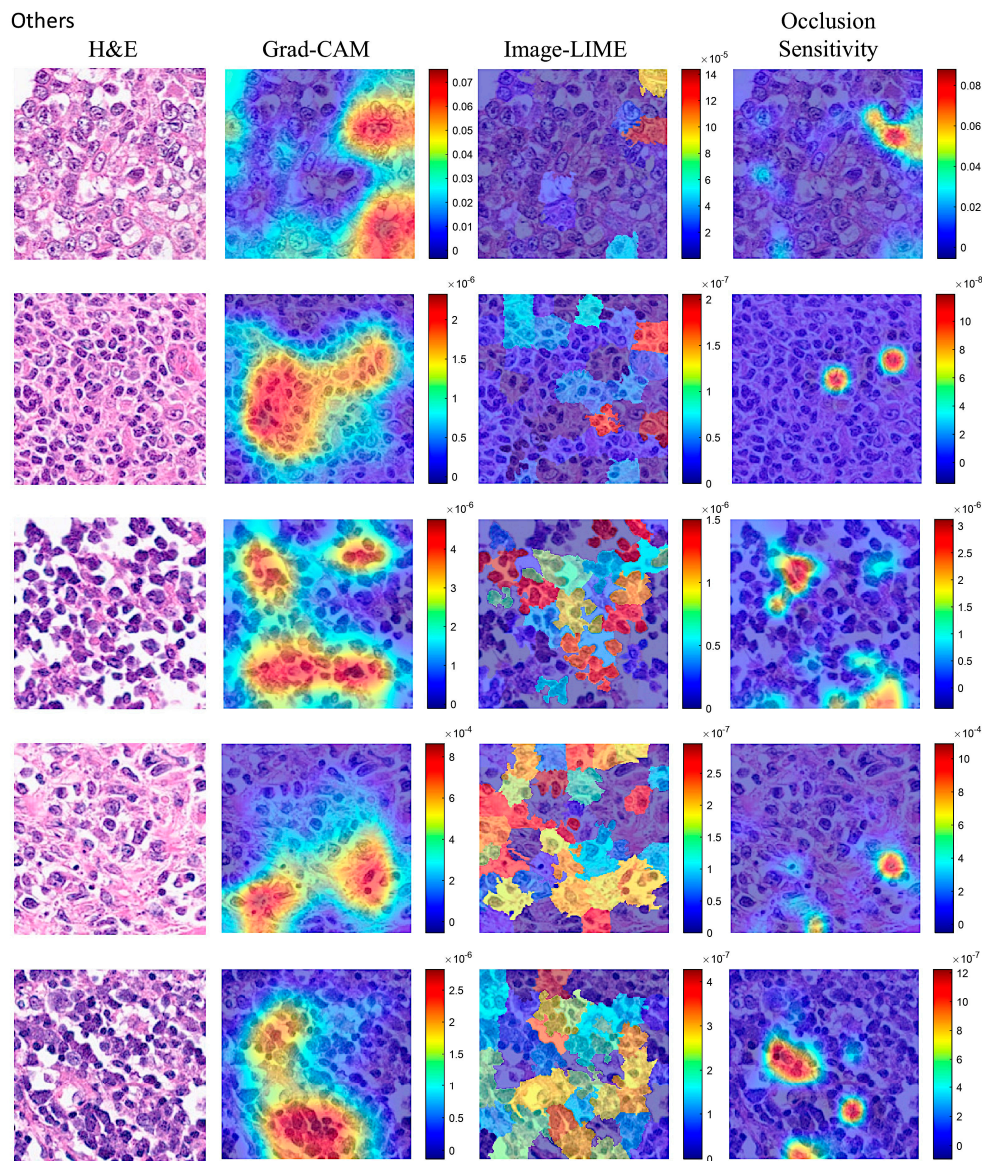


Figure 14. Explainable artificial intelligence (XAI) of the Others group. This figure shows 4 different image patches analyzed with XAI techniques, including Grad-CAM, ImageLIME, and OcclusionSensitivity. XAI techniques were used to identify the areas of an image that the network Darknet-19 used for classification. XAI showed that the convolutional neural networks focused on the epithelial component of the neoplasia.

3.2. Dataset 1: Clinicopathological Characteristics

The Dead < 2 years group was correlated with several clinicopathological characteristics of the patients. The aggressive patients correlated with stage III-IV, International Prognostic Index (IPI) High + High/intermediate, absence of clinical response to treatment, non-GCB cell-of-origin, CD10-, BCL2+, and EBER+. Analysis of the immune microenvironment, cell cycle, and germinal center markers showed that aggressive had higher CD163, IL10, PD-L1, and lower E2F1 protein expressions (all p values < 0.05) (Table 5 and Figures 15 and 16).

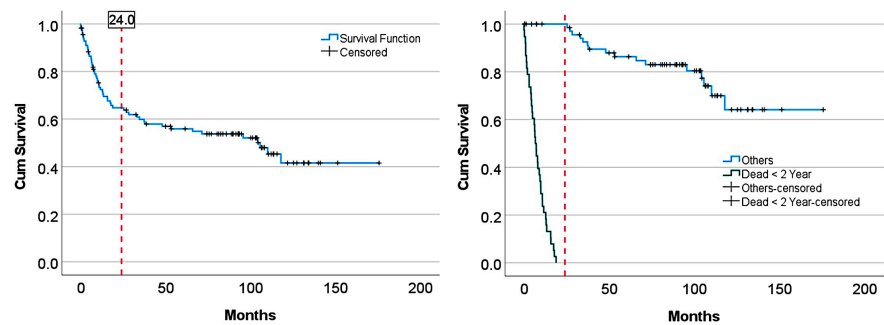


Figure 15. Overall survival of the series. The survival curve showed a point of inflection at 2 years. Two groups were defined: patients who died before 2 years and others. The others included 8 patients who were alive (censored) before the 2 years, and patients who died or were alive after the 2 years. .

Table 5. Correlation with the clinicopathological characteristics.

Variable	All cases	Dead < 2	Others	P
Frequency	114	38/114	76/114	N/A
Clinical characteristics				
Age > 60 years	81/114 (71.1%)	30/38 (78.9%)	51/76 (67.1)	0.273
Male	60/114 (52.6%)	19/38 (50%)	41/76	0.697
Location				
Nodal (+Spleen)	58/114 (50.9%)	16/38 (42.1%)	42/76	0.430
Waldeyer's ring	11/114 (9.6%)	3/38 (7.9%)	8/76 (10.5%)	
Gastrointestinal	13/114 (11.4%)	5/38 (13.2%)	8/76 (10.5%)	
Other extranodal	32/114 (28.1%)	14/38 (36.8%)	18/76	
Stage III-IV	46/97 (47.4%)	18/28 (64.3%)	28/69	0.044
IPI High+High/Intermediate	31/91 (34.1%)	14/27 (51.9%)	17/64	0.029
RCHOP/RCHOP-like treatment	93/98 (94.9%)	26/28 (92.9%)	67/70	0.513
Clinical response	68/92 (73.9%)	5/24 (20.8%)	63/68	<
Hight sIL2R	79/99 (79.8%)	27/29 (93.1%)	52/70	0.052
Pathological characteristics				
CD3+	0/114 (0%)	0/38 (0%)	0/76 (0%)	1.0
CD20+	114/114 (100%)	38/38 (100%)	76/76	1.0
CD5+	13/113 (11.5%)	4/38 (10.5%)	9/75 (12.0%)	1.0
CD10+	33/113 (29.2%)	2/38 (5.3%)	31/75	<
BCL6+	76/113 (67.3%)	26/38 (68.4%)	50/75	1.0
MUM1+	93/113 (82.3%)	33/38 (86.8%)	60/75 (80%)	0.442
Non-GCB	77/114 (67.5%)	35/38 (92.1%)	42/76	<
BCL2+	89/113 (78.8%)	36/38 (94.7%)	53/75	0.003
MYC rearrangement	9/98 (9.2%)	2/29 (6.9%)	7/69 (10.1%)	1.0
EBER+	28/114 (25%)	15/37 (40.5%)	13/75	0.011
Ki67	16.1% +/- 14.2	15.3% +/-	16.5% +/-	0.959
Immune microenvironment				
IL10	12.2% +/- 15.8 (n =	18.6% +/-	9.2% +/-	0.006
PD-L1 (CD274)	12.2% +/- 15.8% (n	18.5% +/-	9.1% +/-	0.026
CSF1R	33.5% +/- 27.5 (n =	28.7% +/-	35.8 % +/-	0.247
CD163	39.2% +/- 25.9 (n =	48.2% +/-	34.6% +/-	0.008
CASP8	6.7% +/- 8.4 (n =	6.0% +/- 9.4	7.1% +/- 8.0	0.268
TNFAIP8	41.3% +/- 25.6 (n	46.2 +/- 24.0	39.3% +/-	0.223
Cell cycle / GC-related				
LMO2	2.6% +/- 3.5 (n =	2.4% +/- 3.9	2.7% +/- 3.4	0.051
MYC	5.4% +/- 5.7 (n =	6.5% +/- 6.4	4.9% +/- 5.5	0.318
MDM2	10.8% +/- 8.1 (n =	9.7% +/- 6.1	11.3% +/-	0.594
CDK6	5.1% +/- 7.4 (n =	3.6% +/- 5.3	5.7% +/- 8.1	0.056
E2F1	1.8% +/- 1.8 (n =	1.2% +/- 0.9	2.0% +/- 1.9	0.020
BCL2	6.8% +/- 9.7 (n =	3.4% +/- 4.5	8.1% +/-	0.087
TP53	5.2% +/- 8.1 (n =	6.6% +/- 10.3	4.6% +/- 7.0	0.128

In this series of 114 cases, 8 cases of Alive < 2 years were included in the Others group.

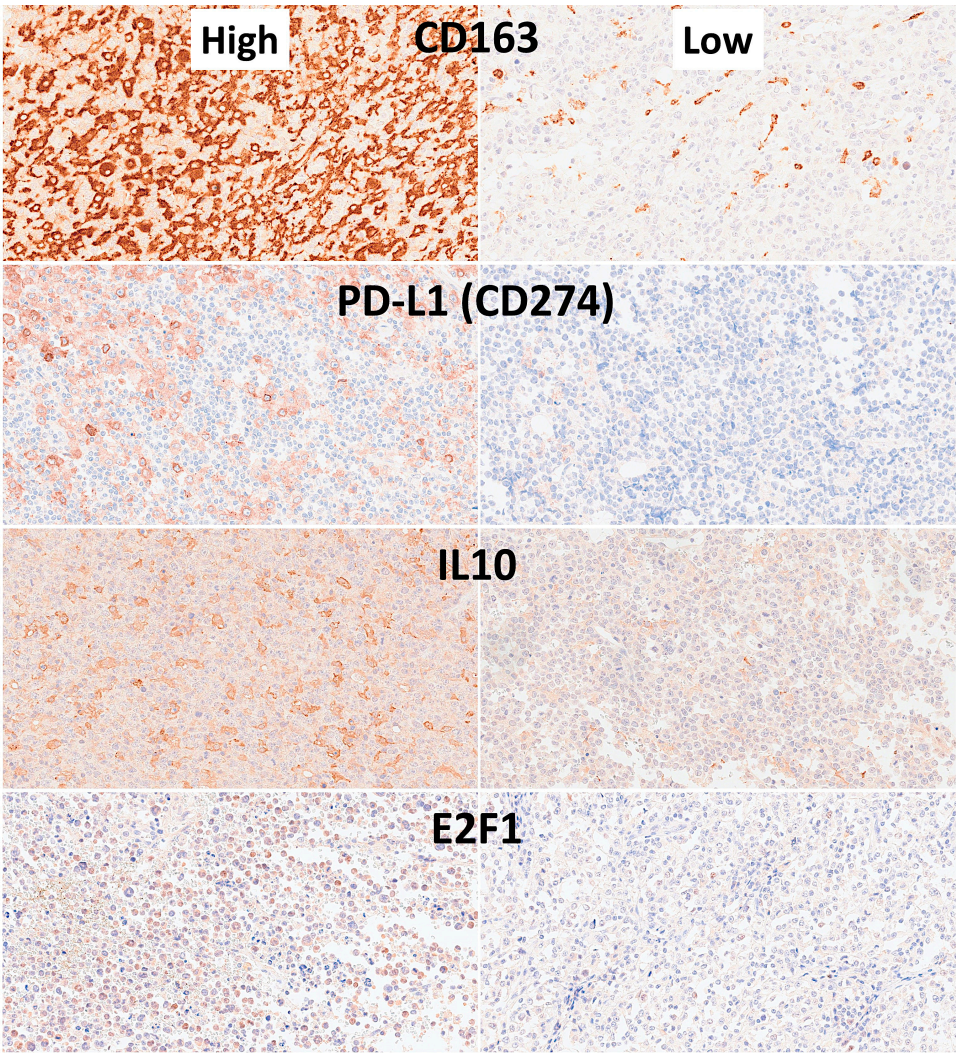


Figure 16. Most relevant immune microenvironment and cell cycle immunohistochemical markers. Analysis of the immune microenvironment, cell cycle, and germinal center markers showed that Dead < 2 years had higher CD163, IL10, and PD-L1, but lower E2F1 (all p values < 0.05).

