





Article

Deep Neural Networks and Machine Learning Radiomics Modelling for Prediction of Relapse in Mantle Cell Lymphoma

Catharina Silvia Lisson ^{1,2,3,*}, Christoph Gerhard Lisson ¹, Marc Fabian Mezger ^{1,3,4}, Daniel Wolf ^{1,3,4}, Stefan Andreas Schmidt ^{1,2} , Wolfgang M. Thaiss ^{1,3,5}, Eugen Tausch ^{6,7}, Ambros J. Beer ^{2,3,5,8,9} , Stephan Stilgenbauer ^{6,7}, Meinrad Beer ^{1,2,3,8,9}  and Michael Goetz ^{1,3,10} 

¹ Department of Diagnostic and Interventional Radiology, University Hospital of Ulm, Albert-Einstein-Allee 23, 89081 Ulm, Germany; christoph.lisson@uniklinik-ulm.de (C.G.L.); marc.mezger@uni-ulm.de (M.F.M.); daniel.wolf@uniklinik-ulm.de (D.W.); stefan.schmidt@uniklinik-ulm.de (S.A.S.); wolfgang.thaiss@uniklinik-ulm.de (W.M.T.); meinrad.beer@uniklinik-ulm.de (M.B.); michael.goetz@uni-ulm.de (M.G.)

² Center for Personalized Medicine (ZPM), University Hospital of Ulm, Albert-Einstein-Allee 23, 89081 Ulm, Germany; ambros.beer@uniklinik-ulm.de

³ Artificial Intelligence in Experimental Radiology (XAIRAD)

⁴ Visual Computing Group, Institute of Media Informatics, Ulm University, James-Franck-Ring, 89081 Ulm, Germany

⁵ Department of Nuclear Medicine, University Hospital of Ulm, Albert-Einstein-Allee 23, 89081 Ulm, Germany

⁶ Department of Internal Medicine III, University Hospital of Ulm, Albert-Einstein-Allee 23, 89081 Ulm, Germany; eugen.tausch@uniklinik-ulm.de (E.T.); stephan.stilgenbauer@uniklinik-ulm.de (S.S.)

⁷ Comprehensive Cancer Center Ulm (CCCU), University Hospital of Ulm, Albert-Einstein-Allee 23, 89081 Ulm, Germany

⁸ Center for Translational Imaging “From Molecule to Man” (MoMan), Department of Internal Medicine II, University Hospital of Ulm, Albert-Einstein-Allee 23, 89081 Ulm, Germany

⁹ i2SouI—Innovative Imaging in Surgical Oncology Ulm, University Hospital of Ulm, Albert-Einstein-Allee 23, 89081 Ulm, Germany

¹⁰ German Cancer Research Center (DKFZ), Division Medical Image Computing, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany

* Correspondence: catharina.lisson@uniklinik-ulm.de; Tel.: +49-731-50061171



Citation: Lisson, C.S.; Lisson, C.G.; Mezger, M.F.; Wolf, D.; Schmidt, S.A.; Thaiss, W.M.; Tausch, E.; Beer, A.J.; Stilgenbauer, S.; Beer, M.; et al. Deep Neural Networks and Machine Learning Radiomics Modelling for Prediction of Relapse in Mantle Cell Lymphoma. *Cancers* **2022**, *14*, 2008. <https://doi.org/10.3390/cancers14082008>

Academic Editor: Alberto Zamò

Received: 3 February 2022

Accepted: 12 April 2022

Published: 15 April 2022

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Simple Summary: Mantle cell lymphoma (MCL) is an aggressive lymphoid tumour with a poor prognosis. There exist no routine biomarkers for the early prediction of relapse. Our study compared the potential of radiomics-based machine learning and 3D deep learning models as non-invasive biomarkers to risk-stratify MCL patients, thus promoting precision imaging in clinical oncology.