3.3. Alternative Dataset 2: Exclusion of Alive Cases with Follow-up of Less Than 2 Years

In the original dataset, there were 8 alive cases (censored) with a follow-up of less than 2 years that we included in the “others” group. It may be considered that these cases should be excluded from the series. Therefore, the analysis was repeated with these 8 cases excluded using the DarkNet-19 CNN.

In the design of the DarkNet-19 architecture, changes in the convolution2dLayer number 19 are necessary. The following parameters were used: name (conv19), filter size (1,1), number of filters (2), stride (1,1), dilation factor (1,1), padding (same), padding value (0), weights [], bias [], weight learning rate factor (10), weightL2 factor (1), bias learning rate factor (10), biasL2 factor (0), weights initializer (glorot), and bias initializer (zeros). In the last classification layer, parameter classes and output size are set at (auto).

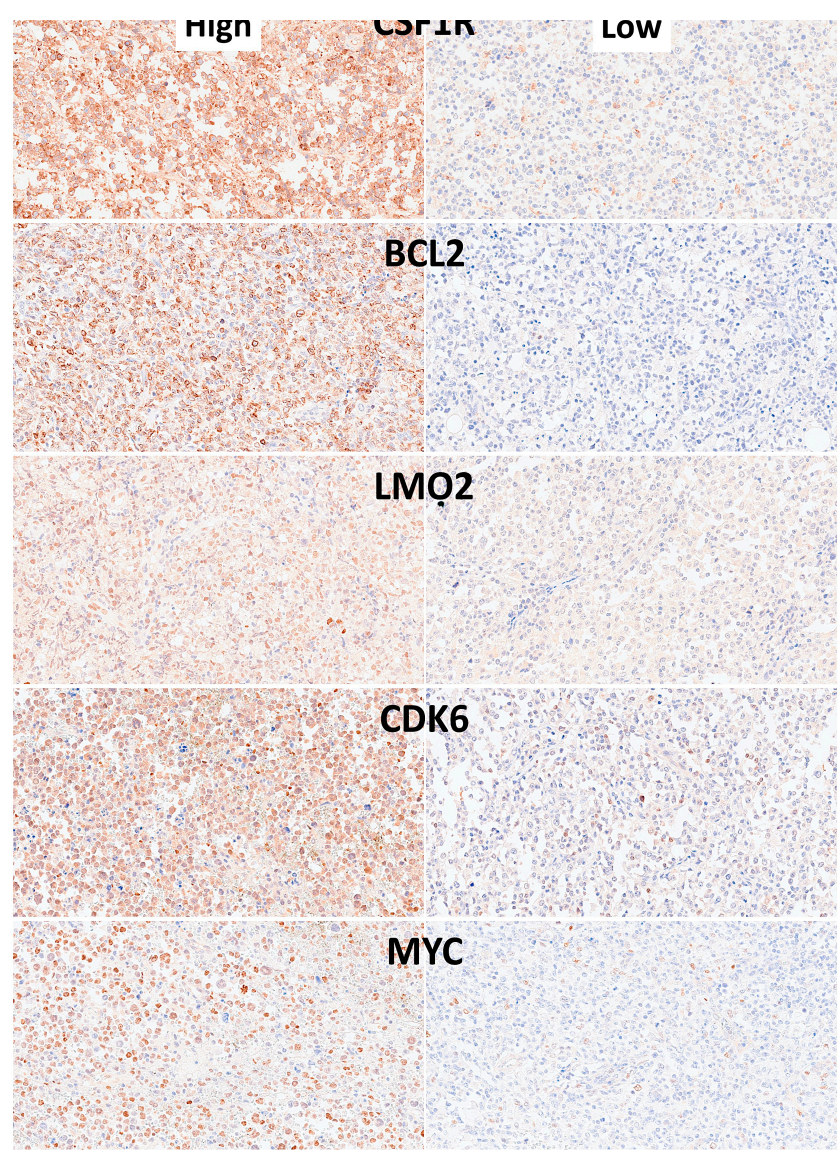


Figure 17. Additional immune microenvironment and cell cycle immunohistochemical markers. For these markers, the analysis of the immune microenvironment, cell cycle, and germinal center markers showed no statistically differences at quantitative level (all p values > 0.05).

The training options were the following: solver (sgdm), initial learning rate (0.001), minibatch size (128), max epochs (5), validation frequency (50). Solver momentum (0.9). Learning rate schedule (none), drop factor (0.1), period (10). Normalization and regularization, L2 regularization (0.0001), reset input normalization (yes), batch normalization statistics (population). Mini-batch shuffle (every-epoch). Validation and output, validation patience (Inf), output network (last-iteration). Gradient clipping, gradient threshold method (l2norm), gradient threshold (Inf). Hardware execution environment (auto). Checkpoint path (), frequency (1), frequency unit (epoch).

The training elapsed 13 min and 6 seconds after the 5 epochs were completed. Training cycle comprised of 1080 iterations, 216 iterations per epoch. The validation accuracy was 95.40%. In the testing set, the accuracy was 95.5%. Other performance parameters were precision 93.34%, recall 93.34%, false positive rate 4.1%, specificity 95.9%, and F1 score 93.34% (Figure 18).

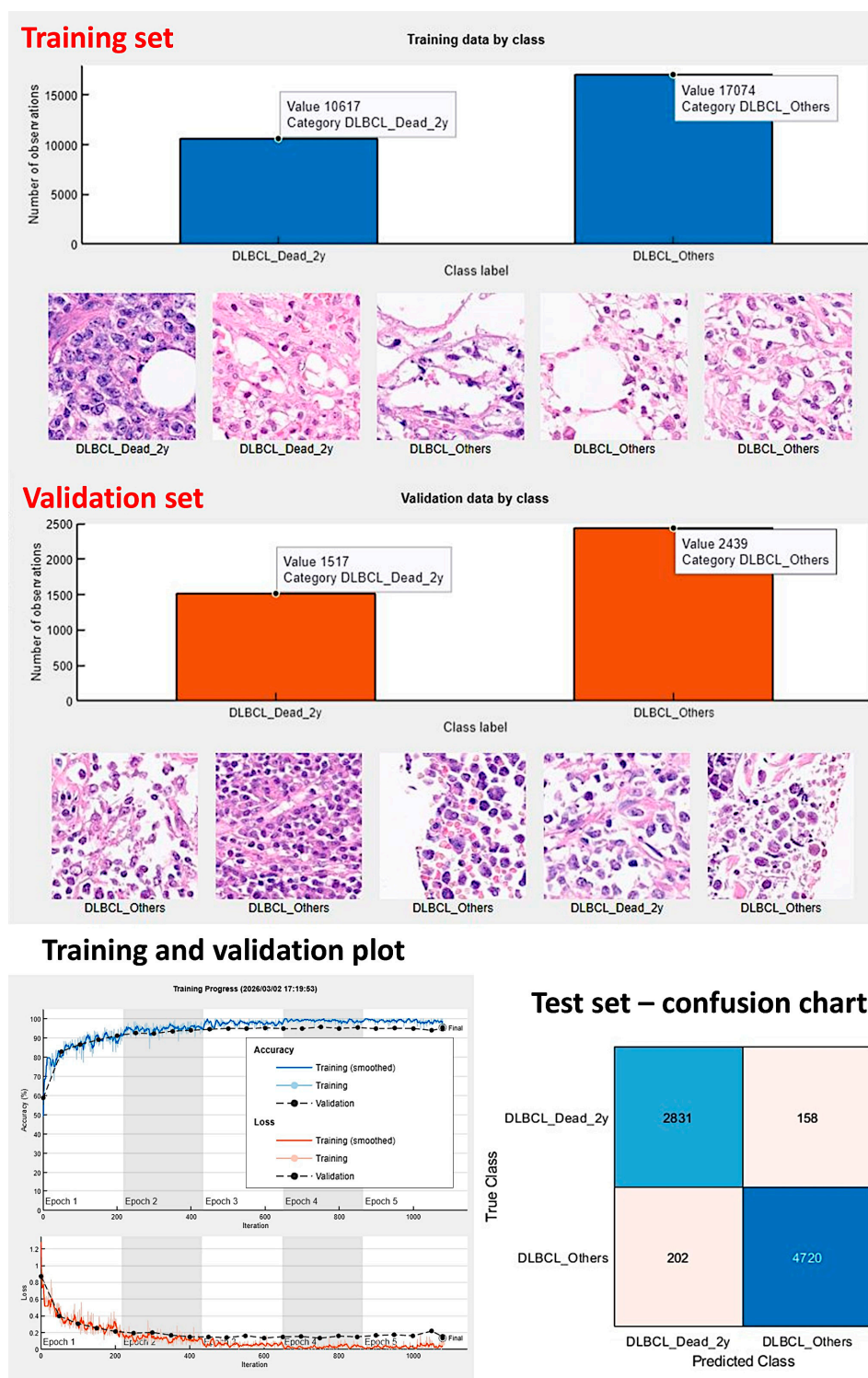


Figure 18. Analysis using modified dataset 1. The analysis was repeated with the exclusion of 8 cases of alive patients that had a follow-up of less than 2 years and were previously included in the “Others” group. This analysis, also based at patch-level because the aim was to create a trained CNN for transfer learning in the future, was performed using the DarkNet-19 architecture. In the test set, the accuracy was 95.5%. Patch-level analysis.

The characteristics of the survival curve are shown in Figure 19.

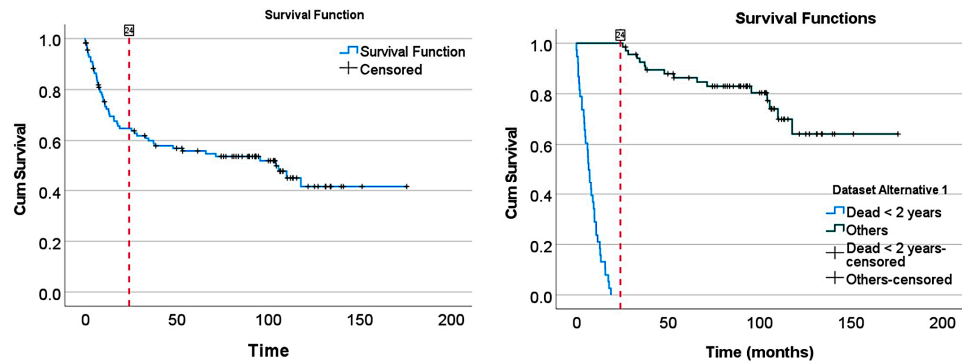


Figure 19. Overall survival of alternative dataset 1. In the original dataset, there were 8 alive cases (censored) with a follow-up of less than 2 years that we included in the “others” group. It may be considered that these cases should be excluded from the series. Therefore, the analysis was repeated with these 8 cases excluded using the DarkNet-19 CNN.

Statistical analyses were performed to describe the clinicopathological characteristics of the series as well as to identify what variables were different between the 2 groups. In comparison to Others (i.e. Dead or Alive > 2 years), the Dead < 2 years patients were characterized by IPI High/High-Intermediate, absence of clinical response, higher sIL2R, and Epstein-Barr virus (EBER) positivity; the phenotype was CD10 negative, BCL2 positive, and non-GCB according to the cell-of-origin surrogate Hans’classifier; and the immune-oncology markers of high IL10, PD-L1 and CD163, but lower E2F1 (Table 6).

Table 6. Correlation with the clinicopathological characteristics (Alive < 2 years excluded).

Variable	All cases	Dead < 2	Others	P
Frequency	106	38/106	68/106	N/A
Clinical characteristics				
Age > 60 years	75/106 (70.8%)	30/38 (78.9%)	45/68	0.188
Male	56/106 (52.8%)	19/38 (50%)	37/68	0.689
Location				
Nodal (+Spleen)	54/106 (50.9%)	16/38 (42.1%)	38/68	0.430
Waldeyer’s ring	10/106 (9.4%)	3/38 (7.9%)	7/68 (10.3%)	
Gastrointestinal	12/106 (11.3%)	5/38 (13.2%)	7/68 (10.3%)	
Other extranodal	30/106 (28.3%)	14/38 (36.8%)	16/68	
Stage III-IV	44/90 (48.9%)	18/28 (64.3%)	26/62	0.069
IPI High+High/Intermediate	30/85 (35.3%)	14/27 (51.9%)	16/58	0.050
RCHOP/RCHOP-like treatment	88/92 (95.6%)	26/28 (92.9%)	62/64	0.597
Clinical response	65/87 (74.7%)	5/24 (20.8%)	60/63	<
Hight sIL2R	75/93 (80.6%)	27/29 (93.1%)	48/64 (75%)	0.049
Pathological characteristics				
CD3+	0/106 (0%)	0/38 (0%)	0/68 (0%)	1.0
CD20+	106/106 (100%)	38/38 (100%)	68/68	1.0
CD5+	13/105 (12.4%)	4/38 (10.5%)	9/67 (13.4%)	0.766
CD10+	29/105 (27.6%)	2/38 (5.3%)	27/67	<
BCL6+	70/105 (66.7%)	26/38 (68.4%)	44/67	0.832
MUM1+	85/105 (81.0%)			
Non-GCB	73/104 (70.2%)	35/38 (92.1%)	38/66	<
BCL2+	84/105 (80.0%)	36/38 (94.7%)	48/67	0.005
MYC rearrangement	9/92 (9.8%)			
EBER+	25/104 (24.0%)	15/37 (40.5%)	10/67	0.007
Ki67	15.9% +/- 14.5	14.9% +/-	16.4% +/-	0.935
Immune microenvironment				
IL10	9.7% +/- 12.2	14.2% +/-	7.3% +/- 9.9	0.010
PD-L1 (CD274)	12.3% +/- 15.9	18.5% +/-	8.9% +/-	0.027
CSF1R	33.7% +/- 27.0	28.1% +/-	36.4% +/-	0.220

CD163	39.9% +/- 25.9	48.2% +/-	35.4% +/-	0.014
CASP8	7.1% +/- 8.6	5.8% +/- 9.2	7.6% +/- 8.3	0.139
TNFAIP8	42.3% +/- 25.9	46.2% +/-	40.5% +/-	0.336
Cell cycle / GC-related				
LMO2	2.6% +/- 3.6%	2.3% +/- 3.9	2.7% +/- 3.4	0.063
MYC	5.4% +/- 5.8	6.3% +/- 6.4	4.9% +/- 5.5	0.304
MDM2	10.9% +/- 8.2	9.4% +/- 6.2	11.5% +/-	0.542
CDK6	4.7% +/- 6.8	3.5% +/- 5.3	5.2% +/- 7.4	0.088
E2F1	1.8% +/- 1.8	1.1% +/- 0.9	2.1% +/- 2.0	0.014
BCL2	7.1% +/- 9.9	3.5% +/- 4.4	8.8% +/-	0.037
TP53	5.4% +/- 8.4	6.6% +/- 10.3	4.8% +/- 7.4	0.108

In this series, the 8 cases of Alive < 2 years were excluded.

3.4. Alternative Dataset 2: Exclusion of Alive DLBCL Cases with a Follow-up of Less than 2 years and Hybrid Partitioning to Avoid Information Leakage

Hybrid partitioning was performed to confirm if information leakage was present in the classification model. Ten random cases, 5 of Dead < 2 years and 5 of Others were selected as test set number 2. The training/validation was performed at the patch level. The test set 1 was also at the patch level. However, the test set 2 was performed at the patient-level. At patient-level analysis, ROC analysis was used to determine the best cutoff to differentiate between the 2 groups. In this analysis, the 8 cases of alive DLBCL before 2 years were excluded. Data is shown in Figure 20 and 21, and Table 7.

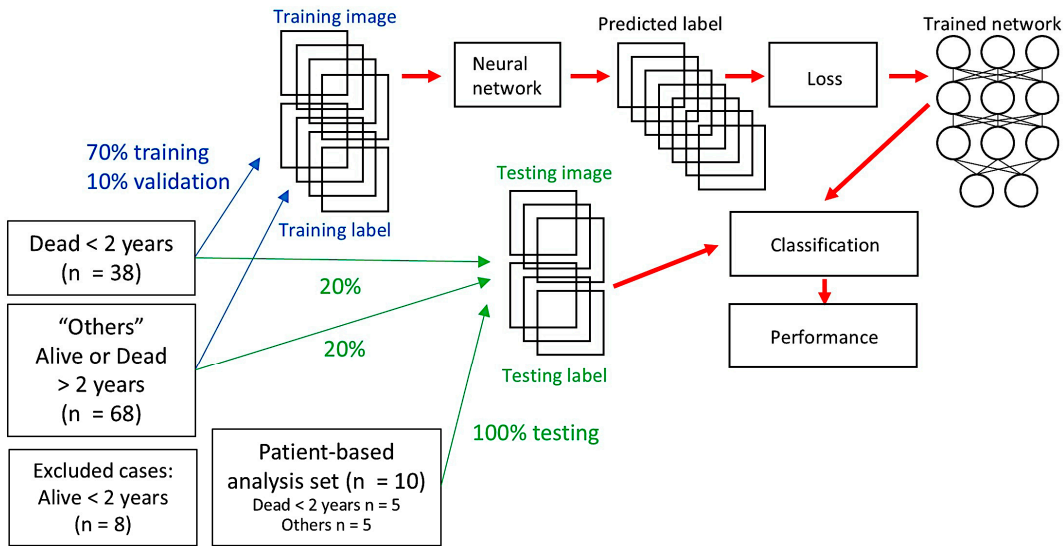


Figure 20. Hybrid partitioning. This hybrid partitioning strategy aimed to prevent information leakage. Initial analysis was patched-based and used 38 cases of Dead < 2 years and 68 cases of “Others”. The “Others” group was composed of Alive or Dead > 2 years patients. Eight alive cases < 2 years were excluded from the series. The test set 2 comprised 10 cases and the analysis was patient-based.

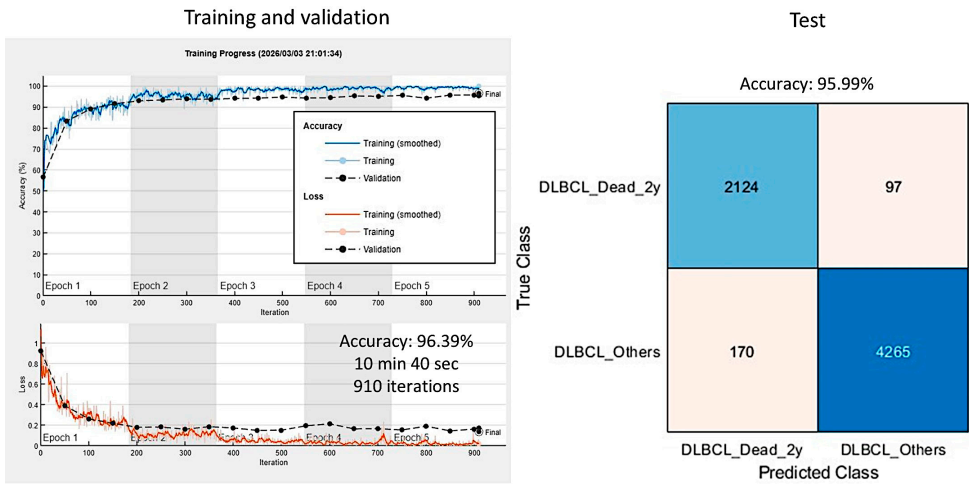


Figure 21. Hybrid partitioning strategy. This Figure shows the training/validation and testing of the DarkNet-19 CNN used in hybrid partitioning, test set 1 (image-patch-based). Ten random cases, 5 of Dead < 2 years and 5 of Others were selected as test set number 2. The training/validation was performed at the patch level. The test set 1 was also at the patch level. However, test set 2 was performed at the patient-level (Table 7).

Table 7. Patient-level analysis in hybrid partitioning.

Case	1	2	3	4	5	6	7	8	9	1
Orig	Ot	Ot	Ot	Ot	Ot	D	D2	D2	D	D
AI	Ot	Ot	D2	Ot	Ot	D	Ot	Ot	D	D
AI	Ot	Ot	D2	Ot	Ot	D	D2	Ot	D	D
Other	380	352	62	804	681	1	905	844	7	3
Other	94.	93.	21.	96.	99.	6.	60.	96.	1	4
Dead	23	23	222	28	4	2	599	30	2	4
Dead	5.7	6.1	78.	2.4	0.6	9	39.	3.4	8	5
All	403	375	284	832	685	2	150	874	3	7
All	100	100	100	100	100	1	100	100	1	1

The standard cutoff for group classification is 50%. By ROC analysis, the best cutoff to differentiate between the 2 groups was 77%. D2Y, Dead < 2 years. Others, Dead or Alive > 2 years.

3.5. Alternative Dataset 3: Exclusion of Alive DLBCL Cases with a Follow-up of Less than 2 Years and Addition of the Third Group of Reactive Lymphoid Tissue

Classification with reactive lymphoid tissue is also important for differential diagnosis. Therefore, an updated series included 44 reactive lymphoid tissue, 38 DLBCL cases in the group “Dead < 2 years”, and 68 DLBCL cases in the group “Other” (which included both Dead and Alive > 2 years) (Figure 22). Of note, the 8 cases of Alive < 2 years were excluded as in section 3.3. At patch level, the accuracy was 99.7% (Figure 23).

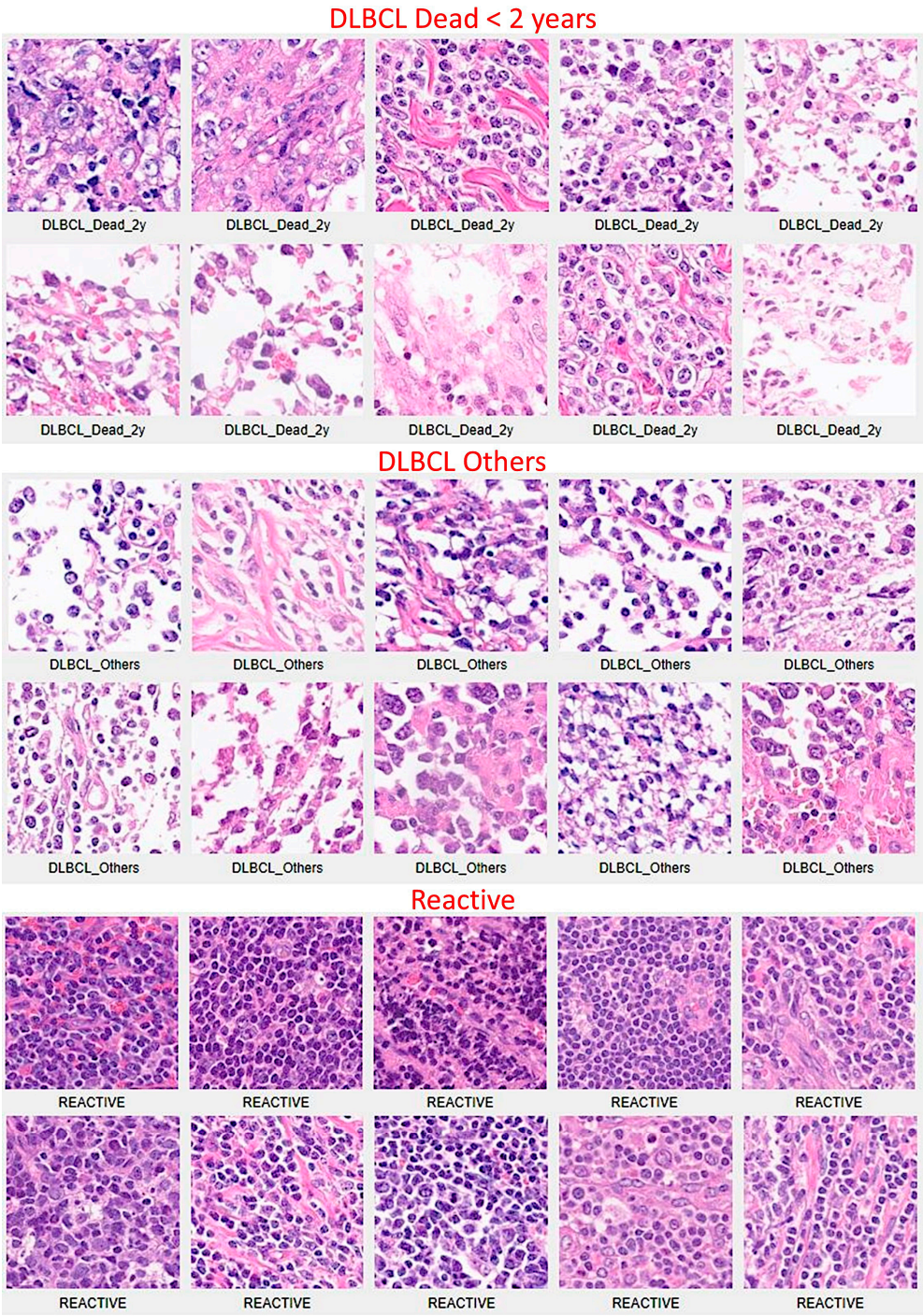


Figure 22. Image patches of DLBCL and reactive lymphoid tissue. The series included 44 cases of reactive lymphoid tissue to expand the classification into three different diagnostic groups: Dead < 2 years, Others, and Reactive lymphoid tissue. The 8 cases of Alive < 2 years were excluded from the analysis.

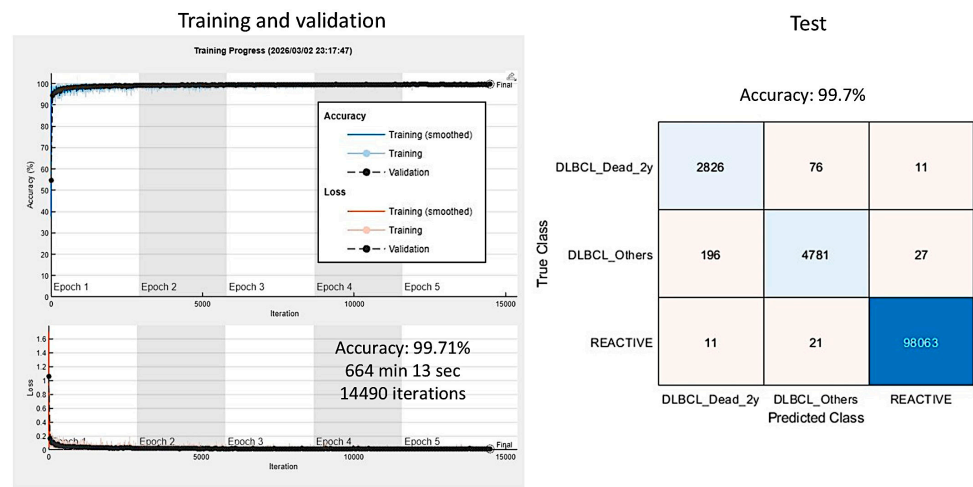


Figure 23. Training progress and test set performance. The accuracy of the training/validation set was 99.71%. It consisted of 14490 iterations, 5 epochs, and lasted for 664 min and 13 seconds. In the test set, the accuracy was 99.7%. Patch-level analysis to classify between DLBCL dead < 2 years, DLBCL Others, and reactive lymphoid tissue.