Abstract: Mantle cell lymphoma (MCL) is a rare lymphoid malignancy with a poor prognosis characterised by frequent relapse and short durations of treatment response. Most patients present with aggressive disease, but there exist indolent subtypes without the need for immediate intervention. The very heterogeneous behaviour of MCL is genetically characterised by the translocation t(11;14)(q13;q32), leading to Cyclin D1 overexpression with distinct clinical and biological characteristics and outcomes. There is still an unfulfilled need for precise MCL prognostication in real-time. Machine learning and deep learning neural networks are rapidly advancing technologies with promising results in numerous fields of application. This study develops and compares the performance of deep learning (DL) algorithms and radiomics-based machine learning (ML) models to predict MCL relapse on baseline CT scans. Five classification algorithms were used, including three deep learning models (3D SEResNet50, 3D DenseNet, and an optimised 3D CNN) and two machine learning models based on K-nearest Neighbor (KNN) and Random Forest (RF). The best performing method, our optimised 3D CNN, predicted MCL relapse with a 70% accuracy, better than the 3D SEResNet50 (62%) and the 3D DenseNet (59%). The second-best performing method was the KNN-based machine learning model (64%) after principal component analysis for improved accuracy. Our optimised CNN developed by ourselves correctly predicted MCL relapse in 70% of the patients on baseline CT imaging. Once prospectively tested in clinical trials with a larger sample size, our proposed 3D deep learning model could facilitate clinical management by precision imaging in MCL.

Keywords: radiomics; machine learning; deep learning; deep neural networks; personalised oncology; precision imaging

1. Introduction

Mantle cell lymphoma (MCL) is a rare subtype of lymphoid malignancy, representing 2.5–6% of all non-Hodgkin lymphoma (NHL) [1,2]. MCL is considered a heterogeneous disease with various clinical presentations ranging from minimal symptoms to progressive generalised lymphadenopathy [3–6]. Most MCL patients present with aggressive disease in an advanced stage (Stage III-IV); however, a subset of patients exhibit an indolent evolution [7]. As a result, the World Health Organization (WHO) updated the classification of MCL in 2017, describing two main subtypes, “classical MCL” and “indolent leukemic non-nodal/smouldering MCL”, with significant differences in the molecular characteristics, clinical features and prognosis [8]. According to follow-up studies, early identification of high-risk patients may modify their therapeutic management, reduce unnecessary toxicity, and improve prognosis [9–12].

As initial staging, the ESMO guidelines recommend performing a contrast-enhanced computed tomography (CT) scan of the neck, thorax, abdomen and pelvis as treatment differs depending on the stage of the disease [7]. A PET/CT scan is especially recommended in the rare limited stages I/II before localised radiotherapy, although reimbursement by the domestic health insurances is not ensured. There is consensus to base diagnosis upon histopathological evaluation of a lymph node biopsy rather than a fine needle aspiration, keeping in mind the heterogeneity of MCL [7].

Novel therapeutic options are constantly evolving, but until today, management of MCL is challenging due to its pattern of resistance, early relapse and poor long-term survival [13]. Despite improvements in remission durations, MCL is regarded as an incurable malignancy, despite the novel drug strategies developed in recent years [14,15].

The “mantle cell lymphoma International Prognostic Index (MIPI)”, based on age, ECOG performance status, lactate dehydrogenase and leukocyte count, assigns patients to low-risk, intermediate-risk and high-risk groups with 5-year survival rates of 60%, median overall survival of 51 months or median overall survival OS 29 months, respectively [16]. Furthermore, the cell proliferation antigen Ki-67 is considered the most established histomorphological risk factor in MCL [2]. Combining the Ki-67 index with the MIPI led to the more refined MIPI-c [17]. Additional TP53 gene mutation at diagnosis is associated with poor response to upfront intensive chemotherapy and poor prognosis [4].

The prognostic potential of minimal residual disease (MRD) has been confirmed in numerous studies [18,19]. Due to current limitations, for example, how to react in MRD-positive patients, MRD is tested in clinical trials but not recommended in clinical routine. Currently, no biomarkers definitely predict indolent behaviour; instead, a short ‘watch and wait’ period under close observation seems appropriate in indolent cases with low tumour burden [14,20].

The treatment selection of any given MCL patient is based mainly on clinical and biological risk factors, symptoms and tumour load [7]. In the era of precision oncology, there is great enthusiasm to identify tumour-specific targetable alterations for precise management for the individual patient, maximising the success of treatment by minimising the therapy-related toxicities.

Most research in precision medicine focuses on molecular tumour profiling using genomics-based arrays from biopsies with promising results in clinical oncology [11], including haematological diseases [21]. However, tumours are spatially and temporally heterogeneous, i.e., demanding repeated biopsies to capture their molecular heterogeneity with an increased risk for the patient [22–25].