Further analysis was performed to train the network to classify between DLBCL and reactive lymphoid tissue using a hybrid partitioning. Ten aleatory cases of DLBCL and 5 cases of reactive lymphoid tissue were excluded from the main series and analyzed at patient-level as test set 2. The rest of the cases were analyzed at patch-level and the series was split into a training set (70%) and a validation set (10%) to help prevent over-fitting during training and a test set 1 (20%).

The training elapsed for 179 min and 24 sec. The training cycle included 5 epochs, with 1983 iterations per epoch, and a total of 9915 iterations. The validation accuracy was 99.83%. In the test set 1, the accuracy was 95.99%. This analysis was patch-based (Figure 24).

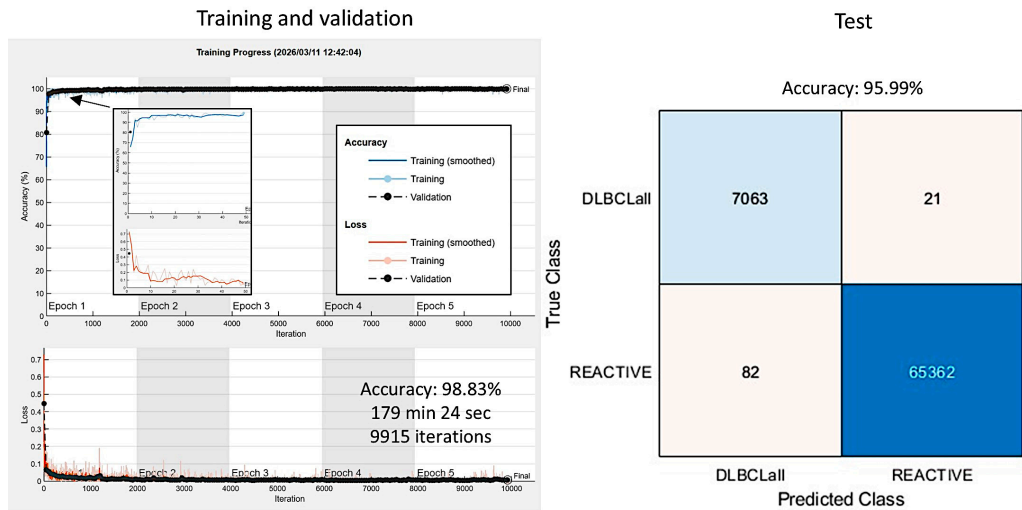


Figure 24. Training progress and test set performance. The accuracy of the training/validation set was 98.83%. It consisted of 9915 iterations, 5 epochs, and lasted for 179 min and 24 seconds. In the test set, the accuracy was 95.99%. Patch-level analysis to classify between DLBCL and reactive lymphoid tissue.

The test set 2 analysis was performed at the patient-level, including 10 cases of DLBCL and 5 cases of reactive lymphoid tissue. All cases but one (90%) of DLBCL cases were correctly classified as DLBCL. All cases of reactive lymphoid tissue were correctly classified (100%). This analysis was designed to avoid information leakage (Table 8).

Table 8. Patient-level analysis in hybrid partitioning.

Case	D1	D2	D3	D4	D5	D6	D7	D8
Origin	DLB	DLB	DLB	DLB	DLB	DLB	DLB	DLB
AI	DLB	DLB	DLB	DLB	DLB	DLB	DLB	React
DLBC	208	710	1498	305	472	284	719	303
DLBC	98.6	96.3	99.6	91.3	79.6	100%	85.7	41.9
Reacti	3	27	6	29	121	0	120	420
Reacti	1.4%	3.7%	0.4%	8.7%	20.4	0%	14.3	58.1
All	211	737	1504	334	593	284	839	723
All	100	100	100	100	100	100	100	100

Case	D9	D10	R1	R2	R3	R4	R5
Origin	DLB	DLB	React	React	React	React	React
AI	DLB	DLB	React	React	React	React	React
DLBC	678	546	7	15	2	3	55
DLBC	98.9	66.0	0.1%	0.03	0.01	0.007	2.59
Reacti	7	281	3872	5972	1715	4583	2069
Reacti	1.1%	34.0	99.9	99.97	99.99	99.99	97.41
All	685	827	3872	5974	1715	4583	2124
All	100	100	100	100	100	100	100

The standard cutoff for group classification is 50%. D, DLBCL. Reactive, reactive lymphoid tissue.

Finally analyses of groups Dead < 2 years and Dead > 2 years and hyperparameter tunning are shown in Appendix C and D. Appendix E shows hyperparameter tuning in DLBCL vs. reactive lymphoid tissue. Appendix F shows the architecture of DarkNet-19.

4. Discussion

DLBCL is one of the most common diagnostic categories of non-Hodgkin lymphoma (NHL), accounting for approximately 25% of NHL cases in the developed world [2,5]. Histologically, it is characterized by large transformed B cells that depict a diffuse growth pattern and efface the normal architecture of the underlying histological structure [2,5]. The diagnostic category of DLBCL, NOS, refers to conventional DLBCL with a mature B-cell phenotype and large cell morphology, but lacking none of the criteria that define specific large B cell lymphoma subtypes (i.e., other large B cell variants) [2,5,10].

This study used a series of 114 cases of conventional large B cell lymphoma. Overall, it included not only NOS cases but also 28 cases of EBER-positive DLBCL. Most of the cases were nodal (58/114, 50.9%), followed by other extranodal (28.1%) and gastrointestinal (11.4%). The aim of the project was to identify cases with poor prognosis using only H&E staining and CNN. After the evaluation of the overall survival curve, 2 groups were defined: patients who died before 2 years and others. Correlation with the clinicopathological characteristics found that the a < 2 years group was correlated with stage III-IV, International Prognostic Index (IPI) High + High/intermediate, progressive disease, non-GCB cell-of-origin, CD10-, BCL2+, and EBER+. Analysis of the immune microenvironment, cell cycle, and germinal center markers showed that Dead < 2 years had higher IL10, CD163, PD-L1, and lower E2F1 expressions. Therefore, the CNN managed to identify the histological features that correlated with the prognosis of the patients. Features that were not very obvious under conventional histological examination under an optical microscope.

The deep learning workflow includes preprocessing data, importing and building the network, training the network, tuning the network, and visualizing the results. In this study, we used transfer learning to take advantage of the knowledge provided by a pre-trained network to learn new patterns in new data. Fine-tuning a pretrained network with transfer learning is typically much faster and easier than training from the beginning. The reuse of the pretrained network includes the following steps: load the pretrained network, replace the final layers, train the network, predict and assess the network accuracy, and deploy the results. This study used 20 pretrained networks, including AlexNet, DarkNet-19, DarkNet-53, DenseNet-201, EfficientNet-b0, GoogleLeNet, GoogleLeNet-places365, Inception-ResNet-v2, Inception-v3, MobileNet-v2, NasNet-Large, NasNet-Mobile, ResNet-101,

ResNet-18, ResNet-50, Shufflenet, SqueezeNet, VGG-16, VGG-19, and Xception. The CNN architectures were different and the performances ranged from 79.16% of SqueezeNet to 96.26% of DarkNet-19. The training times differed as well, being 1429 min for NasNet-Large and 2 min for AlexNet. The final analysis was performed using DarkNet-19. The final model based on DarkNet-19 predicted prognosis groups with high performance (test set accuracy = 96.26%). The other performance parameters were precision (94.46%), recall (95.02%), false positive rate (3.07%), specificity (96.93%), and F1 score (94.74%).

A machine learning model is often referred to as a "black box" model because understanding how the model makes predictions can be difficult. Interpretability tools help you overcome this aspect of machine learning algorithms and reveal how predictors contribute (or do not contribute) to predictions. Moreover, you can validate whether the model uses the correct evidence for its predictions and find model biases that are not immediately apparent. In this study, explainable artificial intelligence (XAI) was performed using Grad-CAM, occlusion sensitivity and image Lime on the network DarkNet-19. Overall, XAI showed that the CNNs focused on the neoplastic B lymphocytes, but some components of the microenvironment may also have played an influence. Of note, current XAI techniques do not allow a more detailed analysis of this type of histological tissue.

In conclusion, narrow artificial intelligence (i.e., trained to perform a specific or a set of closely related tasks) can predict the prognosis of DLBCL based on the computer vision CNN histological analysis of H&E images, but it is process-specific and operates within limited constraints.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org.

Author Contributions: All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded to J.C. by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) and the Japan Society for the Promotion of Science (JSPS), grant numbers KAKEN 15K19061, 18K15100, and 23K06454.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of TOKAI UNIVERSITY, SCHOOL OF MEDICINE (protocol codes IRB20-156).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: All data is available upon request to Joaquim Carreras (joaquim.carreras@tokai.ac.jp) and also shown as supplementary file.

Acknowledgments: N/A.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CNIO	Spanish National Cancer Research Center
CNN	Convolutional neural network
DLBCL	Diffuse large B-cell lymphoma
H&E	Hematoxylin and eosin
NHL	Non-Hodgkin lymphoma
XAI	Explainable artificial intelligence

Appendix A

Useful MATLAB R2023b Update 10 (23.2.0.2859533) code:

%L1 is image 1


```

%% Resize to 299x299x3 because trainedNetwork is Inceptionv3
B = imresize (L1, [299 NaN])

%% Check size of image
sizesz = size(B)

%% Grad-CAM
imds = B
label = classify(trainedNetwork_1,imds)
scoreMap = gradCAM(trainedNetwork_1,imds,label);
figure
imshow(imds)
hold on
imagesc(scoreMap,AlphaData,0.5)
colormap jet
colorbar

%% imageLIME
label = classify(trainedNetwork_1,B)
scoreMap = imageLIME(trainedNetwork_1,B,label);
figure
imshow(B)
hold on
imagesc(scoreMap,AlphaData,0.5)
colormap jet
colorbar

%% OcclusionSensitivity
label = classify(trainedNetwork_1,B)
scoreMap = occlusionSensitivity(trainedNetwork_1,B,label);
figure
imshow(B)
hold on
imagesc(scoreMap,AlphaData,0.5)
colormap jet
colorbar

%% To classify and show image
net = trainedNetwork_1
classes = net.Layers(end).Classes;
[YPred, scores] = classify(net, B);
[score, classIdx] = max(scores);

predClass = classes (classIdx);

imshow(B);
title(sprintf("%s (%.2f)",string(predClass),score));

%% To display an image
imshow(B)

%%
imageLoc = "1104802"
imds = imageDatastore(imageLoc)
imds2 = augmentedImageDatastore([299 299 3], imds);
YPred = classify(trainedNetwork_1,imds2)
scores = predict(trainedNetwork_1,imds2)

```

Appendix B

Immunohistochemistry was performed using Leica Bond-Max system and reagents following the manufacturer's instruction. Slides were digitalized using a NanoZoomer S360 digital slide scanner (Hamamatsu Photonics). Digital image quantification was performed using Fiji software (ImageJ). The primary antibodies and the staining conditions for the target markers were the following:

Supplementary Table B. Immunohistochemical markers and primary antibodies

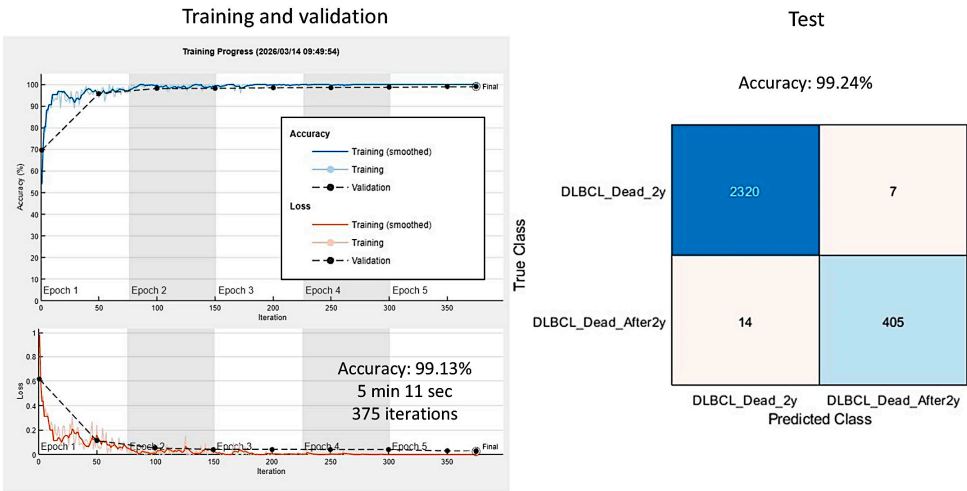
Marker	Target	Clone	Company
CD3	T lymphocytes	LN10	Novocastra
CD20	B lymphocytes	L26	Novocastra
CD5	T lymphocytes	4C7	Novocastra
CD10	Germinal center	56C6	Novocastra
BCL6	Germinal center	LN22	Novocastra
MUM1 / IRF4	Plasma cell differentiation	EAU32	Novocastra
BCL2	Apoptosis	Bcl2/10/D5	Novocastra
EBER	EBV-encoded mRNA	BP0589/AR0833	Novocastra
Ki67	Cell proliferation	MM1	Novocastra
IL10	Immuno-oncology	LS-B7432	Lifespan
PD-L1	Immuno-oncology	E1J2	Cell Signaling
CSF1R	Immuno-oncology	FER216	CNIO
CD163	Tumor-associated	10D6	Novocastra
CASP8	Active subunit p18	11B6	Novocastra
TNFAIP8	Apoptosis	14559-MM0	Sino Biological
LMO2	Hematopoietic development	299B	CNIO
MYC	Proto-oncogene	Y69	Abcam
MDM2	Proto-oncogene	IF2	Invitrogen
CDK6	Cell cycle	98D	CNIO
E2F1	Cell cycle	Agro368V	CNIO
TP53	Cell regulation	DO-7	Novocastra

CNIO, Spanish National Cancer Research Center (CNIO), Madrid, Spain.

Appendix C Dead <2 Years vs. > 2 Years

Same type of partition analysis was performed to classify and compare the cases of Dead < 2 years with Dead > 2 years. The analysis included the training, validation, test1 and test 2 sets following the hybrid partitioning strategy.

Figure Appendix C1 shows the results of the part of the analysis that is made using image patches.



Appendix Figure C. Image classification between Dead < 2 years and Dead > 2 years.

The accuracy of the training/validation was 99.13% and the test set 1 accuracy was 99.24%. Analysis at patient level was also performed with 5 cases of Dead < 2 years and 5 cases of Dead > 2 years as test set 2. At patient level, all 5 cases of Dead < 2 years were correctly classified, but cases Dead > 2 years not (due to few numbers of cases in the group).

Appendix Table C shows the clinicopathological characteristics and differences between the two groups. In comparison to Dead > 2 years, Dead < 2 years was characterized by less clinical response, less CD10 and BCL2 protein expressions and non-GCB phenotype (all p values < 0.05).

Appendix Table C. Correlation with the clinicopathological characteristics.

Variable	All dead cases	Dead < 2	Dead > 2 years	P value
Frequency	54/54 (100%)	38/54 (70.4%)	16/54 (29.5%)	N/A
Clinical characteristics				
Age > 60 years	45/54 (83.3%)	30/38 (78.9%)	15/16 (93.8%)	0.253
Male	28/54 (51.9%)	19/38 (50%)	9/16 (56.3%)	0.770
Location				
Nodal (+Spleen)	24/54 (44.4%)	16/38 (42.1%)	8/16 (50%)	0.486
Waldeyer's ring	4/54 (7.4%)	3/38 (7.9%)	1/16 (6.3%)	
Gastrointestinal	5/54 (48.8%)	5/38 (13.2%)	0/16 (0%)	
Other extranodal	21/54 (38.9%)	14/38 (36.8%)	7/16 (43.8%)	
Stage III-IV		18/28 (64.3%)	6/14 (42.9%)	0.208
IPI High+High/Intermediate	20/41 (48.8%)	14/27 (51.9%)	6/14 (42.9%)	0.744
RCHOP/RCHOP-like treatment		26/28 (92.9%)	13/14 (92.9%)	0.891
Clinical response	39/54 (72.2%)	5/24 (20.8%)	12/14 (85.7%)	< 0.001
Hight sIL2R		27/29 (93.1%)	14/15 (93.3%)	1.000
Pathological characteristics				
CD3+	0/54 (0%)	0/38 (0%)	0/16 (0%)	1.000
CD20+	54/54 (100%)	38/38 (100%)	16/16 (100%)	1.000
CD5+	7/54 (13.0%)	4/38 (10.5%)	3/16 (18.8%)	0.410
CD10+	7/54 (13.0%)	2/38 (5.3%)	5/16 (31.3%)	0.019
BCL6+	37/54 (68.5%)	26/38 (68.4%)	11/16 (68.8%)	1.000
MUM1+	47/54 (87.0%)	33/38 (86.8%)	14/16 (87.5%)	1.000
Non-GCB	46/54 (85.2%)	35/38 (92.1%)	11/16 (68.8%)	0.041
BCL2+	48/54 (88.9%)	36/38 (94.7%)	12/16 (75.0%)	0.056
MYC rearrangement	3/45 (6.7%)	2/29 (6.9%)	1/16 (6.3%)	1.000
EBER+	19/52 (36.5%)	15/37 (40.5%)	4/15 (26.7%)	0.526
Ki67	15.8% +/- 12.9	15.7% +/- 12.4	16.2% +/- 14.5	0.919
Immune microenvironment				
IL10	11.7% +/- 14.4	14.7% +/- 16.1	5.2% +/- 5.9	0.081
PD-L1 (CD274)	13.5% +/- 16.1	16.2% +/- 18.2	7.8% +/- 7.8	0.566
CSF1R	29.7% +/- 25.0	27.6% +/- 25.3	34.1% +/- 24.9	0.405
CD163	39.4 +/- 24.9	43.5% +/- 23.2	30.6% +/- 27.2	0.058
CASP8	5.3% +/- 8.1%	6.3% +/- 9.5	3.2% +/- 3.5	0.738
TNFAIP8	47.2% +/- 25.5	46.6% +/- 24.5	48.5% +/- 28.6	0.685
Cell cycle / GC-related				
LMO2	2.2% +/- 3.5	2.3% +/- 3.9	1.9% +/- 2.2	0.447
MYC	5.3% +/- 5.8	6.5% +/- 6.4	2.9% +/- 3.2	0.080
MDM2	10.0% +/- 6.8	9.7% +/- 6.1	10.6% +/- 8.2	0.964
CDK6	3.8% +/- 5.1	3.6% +/- 5.3	4.1% +/- 4.9	0.178
E2F1	1.3% +/- 1.0	1.2% +/- 0.9	1.7% +/- 1.2	0.210
BCL2	5.5% +/- 7.4	3.4% +/- 4.5	10.2% +/- 10.3	0.022
TP53	5.7% +/- 9.0	6.9% +/- 10.6	3.2% +/- 3.3	0.247

In this series, the 8 cases of Alive < 2 years were excluded.

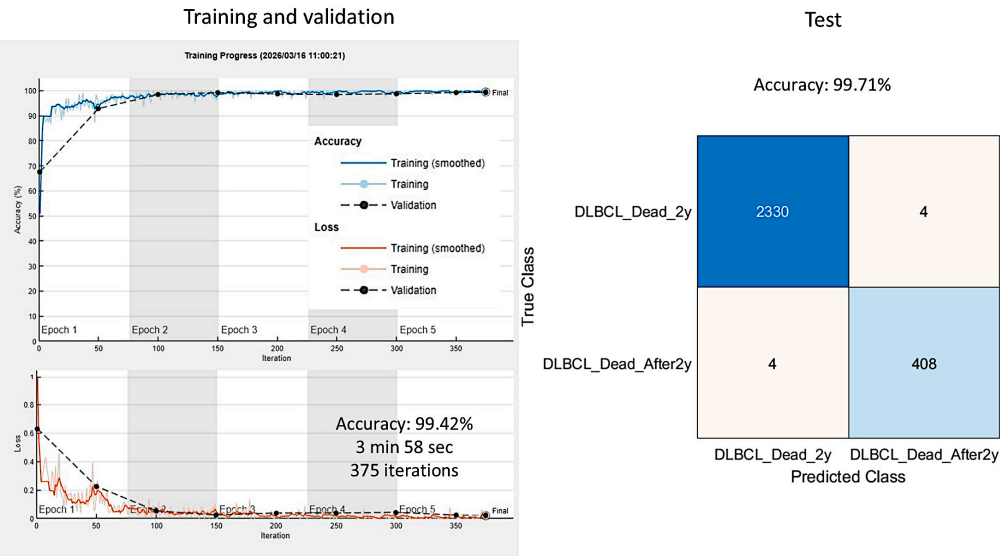
Appendix D Parameter Tuning in Dead <2 Years vs. > 2 Years

The analysis of Appendix C was further developed with parameter tuning.

First, data augmentation for the training set was setup using random reflection x and y axis, leaving the standard training options such as sgdm solver, initial learning rate 0.001, minibatch size 128, maxepochs 5, and validation frequency 50.

The training and validation lasted for 3 min 58 sec and the validation accuracy was 99.42% at image patch-level. In comparison to previous analysis shown in Appendix C, the accuracy increased slightly. At patient level, all 5 cases of Dead < 2 years were correctly classified. However, Dead > 2 years were not; this may be due to the small number of samples in that group. Appendix Figure D shows the results of the training/validation and test 1 (patch-level analysis). Appendix Table 1 and

Figure D2 show the exhaustive sweep hyperparameter tuning. Bayesian optimization is shown in Figure D3.

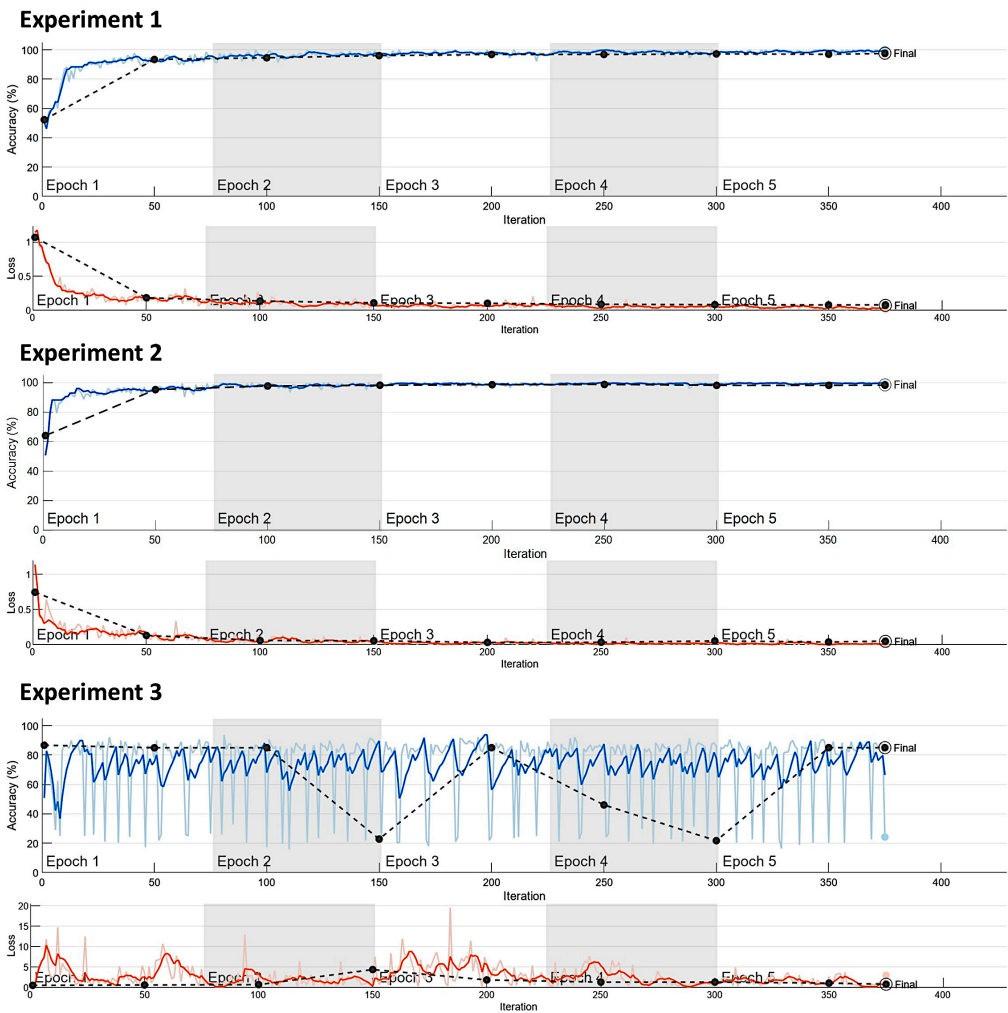


Appendix Figure D1. Training/validation and test 1 using training data augmentation of random reflection x and y axis.

Appendix D Appendix 1. Exhaustive sweep hyperparameter strategy.

Experi ment	Exhaus tive	Swee p	Hyperpara meters	Metri cs			
Experi ment 1	Status	Elap sed	Initial Learning	Traini ng	Train ing	Valida tion	Valida tion
1	Max Epochs	4 min	0.0001	100.00 00	0.015 1	97.669 3	0.0735
2	Max Epochs	4 min	0.0010	99.218 8	0.015 9	98.543 3	0.0478
3	Max Epochs	4 min	0.0100	24.218 8	3.012 9	84.996 4	0.8352

The hyperparameter learning rate was tuned to train networks under different training conditions and compare the results. For example, sweep through a range of hyperparameter values.