Advanced medical imaging addresses the challenges of biopsy being non-invasive, repeatable and applicable for difficult-to-reach lesions within the body. The emergence

of high-resolution image acquisition and modern computational technologies amid the requirements of precision medicine have given rise to artificial intelligence (AI)—based image analysis in radiology. In general, AI techniques detect specific patterns in medical images using data characterisation algorithms to convert conventional imaging information into quantitative mineable “big data” [11,26]. Radiomics uses imaging-based texture analysis to generate imaging features extracted from a region of interest (ROI) to reflect tumour physiology and radiographic phenotype for classification, prognosis and monitoring of treatment response in many cancer types, including lymphoma [12,26–53]. Additionally, there are indications of an association between radiomic features and the underlying gene expression patterns [54]. Combining radiomic image analysis with classic machine learning algorithms is an emerging technique. Machine-learning (ML) refers to algorithms that analyse and learn on a dataset before making decisions or predictions based on the learned information [55,56]. Radiomics-based ML has been shown to predict primary treatment failure in diffuse large B-cell lymphoma (DLBCL) better than standard clinical criteria [45]. However, high-throughput radiomic analysis is limited by manual delineation of tumour boundaries, which is labour intensive with poor reproducibility.

Deep learning (DL) algorithms are considered advanced AI techniques and have emerged as a promising tool in many research fields [57,58]. DL is based on complex artificial neural networks that detect patterns in the data and learn efficient features from the raw data without the preprocessing steps required in classic machine learning models [59]. For image analysis, deep neural networks based on convolutional neural networks (CNNs) have proven efficient in accurately locating and segmenting tumour lesions, useful for diagnosis and image analysis tasks, and outperforming other traditional automated algorithms [60].

To the best of our knowledge, this is the first study to develop and compare 3D deep neural networks with radiomics-based machine learning methods on routine CT imaging at baseline for predicting relapse of mantle cell lymphoma (MCL), thus promoting precision imaging in clinical oncology.

2. Materials and Methods

2.1. Patients and Imaging Protocol

Enrolled in this retrospective study were 30 treatment-naive patients with histologically proven mantle cell lymphoma who underwent contrast-enhanced CT or PET/CT scans at our institution between January 2005 and January 2016. Demographic patient data, laboratory and clinical data (such as the white blood cell count, lactate dehydrogenase levels and Ki-67 proliferation), disease stage according to the Ann Arbor staging system, treatment regimen and clinical outcome data were recorded by thoroughly reviewing electronic charts and the radiology information system.

Incomplete clinical and imaging data records and lack of histological confirmation were exclusion criteria. Further exclusion criteria included patients whose disease status was not confirmed at the end of therapy and those who had disease progression within the first line of treatment.

Additional inclusion criteria were treatment with an R-CHOP-based regimen (rituximab, cyclophosphamide, doxorubicin, vincristine and prednisone) alternating with or instead of R-DHAP (rituximab, dexamethasone, high-dose cytarabine and cisplatin) or R-FC (rituximab, fludarabine and cyclophosphamide).

The primary endpoint was relapse of MCL based on clinical assessment and imaging studies. Follow-up recordings in the electronic charts determined each patient’s status.

All enrolled MCL subjects were followed up for at least five years after therapy, except those who experienced death by any cause. A flow diagram of the cohort selection is presented in Figure 1.

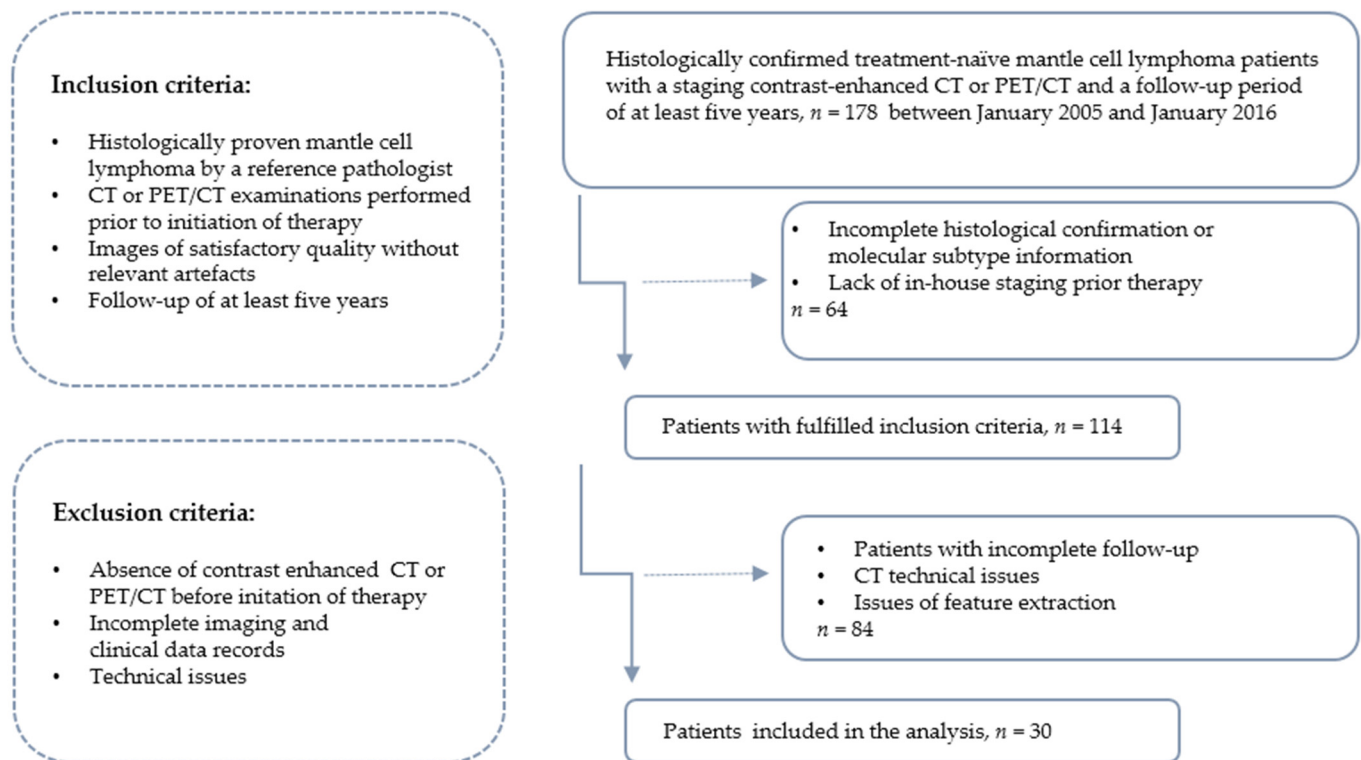


Figure 1. Recruitment pathway of the study.

Contrast-enhanced PET/CT or CT images for assessment of disease status were performed close to therapy initiation. The images used for analysis were acquired on the multiple-row detector CT scanners Philips Brilliance CT 64-channel scanner or Philips Brilliance iCT 256-channel scanner (Philips Healthcare, Cleveland, OH, USA) and the PET/CT scanner Siemens Biograph mCT(40)S (Siemens Healthineers, Erlangen, Germany). Because most MCL patients in our retrospective study had been enrolled in clinical trials, such as the MCL R2 Elderly trial (NCT01865110), or had been treated according to them, all CT and PET/CT imaging used the acquisition and reconstruction parameters according to the standard protocol:

CT scans were performed after intravenous contrast agent injection of Ultravist® 370 (Bayer Schering Pharma, Berlin, Germany) in a weight-adopted dose with a delay of 70–80 s to represent the portovenous phase of chest and abdomen (tube voltage 100–120 kV with automatically calculated tube current; collimation 40×0.625 mm and 64×0.625 mm; 1 mm reconstruction thickness, increment 0.5 mm, matrix 512×512). Thin-slice CT images were reconstructed using a soft tissue kernel (filter, B31f) for further analysis.

^{18}F -FDG-PET/CT scans were performed approximately 60 min after intravenous administration of ^{18}F -FDG with a mean of 282 MBq of (range 223–334 MBq). The contrast-enhanced CT image was acquired in the portal venous phase 70–80 s after intravenous contrast agent injection (Ultravist 370, Bayer Schering Pharma, Berlin, Germany) in a weight-adopted dose. Attenuation-based online modulation of tube current (CARE Dose) was used with a quality reference tube current setting (reference mAs) of 210 mAs (tube voltage 120 kV; 16×1.2 -mm collimation; 4 mm slice thickness; matrix 512×512) followed by the PET scan from the mid-thighs to the base of the skull in 5 to 8 bed positions.