Appendix Figure D2. Exhaustive sweep hyperparameter strategy. The hyperparameter learning rate was tuned to train networks under different training conditions and compare the results. For example, sweep through a range of hyperparameter values.

Elapsed Time	Hyperparameters			Metrics				
	myInitialLearnRate	Momentum	L2Regularization	Training Accuracy (%)	Training Loss	Validation Accuracy (%)	Validation Loss	ErrorRate
0 hr 4 min 48 sec	0.0025	0.9088	0.0054	99.2188	0.0140	99.4902	0.0194	0.0000
0 hr 4 min 51 sec	0.0027	0.8142	0.0030	100.0000	0.0099	98.9075	0.0278	0.0000
0 hr 4 min 49 sec	0.0037	0.9243	0.0056	98.4375	0.0607	98.8347	0.0360	0.0000
0 hr 4 min 51 sec	0.0035	0.9757	0.0057	99.2188	0.0237	98.7618	0.0344	0.0000
0 hr 4 min 51 sec	0.0003	0.8038	0.0002	100.0000	0.0080	98.6890	0.0478	0.0000
0 hr 4 min 50 sec	0.0032	0.8113	0.0085	98.4375	0.0442	98.5433	0.0345	0.0000
0 hr 4 min 50 sec	0.0003	0.9783	0.0003	100.0000	0.0087	98.4705	0.0453	0.0000
0 hr 4 min 49 sec	0.0002	0.9742	0.0100	100.0000	0.0073	98.3977	0.0527	0.0000
0 hr 4 min 50 sec	0.0038	0.9310	0.0009	99.2188	0.0411	98.3977	0.0353	0.0000
0 hr 4 min 50 sec	0.0002	0.8077	4.5287e-5	100.0000	0.0089	98.1063	0.0559	0.0000
0 hr 4 min 49 sec	0.0002	0.8008	0.0100	100.0000	0.0104	98.0335	0.0615	0.0000
0 hr 4 min 51 sec	0.0002	0.8002	0.0098	100.0000	0.0083	97.9607	0.0578	0.0000
0 hr 4 min 50 sec	0.0053	0.9345	0.0018	96.8750	0.0559	97.3052	0.0733	0.0000
0 hr 4 min 50 sec	0.0055	0.9216	0.0012	99.2188	0.0548	97.0138	0.0770	0.0000
0 hr 4 min 48 sec	0.0057	0.9719	0.0047	96.8750	0.0712	96.0670	0.1088	0.0000
0 hr 4 min 49 sec	0.0068	0.9733	0.0068	82.8125	1.9955	84.9964	1.9564	0.0000
0 hr 4 min 49 sec	0.0099	0.8062	0.0078	82.8125	5.3981	84.9964	5.8279	0.0000
0 hr 4 min 49 sec	0.0078	0.8256	0.0044	82.8125	0.7085	84.9964	0.5502	0.0000
0 hr 4 min 48 sec	0.0076	0.8006	4.7092e-6	82.8125	1.5439	84.9964	1.5599	0.0000
0 hr 4 min 51 sec	0.0086	0.9028	0.0072	82.8125	1.6735	84.9964	0.6769	0.0000
0 hr 4 min 49 sec	0.0064	0.9601	0.0036	82.8125	0.4841	84.9964	0.4434	0.0000
0 hr 4 min 49 sec	0.0099	0.8022	0.0099	82.8125	2.0547	84.9964	2.0908	0.0000
0 hr 4 min 50 sec	0.0100	0.8033	0.0003	82.8125	0.8548	84.9964	1.2911	0.0000
0 hr 4 min 50 sec	0.0099	0.9786	0.0099	82.8125	2.5191	84.9964	2.5683	0.0000
0 hr 4 min 49 sec	0.0071	0.9046	0.0006	82.8125	0.8221	84.9964	0.5853	0.0000
0 hr 4 min 49 sec	0.0090	0.8395	0.0032	82.8125	2.6673	84.9964	1.0918	0.0000
0 hr 4 min 50 sec	0.0067	0.9142	0.0047	82.8125	0.4931	84.9964	0.4572	0.0000
0 hr 4 min 50 sec	0.0069	0.8126	0.0046	82.8125	1.0606	84.9964	0.9265	0.0000
0 hr 4 min 49 sec	0.0087	0.9305	0.0012	82.8125	0.5112	84.9964	0.5707	0.0000
0 hr 4 min 48 sec	0.0095	0.9795	0.0030	82.8125	0.6563	18.6453	4.4885	0.0000

Appendix Figure D3. Dead < 2 years vs. > 2 years DarkNet19 Hyperparameter optimization. The following Bayesian optimization included: Initial learning rate [1e-4 1e-2], Momentum [0.8 0.98], L2Regularization [1e-10 1e-2], Maximum time (in seconds): Inf, Maximum number of trials: 30, and custom metrics: ErrorRate (direction, minimize).

Appendix E Parameter Tuning in DLBCL vs. Reactive

The analysis of image classification between DLBCL and reactive lymphoid tissue was expanded using DarkNet-19 CNN, data augmentation, and hyperparameter tuning. The following Bayesian optimization included: Initial learning rate [1e-4 1e-2], Momentum [0.8 0.98], L2Regularization [1e-10 1e-2], Maximum time (in seconds): Inf, Maximum number of trials: 30, and custom metrics: ErrorRate (direction, minimize).

Elapsed Time	Hyperparameters			Metrics				
	myInitialLearnRate	Momentum	L2Regularization	Training Accuracy (%)	Training Loss	Validation Accuracy (%)	Validation Loss	ErrorRate
3 hr 33 min 11 sec	0.0011	0.8005	0.0002	99.2188	0.0196	99.9145	0.0030	0.0000
3 hr 32 min 46 sec	0.0020	0.9782	0.0099	99.2188	0.0151	99.8842	0.0033	0.0000
3 hr 35 min 14 sec	0.0031	0.8478	0.0075	99.2188	0.0194	99.8842	0.0031	0.0000
3 hr 32 min 11 sec	0.0008	0.9800	0.0097	100.0000	0.0017	99.8759	0.0039	0.0000
3 hr 30 min 12 sec	0.0011	0.8001	0.0100	99.2188	0.0219	99.8401	0.0052	0.0000
3 hr 27 min 48 sec	0.0048	0.9724	1.0533e-5	99.2188	0.0134	99.7766	0.0063	0.0000
3 hr 29 min 20 sec	0.0063	0.8097	0.0001	99.2188	0.0388	99.4623	0.0151	0.0000
3 hr 33 min 5 sec	0.0131	0.9645	0.0005	98.4375	0.0453	98.9576	0.0277	0.0000
3 hr 25 min 50 sec	0.0179	0.9410	0.0030	98.4375	0.0429	98.6819	0.0353	0.0000
3 hr 26 min 48 sec	0.0183	0.8164	0.0074	97.6563	0.0704	98.6709	0.0357	0.0000
3 hr 35 min 12 sec	0.0194	0.8872	0.0034	98.4375	0.0670	98.1938	0.0491	0.0000
3 hr 28 min 8 sec	0.0184	0.8634	0.0062	97.6563	0.0514	98.0642	0.0560	0.0000
3 hr 26 min 8 sec	0.0333	0.9073	0.0064	98.4375	0.0405	97.9043	0.0560	0.0000
3 hr 30 min 18 sec	0.0737	0.9581	0.0053	98.4375	0.0639	97.5292	0.0700	0.0000
3 hr 36 min 37 sec	0.0534	0.9473	0.0044	97.6563	0.0750	97.4024	0.0752	0.0000
3 hr 27 min 42 sec	0.0462	0.8517	0.0089	96.8750	0.0761	97.3141	0.0676	0.0000
3 hr 36 min 4 sec	0.0331	0.9131	0.0085	99.2188	0.0765	97.2921	0.0695	0.0000
3 hr 22 min 32 sec	0.0404	0.8849	0.0055	96.8750	0.0816	97.0935	0.0734	0.0000
3 hr 29 min 55 sec	0.0999	0.9764	0.0099	97.6563	0.0659	96.7240	0.0825	0.0000
3 hr 25 min 31 sec	0.0903	0.9566	0.0025	98.4375	0.0808	96.6358	0.0875	0.0000
3 hr 35 min 10 sec	0.0308	0.8020	0.0066	96.8750	0.0878	96.5834	0.0891	0.0000
3 hr 32 min 12 sec	0.0984	0.8042	0.0098	97.6563	0.0851	96.5558	0.0918	0.0000
3 hr 27 min 53 sec	0.0850	0.9480	0.0071	97.6563	0.1157	96.5227	0.0968	0.0000
3 hr 34 min 7 sec	0.0856	0.8922	0.0005	96.0938	0.1046	96.5172	0.0943	0.0000
3 hr 23 min 45 sec	0.0770	0.8024	0.0082	97.6563	0.1061	96.5117	0.1042	0.0000
3 hr 31 min 21 sec	0.0997	0.8088	0.0003	96.0938	0.1134	96.1201	0.1188	0.0000
3 hr 28 min 29 sec	0.0612	0.8676	0.0055	96.0938	0.0970	96.0374	0.1044	0.0000
3 hr 31 min 48 sec	0.0999	0.9786	0.0001	95.3125	0.1188	96.0319	0.1153	0.0000
3 hr 26 min 8 sec	0.0963	0.8249	0.0006	96.0938	0.0936	95.8912	0.1127	0.0000
3 hr 28 min 30 sec	0.0901	0.8626	0.0059	95.3125	0.1314	95.8333	0.1090	0.0000

Appendix Figure E. Parameter tuning in DLBCL vs. reactive lymphoid tissue. The analysis of image classification between DLBCL and reactive lymphoid tissue was expanded using DarkNet-19 CNN, data augmentation, and hyperparameter tuning. The following Bayesian optimization included: Initial learning rate [1e-4 1e-2], Momentum [0.8 0.98], L2Regularization [1e-10 1e-2], Maximum time (in seconds): Inf, Maximum number of trials: 30, and custom metrics: ErrorRate (direction, minimize).