All PET scans were acquired in 3D mode with an acquisition time of 3 min per bed position in the time of flight technique. PET data were reconstructed with attenuation correction using dedicated standard software (PETsyngo, Siemens, Erlangen, Germany).

2.2. Radiomic-Based Machine Learning as Prediction Model

2.2.1. Tumour Segmentation, Data Processing and Feature Extraction

Two board-certified radiologists analysed all images, both with over 10 years of experience in oncologic imaging and over 5-year experience in texture analysis. Consent was established by joint consultation in cases of disagreement. Target lymphoma lesions were selected based on Cheson criteria [40]. Segmentation, texture analysis and feature extraction were performed using the software mint Lesion™ (mint Medical GmbH, Heidelberg, Germany), which allows the three-dimensional size and whole lesion radiomic measurements.

All target lymph nodes were segmented in a semiautomatic process by manually delineating superior and inferior tumour boundaries. The software algorithm computed the tumour boundaries on the slices in between for “whole tumour volume” data. The final 3D segmentation was thoroughly reviewed, and if necessary, the tumour boundaries were manually modified. In the final processing step, radiomic features were quantified concerning their characteristic pattern of grey levels within the ROI using texture feature descriptors according to the Image Biomarker Standardization Initiative (IBSI) guidelines [38]. There were 22 texture features of first-order and 26 features of second-order derived from the grey-level co-occurrence matrix were chosen for further analysis (see Table A2 in Appendix A). First-order statistics are histogram-based methods that describe the distribution of values of individual voxels and reduce a region of interest to single values for mean, median, maximum, minimum, uniformity and others. Second-order statistics consider spatial relationships as they describe statistical interrelationships between voxels with similar or dissimilar contrast values. Image-based texture analysis was first introduced in 1973 by Haralick et al. [61].

The 3D volumetric radiomic features were extracted and used as input volume for machine learning model development. See Figure 2 for a depiction of the workflow.

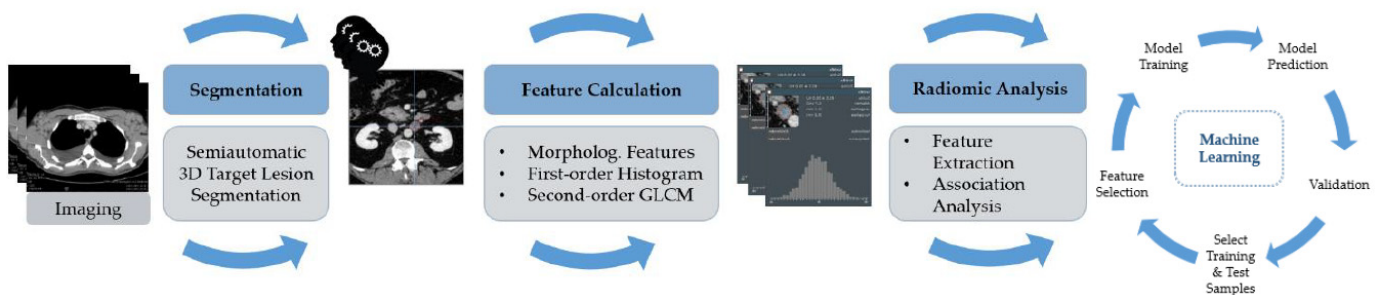


Figure 2. The schematic diagram for processing and analysis in the machine learning approach. Morphological features, including volume and textural features of first and second-order, were obtained. Details regarding the extraction settings are listed in Appendix A, Table A1.

2.2.2. Features Selection

After preprocessing, the number of features was reduced using a filtering-based feature selection. Like any other data-mining application, radiomics underlies the curse of dimensionality [62], and an appropriate feature selection strategy is crucial to reduce the great number of initially extracted quantitative radiomic features. By choosing a subset of optimal features, overfitting is reduced, and thus models experience increased generalizability for simpler and faster ML models with enhanced performance [63]. After preprocessing, feature selection using filter feature selection techniques was applied [63]. Like any other data-mining application, radiomics underlies the curse of dimensionality [62], as it extracts a large number of quantitative features from the regions of interest (ROIs).