ANALYSIS RESULT					
	Name	Type	Activations	Learnable Properties	States
1	input 256-256x3 images with 'rescale-zero-one' normalization	Image Input	$256(S) \times 256(S) \times 3(C) \times 1(B)$	-	-
2	conv1 12 3x3x3 convolutions with stride [1 1] and padding 'same'	2-D Convolution	$256(S) \times 256(S) \times 32(C) \times 1(B)$	Weights $3 \times 3 \times 3 \times 32$ Bias $1 \times 1 \times 32$	-
3	batchnorm1 Batch normalization with 32 channels	Batch Normalization	$256(S) \times 256(S) \times 32(C) \times 1(B)$	Offset $1 \times 1 \times 32$ Scale $1 \times 1 \times 32$	TrainedMean $1 \times 1 \times 32$ TrainedVariance $1 \times 1 \times 32$
4	leaky1 Leaky ReLU with scale 0.1	Leaky ReLU	$256(S) \times 256(S) \times 32(C) \times 1(B)$	-	-
5	pool1 2x2 max pooling with stride [2 2] and padding [0 0 0 0]	2-D Max Pooling	$128(S) \times 128(S) \times 32(C) \times 1(B)$	-	-
6	conv2 64 3x3x32 convolutions with stride [1 1] and padding 'same'	2-D Convolution	$128(S) \times 128(S) \times 64(C) \times 1(B)$	Weights $3 \times 3 \times 32 \times 64$ Bias $1 \times 1 \times 64$	-
7	batchnorm2 Batch normalization with 64 channels	Batch Normalization	$128(S) \times 128(S) \times 64(C) \times 1(B)$	Offset $1 \times 1 \times 64$ Scale $1 \times 1 \times 64$	TrainedMean $1 \times 1 \times 64$ TrainedVariance $1 \times 1 \times 64$
8	leaky2 Leaky ReLU with scale 0.1	Leaky ReLU	$128(S) \times 128(S) \times 64(C) \times 1(B)$	-	-
9	pool2 2x2 max pooling with stride [2 2] and padding [0 0 0 0]	2-D Max Pooling	$64(S) \times 64(S) \times 64(C) \times 1(B)$	-	-
10	conv3 128 3x3x64 convolutions with stride [1 1] and padding 'same'	2-D Convolution	$64(S) \times 64(S) \times 128(C) \times 1(B)$	Weights $3 \times 3 \times 64 \times 128$ Bias $1 \times 1 \times 128$	-
11	batchnorm3 Batch normalization with 128 channels	Batch Normalization	$64(S) \times 64(S) \times 128(C) \times 1(B)$	Offset $1 \times 1 \times 128$ Scale $1 \times 1 \times 128$	TrainedMean $1 \times 1 \times 128$ TrainedVariance $1 \times 1 \times 128$
12	leaky3 Leaky ReLU with scale 0.1	Leaky ReLU	$64(S) \times 64(S) \times 128(C) \times 1(B)$	-	-
13	conv4 64 1x1x128 convolutions with stride [1 1] and padding 'same'	2-D Convolution	$64(S) \times 64(S) \times 64(C) \times 1(B)$	Weights $1 \times 1 \times 128 \times 64$ Bias $1 \times 1 \times 64$	-
14	batchnorm4 Batch normalization with 64 channels	Batch Normalization	$64(S) \times 64(S) \times 64(C) \times 1(B)$	Offset $1 \times 1 \times 64$ Scale $1 \times 1 \times 64$	TrainedMean $1 \times 1 \times 64$ TrainedVariance $1 \times 1 \times 64$
15	leaky4 Leaky ReLU with scale 0.1	Leaky ReLU	$64(S) \times 64(S) \times 64(C) \times 1(B)$	-	-
16	conv5 128 3x3x64 convolutions with stride [1 1] and padding 'same'	2-D Convolution	$64(S) \times 64(S) \times 128(C) \times 1(B)$	Weights $3 \times 3 \times 64 \times 128$ Bias $1 \times 1 \times 128$	-
17	batchnorm5 Batch normalization with 128 channels	Batch Normalization	$64(S) \times 64(S) \times 128(C) \times 1(B)$	Offset $1 \times 1 \times 128$ Scale $1 \times 1 \times 128$	TrainedMean $1 \times 1 \times 128$ TrainedVariance $1 \times 1 \times 128$
18	leaky5 Leaky ReLU with scale 0.1	Leaky ReLU	$64(S) \times 64(S) \times 128(C) \times 1(B)$	-	-
19	pool3 2x2 max pooling with stride [2 2] and padding [0 0 0 0]	2-D Max Pooling	$32(S) \times 32(S) \times 128(C) \times 1(B)$	-	-
20	conv6 256 3x3x128 convolutions with stride [1 1] and padding 'same'	2-D Convolution	$32(S) \times 32(S) \times 256(C) \times 1(B)$	Weights $3 \times 3 \times 128 \times 256$ Bias $1 \times 1 \times 256$	-
21	batchnorm6 Batch normalization with 256 channels	Batch Normalization	$32(S) \times 32(S) \times 256(C) \times 1(B)$	Offset $1 \times 1 \times 256$ Scale $1 \times 1 \times 256$	TrainedMean $1 \times 1 \times 256$ TrainedVariance $1 \times 1 \times 256$
22	leaky6 Leaky ReLU with scale 0.1	Leaky ReLU	$32(S) \times 32(S) \times 256(C) \times 1(B)$	-	-
23	conv7 128 1x1x256 convolutions with stride [1 1] and padding 'same'	2-D Convolution	$32(S) \times 32(S) \times 128(C) \times 1(B)$	Weights $1 \times 1 \times 256 \times 128$ Bias $1 \times 1 \times 128$	-
24	batchnorm7 Batch normalization with 128 channels	Batch Normalization	$32(S) \times 32(S) \times 128(C) \times 1(B)$	Offset $1 \times 1 \times 128$ Scale $1 \times 1 \times 128$	TrainedMean $1 \times 1 \times 128$ TrainedVariance $1 \times 1 \times 128$
25	leaky7 Leaky ReLU with scale 0.1	Leaky ReLU	$32(S) \times 32(S) \times 128(C) \times 1(B)$	-	-
26	conv8 256 3x3x128 convolutions with stride [1 1] and padding 'same'	2-D Convolution	$32(S) \times 32(S) \times 256(C) \times 1(B)$	Weights $3 \times 3 \times 128 \times 256$ Bias $1 \times 1 \times 256$	-
27	batchnorm8 Batch normalization with 256 channels	Batch Normalization	$32(S) \times 32(S) \times 256(C) \times 1(B)$	Offset $1 \times 1 \times 256$ Scale $1 \times 1 \times 256$	TrainedMean $1 \times 1 \times 256$ TrainedVariance $1 \times 1 \times 256$
28	leaky8 Leaky ReLU with scale 0.1	Leaky ReLU	$32(S) \times 32(S) \times 256(C) \times 1(B)$	-	-
29	pool4 2x2 max pooling with stride [2 2] and padding [0 0 0 0]	2-D Max Pooling	$16(S) \times 16(S) \times 256(C) \times 1(B)$	-	-
30	conv9 512 3x3x256 convolutions with stride [1 1] and padding 'same'	2-D Convolution	$16(S) \times 16(S) \times 512(C) \times 1(B)$	Weights $3 \times 3 \times 256 \times 512$ Bias $1 \times 1 \times 512$	-
31	batchnorm9 Batch normalization with 512 channels	Batch Normalization	$16(S) \times 16(S) \times 512(C) \times 1(B)$	Offset $1 \times 1 \times 512$ Scale $1 \times 1 \times 512$	TrainedMean $1 \times 1 \times 512$ TrainedVariance $1 \times 1 \times 512$
32	leaky9 Leaky ReLU with scale 0.1	Leaky ReLU	$16(S) \times 16(S) \times 512(C) \times 1(B)$	-	-
33	conv10 256 1x1x512 convolutions with stride [1 1] and padding 'same'	2-D Convolution	$16(S) \times 16(S) \times 256(C) \times 1(B)$	Weights $1 \times 1 \times 512 \times 256$ Bias $1 \times 1 \times 256$	-
34	batchnorm10 Batch normalization with 256 channels	Batch Normalization	$16(S) \times 16(S) \times 256(C) \times 1(B)$	Offset $1 \times 1 \times 256$ Scale $1 \times 1 \times 256$	TrainedMean $1 \times 1 \times 256$ TrainedVariance $1 \times 1 \times 256$
35	leaky10 Leaky ReLU with scale 0.1	Leaky ReLU	$16(S) \times 16(S) \times 256(C) \times 1(B)$	-	-
36	conv11 512 3x3x256 convolutions with stride [1 1] and padding 'same'	2-D Convolution	$16(S) \times 16(S) \times 512(C) \times 1(B)$	Weights $3 \times 3 \times 256 \times 512$ Bias $1 \times 1 \times 512$	-
37	batchnorm11 Batch normalization with 512 channels	Batch Normalization	$16(S) \times 16(S) \times 512(C) \times 1(B)$	Offset $1 \times 1 \times 512$ Scale $1 \times 1 \times 512$	TrainedMean $1 \times 1 \times 512$ TrainedVariance $1 \times 1 \times 512$
38	leaky11 Leaky ReLU with scale 0.1	Leaky ReLU	$16(S) \times 16(S) \times 512(C) \times 1(B)$	-	-
39	conv12 256 1x1x512 convolutions with stride [1 1] and padding 'same'	2-D Convolution	$16(S) \times 16(S) \times 256(C) \times 1(B)$	Weights $1 \times 1 \times 512 \times 256$ Bias $1 \times 1 \times 256$	-
40	batchnorm12 Batch normalization with 256 channels	Batch Normalization	$16(S) \times 16(S) \times 256(C) \times 1(B)$	Offset $1 \times 1 \times 256$ Scale $1 \times 1 \times 256$	TrainedMean $1 \times 1 \times 256$ TrainedVariance $1 \times 1 \times 256$
41	leaky12 Leaky ReLU with scale 0.1	Leaky ReLU	$16(S) \times 16(S) \times 256(C) \times 1(B)$	-	-
42	conv13 512 3x3x256 convolutions with stride [1 1] and padding 'same'	2-D Convolution	$16(S) \times 16(S) \times 512(C) \times 1(B)$	Weights $3 \times 3 \times 256 \times 512$ Bias $1 \times 1 \times 512$	-
43	batchnorm13 Batch normalization with 512 channels	Batch Normalization	$16(S) \times 16(S) \times 512(C) \times 1(B)$	Offset $1 \times 1 \times 512$ Scale $1 \times 1 \times 512$	TrainedMean $1 \times 1 \times 512$ TrainedVariance $1 \times 1 \times 512$
44	leaky13 Leaky ReLU with scale 0.1	Leaky ReLU	$16(S) \times 16(S) \times 512(C) \times 1(B)$	-	-
45	pool5 2x2 max pooling with stride [2 2] and padding [0 0 0 0]	2-D Max Pooling	$8(S) \times 8(S) \times 512(C) \times 1(B)$	-	-
46	conv14 1024 3x3x512 convolutions with stride [1 1] and padding 'same'	2-D Convolution	$8(S) \times 8(S) \times 1024(C) \times 1(B)$	Weights $3 \times 3 \times 512 \times 1024$ Bias $1 \times 1 \times 1024$	-
47	batchnorm14 Batch normalization with 1024 channels	Batch Normalization	$8(S) \times 8(S) \times 1024(C) \times 1(B)$	Offset $1 \times 1 \times 1024$ Scale $1 \times 1 \times 1024$	TrainedMean $1 \times 1 \times 1024$ TrainedVariance $1 \times 1 \times 1024$
48	leaky14 Leaky ReLU with scale 0.1	Leaky ReLU	$8(S) \times 8(S) \times 1024(C) \times 1(B)$	-	-
49	conv15 512 1x1x1024 convolutions with stride [1 1] and padding 'same'	2-D Convolution	$8(S) \times 8(S) \times 512(C) \times 1(B)$	Weights $1 \times 1 \times 1024 \times 512$ Bias $1 \times 1 \times 512$	-
50	batchnorm15 Batch normalization with 512 channels	Batch Normalization	$8(S) \times 8(S) \times 512(C) \times 1(B)$	Offset $1 \times 1 \times 512$ Scale $1 \times 1 \times 512$	TrainedMean $1 \times 1 \times 512$ TrainedVariance $1 \times 1 \times 512$
51	leaky15 Leaky ReLU with scale 0.1	Leaky ReLU	$8(S) \times 8(S) \times 512(C) \times 1(B)$	-	-
52	conv16 1024 3x3x512 convolutions with stride [1 1] and padding 'same'	2-D Convolution	$8(S) \times 8(S) \times 1024(C) \times 1(B)$	Weights $3 \times 3 \times 512 \times 1024$ Bias $1 \times 1 \times 1024$	-
53	batchnorm16 Batch normalization with 1024 channels	Batch Normalization	$8(S) \times 8(S) \times 1024(C) \times 1(B)$	Offset $1 \times 1 \times 1024$ Scale $1 \times 1 \times 1024$	TrainedMean $1 \times 1 \times 1024$ TrainedVariance $1 \times 1 \times 1024$
54	leaky16 Leaky ReLU with scale 0.1	Leaky ReLU	$8(S) \times 8(S) \times 1024(C) \times 1(B)$	-	-
55	conv17 512 1x1x1024 convolutions with stride [1 1] and padding 'same'	2-D Convolution	$8(S) \times 8(S) \times 512(C) \times 1(B)$	Weights $1 \times 1 \times 1024 \times 512$ Bias $1 \times 1 \times 512$	-
56	batchnorm17 Batch normalization with 512 channels	Batch Normalization	$8(S) \times 8(S) \times 512(C) \times 1(B)$	Offset $1 \times 1 \times 512$ Scale $1 \times 1 \times 512$	TrainedMean $1 \times 1 \times 512$ TrainedVariance $1 \times 1 \times 512$
57	leaky17 Leaky ReLU with scale 0.1	Leaky ReLU	$8(S) \times 8(S) \times 512(C) \times 1(B)$	-	-
58	conv18 1024 3x3x512 convolutions with stride [1 1] and padding 'same'	2-D Convolution	$8(S) \times 8(S) \times 1024(C) \times 1(B)$	Weights $3 \times 3 \times 512 \times 1024$ Bias $1 \times 1 \times 1024$	-
59	batchnorm18 Batch normalization with 1024 channels	Batch Normalization	$8(S) \times 8(S) \times 1024(C) \times 1(B)$	Offset $1 \times 1 \times 1024$ Scale $1 \times 1 \times 1024$	TrainedMean $1 \times 1 \times 1024$ TrainedVariance $1 \times 1 \times 1024$
60	leaky18 Leaky ReLU with scale 0.1	Leaky ReLU	$8(S) \times 8(S) \times 1024(C) \times 1(B)$	-	-
61	conv19 2 1x1 convolutions with stride [1 1] and padding 'same'	2-D Convolution	$8(S) \times 8(S) \times 2(C) \times 1(B)$	Weights $1 \times 1 \times 1024 \times 2$ Bias $1 \times 1 \times 2$	-
62	avg1 2x2 global average pooling	Softmax	$1(S) \times 1(S) \times 2(C) \times 1(B)$	-	-
63	softmax softmax	Softmax	$1(S) \times 1(S) \times 2(C) \times 1(B)$	-	-
64	classoutput crossentropyloss	Classification Output	$1(S) \times 1(S) \times 2(C) \times 1(B)$	-	-

Appendix Figure F. DarkNet-19 architecture. DarkNet-19 is a convolutional neural network (CNN) that is 19 layers deep. This CNN designed for image classification is characterized by 19.8 M total learnables, 64 layers and 63 numbers of connections.