An appropriate feature selection strategy is crucial to reduce the dimensionality of radiomic features; by choosing a subset of optimal features, overfitting is reduced, and thus models experience increased generalizability for simpler and faster ML models with enhanced performance [63].

For this study, RelieFF—a well-known and often utilised feature selection method—was used. The top 50 features were chosen as the input variable for further analysis, as shown in the correlation plot in Figure 3. The calculations were done using the open-source software licensed under the GNU General Public License Weka Toolkit (version 3.8, University of Waikato, Waikato, New Zealand) [64–70].

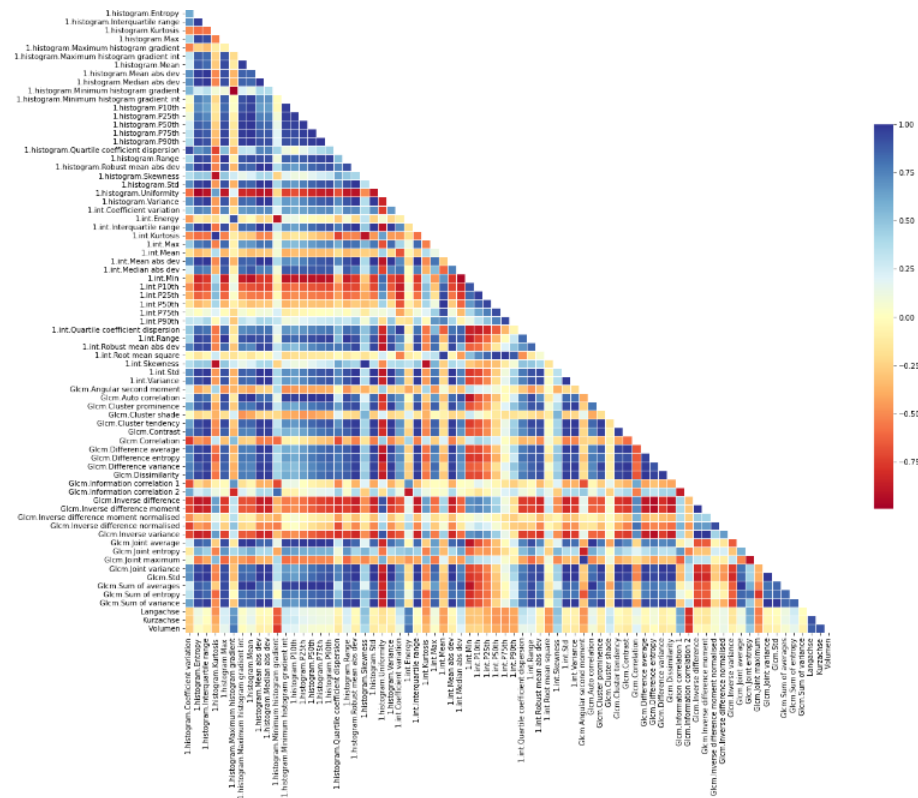


Figure 3. Correlation plot depicting the hierarchical clustering of all extracted features and their discriminatory power between the relapse and the remission group.

2.2.3. Dataset Characteristics and Preprocessing

The dataset consisted of 134 ROIs in the CT examinations of the 30 MCL patients (17 patients with complete remission and 13 patients with relapse of disease). Data were randomly split into training (70%) and testing sets (30%) but stratified to ensure consistent distribution in the training and testing sets to mimic the real-world distribution and reduce overfitting. The splitting was done considering the patient level, i.e., no patient contributed to training and testing.

The selected radiomic features of the 134 ROIs were the input variables. The output variables were binary: The patients who suffered from a relapse of disease within the 5-year observation period were assigned to the high-risk group, whereas those with complete remission were classified in the low-risk group. The selected features were used for the supervised machine learning model, including RF and KNN.

2.2.4. Machine Learning (ML) Classification Architectures

Our machine learning pipeline used two classic machine learning algorithms—Random Forest (RF) and K-nearest Neighbor (KNN) architectures. The medical imaging community has proven these methods efficient and suitable [71–73]. The algorithms were implemented using the open-source Python machine learning library Scikit-learn [74].