References

1. Brown JR, Aster JC, Lister A, Rosmarin A. Pathobiology of diffuse large B cell lymphoma and primary mediastinal large B cell lymphoma. In UpToDate. Website: www.uptodate.com (last accessed on 19th November 2025).
2. Alaggio, R.; Amador, C.; Anagnostopoulos, I.; Attygalle, A.D.; Araujo, I.B.O.; Berti, E.; Bhagat, G.; Borges, A.M.; Boyer, D.; Calaminici, M.; et al. The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Lymphoid Neoplasms. *Leukemia* **2022**, *36*, 1720-1748, doi:10.1038/s41375-022-01620-2.
3. Arber, D.A.; Campo, E.; Jaffe, E.S. Advances in the Classification of Myeloid and Lymphoid Neoplasms. *Virchows Arch* **2023**, *482*, 1-9, doi:10.1007/s00428-022-03487-1.
4. Campo, E. The 2022 classifications of lymphoid neoplasms : Keynote. *Pathologie (Heidelb)* **2023**, *44*, 121-127, doi:10.1007/s00292-023-01247-9.
5. Campo, E.; Jaffe, E.S.; Cook, J.R.; Quintanilla-Martinez, L.; Swerdlow, S.H.; Anderson, K.C.; Brousset, P.; Cerroni, L.; de Leval, L.; Dirnhofer, S.; et al. The International Consensus Classification of Mature Lymphoid Neoplasms: a report from the Clinical Advisory Committee. *Blood* **2022**, *140*, 1229-1253, doi:10.1182/blood.2022015851.
6. de Leval, L.; Alizadeh, A.A.; Bergsagel, P.L.; Campo, E.; Davies, A.; Dogan, A.; Fitzgibbon, J.; Horwitz, S.M.; Melnick, A.M.; Morice, W.G.; et al. Genomic profiling for clinical decision making in lymphoid neoplasms. *Blood* **2022**, *140*, 2193-2227, doi:10.1182/blood.2022015854.
7. Song, J.Y.; Dirnhofer, S.; Piris, M.A.; Quintanilla-Martinez, L.; Pileri, S.; Campo, E. Diffuse large B-cell lymphomas, not otherwise specified, and emerging entities. *Virchows Arch* **2023**, *482*, 179-192, doi:10.1007/s00428-022-03466-6.
8. Li, S.; Young, K.H.; Medeiros, L.J. Diffuse large B-cell lymphoma. *Pathology* **2018**, *50*, 74-87, doi:10.1016/j.pathol.2017.09.006.
9. Crump, M. Management of Relapsed Diffuse Large B-cell Lymphoma. *Hematol Oncol Clin North Am* **2016**, *30*, 1195-1213, doi:10.1016/j.hoc.2016.07.004.
10. Swerdlow, S.H.; Campo, E.; Pileri, S.A.; Harris, N.L.; Stein, H.; Siebert, R.; Advani, R.; Ghielmini, M.; Salles, G.A.; Zelenetz, A.D.; et al. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood* **2016**, *127*, 2375-2390, doi:10.1182/blood-2016-01-643569.
11. Xie, Y.; Pittaluga, S.; Jaffe, E.S. The histological classification of diffuse large B-cell lymphomas. *Semin Hematol* **2015**, *52*, 57-66, doi:10.1053/j.seminhematol.2015.01.006.
12. de Leval, L.; Jaffe, E.S. Lymphoma Classification. *Cancer J* **2020**, *26*, 176-185, doi:10.1097/PPO.0000000000000451.
13. Willemze, R.; Cerroni, L.; Kempf, W.; Berti, E.; Facchetti, F.; Swerdlow, S.H.; Jaffe, E.S. The 2018 update of the WHO-EORTC classification for primary cutaneous lymphomas. *Blood* **2019**, *133*, 1703-1714, doi:10.1182/blood-2018-11-881268.
14. Jaffe, E.S. Diagnosis and classification of lymphoma: Impact of technical advances. *Semin Hematol* **2019**, *56*, 30-36, doi:10.1053/j.seminhematol.2018.05.007.
15. Quintanilla-Martinez, L.; Swerdlow, S.H.; Tousseyn, T.; Barrionuevo, C.; Nakamura, S.; Jaffe, E.S. New concepts in EBV-associated B, T, and NK cell lymphoproliferative disorders. *Virchows Arch* **2023**, *482*, 227-244, doi:10.1007/s00428-022-03414-4.
16. Hans, C.P.; Weisenburger, D.D.; Greiner, T.C.; Gascoyne, R.D.; Delabie, J.; Ott, G.; Muller-Hermelink, H.K.; Campo, E.; Braziel, R.M.; Jaffe, E.S.; et al. Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. *Blood* **2004**, *103*, 275-282, doi:10.1182/blood-2003-05-1545.
17. Collinge, B.; Hilton, L.K.; Wong, J.; Alduaij, W.; Ben-Neriah, S.; Slack, G.W.; Farinha, P.; Boyle, M.; Meissner, B.; Cook, J.R.; et al. High-grade B-cell lymphoma, not otherwise specified: an LLMP study. *Blood Adv* **2025**, *9*, 5409-5422, doi:10.1182/bloodadvances.2025016651.
18. Herlevic, V.; Reynolds, S.B.; Morris, J.D. Gastric Lymphoma. In *StatPearls*; Treasure Island (FL), 2025.
19. Grainger, B.T.; Cheah, C.Y. Primary testicular lymphoma. *Cancer Treat Rev* **2025**, *136*, 102927, doi:10.1016/j.ctrv.2025.102927.

20. Chen, S.Y.; Xu, P.P.; Feng, R.; Cui, G.H.; Wang, L.; Cheng, S.; Mu, R.J.; Zhang, H.L.; Wei, X.L.; Song, Y.P.; et al. Extranodal diffuse large B-cell lymphoma: Clinical and molecular insights with survival outcomes from the multicenter EXPECT study. *Cancer Commun (Lond)* **2025**, *45*, 919-935, doi:10.1002/cac2.70033.
21. International Non-Hodgkin's Lymphoma Prognostic Factors, P. A predictive model for aggressive non-Hodgkin's lymphoma. *N Engl J Med* **1993**, *329*, 987-994, doi:10.1056/NEJM199309303291402.
22. Scott, D.W. Cell-of-Origin in Diffuse Large B-Cell Lymphoma: Are the Assays Ready for the Clinic? *Am Soc Clin Oncol Educ Book* **2015**, e458-466, doi:10.14694/EdBook_AM.2015.35.e458.
23. Choi, W.W.; Weisenburger, D.D.; Greiner, T.C.; Piris, M.A.; Banham, A.H.; Delabie, J.; Braziel, R.M.; Geng, H.; Iqbal, J.; Lenz, G.; et al. A new immunostain algorithm classifies diffuse large B-cell lymphoma into molecular subtypes with high accuracy. *Clin Cancer Res* **2009**, *15*, 5494-5502, doi:10.1158/1078-0432.CCR-09-0113.
24. Alizadeh, A.A.; Eisen, M.B.; Davis, R.E.; Ma, C.; Lossos, I.S.; Rosenwald, A.; Boldrick, J.C.; Sabet, H.; Tran, T.; Yu, X.; et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **2000**, *403*, 503-511, doi:10.1038/35000501.
25. Rosenwald, A.; Wright, G.; Chan, W.C.; Connors, J.M.; Campo, E.; Fisher, R.I.; Gascoyne, R.D.; Muller-Hermelink, H.K.; Smeland, E.B.; Giltman, J.M.; et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med* **2002**, *346*, 1937-1947, doi:10.1056/NEJMoa012914.
26. Wright, G.; Tan, B.; Rosenwald, A.; Hurt, E.H.; Wiestner, A.; Staudt, L.M. A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. *Proc Natl Acad Sci U S A* **2003**, *100*, 9991-9996, doi:10.1073/pnas.1732008100.
27. Berglund, M.; Thunberg, U.; Amini, R.M.; Book, M.; Roos, G.; Erlanson, M.; Linderöth, J.; Dictor, M.; Jerkeman, M.; Cavallin-Stahl, E.; et al. Evaluation of immunophenotype in diffuse large B-cell lymphoma and its impact on prognosis. *Mod Pathol* **2005**, *18*, 1113-1120, doi:10.1038/modpathol.3800396.
28. Aster JC, Abramson JS, Lister A, Rosmarin AG. Prognosis of diffuse large B cell lymphoma. In UpToDate (last accessed on 20th November, 2025).
29. Horn, H.; Ziepert, M.; Becher, C.; Barth, T.F.; Bernd, H.W.; Feller, A.C.; Klapper, W.; Hummel, M.; Stein, H.; Hansmann, M.L.; et al. MYC status in concert with BCL2 and BCL6 expression predicts outcome in diffuse large B-cell lymphoma. *Blood* **2013**, *121*, 2253-2263, doi:10.1182/blood-2012-06-435842.
30. Johnson, N.A.; Slack, G.W.; Savage, K.J.; Connors, J.M.; Ben-Neriah, S.; Rogic, S.; Scott, D.W.; Tan, K.L.; Steidl, C.; Sehn, L.H.; et al. Concurrent expression of MYC and BCL2 in diffuse large B-cell lymphoma treated with rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone. *J Clin Oncol* **2012**, *30*, 3452-3459, doi:10.1200/JCO.2011.41.0985.
31. Green, T.M.; Young, K.H.; Visco, C.; Xu-Monette, Z.Y.; Orazi, A.; Go, R.S.; Nielsen, O.; Gadeberg, O.V.; Mourits-Andersen, T.; Frederiksen, M.; et al. Immunohistochemical double-hit score is a strong predictor of outcome in patients with diffuse large B-cell lymphoma treated with rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone. *J Clin Oncol* **2012**, *30*, 3460-3467, doi:10.1200/JCO.2011.41.4342.
32. Hu, S.; Xu-Monette, Z.Y.; Tzankov, A.; Green, T.; Wu, L.; Balasubramanyam, A.; Liu, W.M.; Visco, C.; Li, Y.; Miranda, R.N.; et al. MYC/BCL2 protein coexpression contributes to the inferior survival of activated B-cell subtype of diffuse large B-cell lymphoma and demonstrates high-risk gene expression signatures: a report from The International DLBCL Rituximab-CHOP Consortium Program. *Blood* **2013**, *121*, 4021-4031; quiz 4250, doi:10.1182/blood-2012-10-460063.
33. Somasundaram, E.; Abramson, J.S. Double hit lymphoma: contemporary understanding and practices. *Leuk Lymphoma* **2025**, *66*, 26-33, doi:10.1080/10428194.2024.2412295.
34. Karmali, R.; Shouse, G.; Torka, P.; Moyo, T.K.; Romancik, J.; Barta, S.K.; Bhansali, R.; Cohen, J.B.; Shah, N.N.; Zurko, J.; et al. Double hit & double expressor lymphomas: a multicenter analysis of survival outcomes with CD19-directed CAR T-cell therapy. *Blood Cancer J* **2025**, *15*, 43, doi:10.1038/s41408-025-01250-8.
35. Qiu, L.; Medeiros, L.J.; Li, S. High-grade B-cell lymphomas: Double hit and non-double hit. *Hum Pathol* **2025**, *156*, 105700, doi:10.1016/j.humpath.2024.105700.

36. Reddy, A.; Zhang, J.; Davis, N.S.; Moffitt, A.B.; Love, C.L.; Waldrop, A.; Leppa, S.; Pasanen, A.; Meriranta, L.; Karjalainen-Lindsberg, M.L.; et al. Genetic and Functional Drivers of Diffuse Large B Cell Lymphoma. *Cell* **2017**, *171*, 481–494 e415, doi:10.1016/j.cell.2017.09.027.
37. Schmitz, R.; Wright, G.W.; Huang, D.W.; Johnson, C.A.; Phelan, J.D.; Wang, J.Q.; Roulland, S.; Kasbekar, M.; Young, R.M.; Shaffer, A.L.; et al. Genetics and Pathogenesis of Diffuse Large B-Cell Lymphoma. *N Engl J Med* **2018**, *378*, 1396–1407, doi:10.1056/NEJMoa1801445.
38. Chapuy, B.; Stewart, C.; Dunford, A.J.; Kim, J.; Kamburov, A.; Redd, R.A.; Lawrence, M.S.; Roemer, M.G.M.; Li, A.J.; Ziepert, M.; et al. Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. *Nat Med* **2018**, *24*, 679–690, doi:10.1038/s41591-018-0016-8.
39. Carreras, J.; Kikuti, Y.Y.; Roncador, G.; Miyaoka, M.; Hiraiwa, S.; Tomita, S.; Ikoma, H.; Kondo, Y.; Ito, A.; Shiraiwa, S.; et al. High Expression of Caspase-8 Associated with Improved Survival in Diffuse Large B-Cell Lymphoma: Machine Learning and Artificial Neural Networks Analyses. *BioMedInformatics* **2021**, *1*, 18–46. <https://doi.org/10.3390/biomedinformatics1010003>.
40. Carreras, J.; Kikuti, Y.Y.; Miyaoka, M.; Hiraiwa, S.; Tomita, S.; Ikoma, H.; Kondo, Y.; Ito, A.; Shiraiwa, S.; Hamoudi, R.; et al. A Single Gene Expression Set Derived from Artificial Intelligence Predicted the Prognosis of Several Lymphoma Subtypes; and High Immunohistochemical Expression of TNFAIP8 Associated with Poor Prognosis in Diffuse Large B-Cell Lymphoma. *AI* **2020**, *1*, 342–360. <https://doi.org/10.3390/ai1030023>
41. Carreras, J.; Kikuti, Y.Y.; Miyaoka, M.; Roncador, G.; Garcia, J.F.; Hiraiwa, S.; Tomita, S.; Ikoma, H.; Kondo, Y.; Ito, A.; et al. Integrative Statistics, Machine Learning and Artificial Intelligence Neural Network Analysis Correlated CSF1R with the Prognosis of Diffuse Large B-Cell Lymphoma. *Hemato* **2021**, *2*, 182–206. <https://doi.org/10.3390/hemato2020011>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.