The Random Forest algorithm (RF) [75,76] is an ensemble classifier that produces multiple decision trees using randomly selected subsamples of the data set. The prediction is achieved by averaging the predictions of all decision trees. This improves the shortcomings

of individual decision tree algorithms—namely, weak generalisation and a tendency to overfitting. We used the default RandomForestClassifier from sklearn.

The K-Nearest Neighbors algorithm (KNN) [77] is also known as “Instance-based” learning as the hypotheses are built from the training instances. KNN is based on distance calculations between instances: Target data points are classified according to the labels of the closest neighbours from the training data. KNN is a conceptually simple but powerful algorithm that is easy to use, interpret and implement. We used the grid search optimisation algorithm to fine-tune the model and selected five neighbours in our study.

2.2.5. ML Model Development and Evaluation

RF and KNN algorithms were trained and selected by comparing performance on the test set’s evaluation. Each algorithm used ten seeds for the training and testing split to ensure inter-model reproducibility. Results of all runs were averaged and standard deviation and confidence intervals were calculated.

All computing was done on a single workstation running Ubuntu 20.04 LTS with 32 GB memory, AMD Ryzen 7 3700X (AMD, Sunnyvale, CA, USA) and NVidia RTX 2070 Super 8 GB VRAM (Santa Clara, CA, USA). The following software modules were installed: Scikit-learn 0.23.3, Python 3.8.5, JupyterLab 2.2.6. (Further details see below).

2.3. Neural Network Approach as Prediction Model

Convolutional neural networks (CNNs) are deep learning techniques useful for diagnosis and image analysis tasks [78,79]. In general, a CNN consists of a series of layers of convolution filters followed by fully connected layers at the very end of the network. In contrast to traditional ML models, CNNs do not require the quantification and selection of radiomic features as they are trained directly on the images.

Three different network architectures were trained and evaluated—an optimised 3D CNN designed by ourselves and two larger state-of-the-art 3D CNNs, consisting of a 3D DenseNet and a 3D Squeeze-and-Excitation ResNet 50 (3D SE ResNet). Details of the three networks are described in the “network architecture” section. See Figure 4 for a depiction of the workflow.

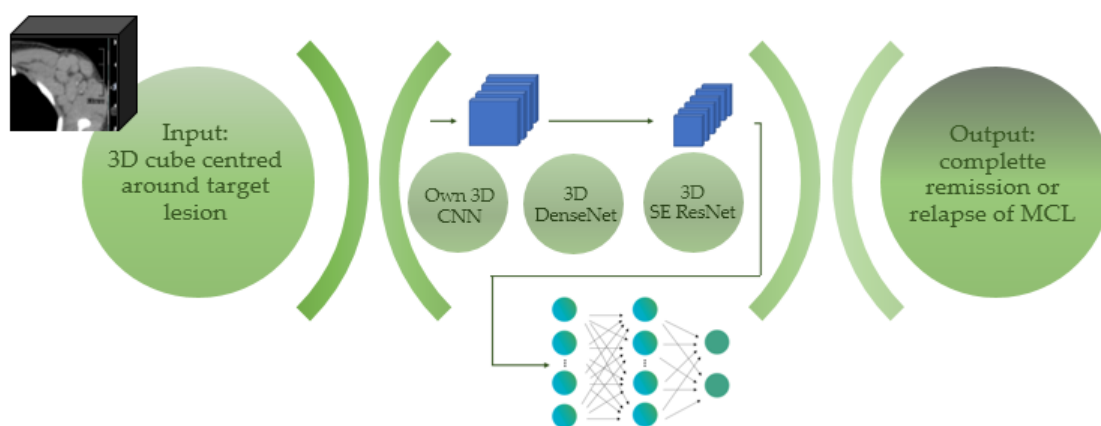


Figure 4. The schematic diagram for data processing in deep neural networks.

2.3.1. Dataset, Training and Validation

The dataset consisted of CT examinations containing 134 ROIs of 30 MCL patients. ROIs were dichotomised into MCL in complete remission (CR) and MCL with relapse of disease (RD); 35% were randomly selected for testing, while the remaining 65% were used for training and validation. The training set was utilised for training the network, while the validation set was used to monitor the effectiveness of the training. ROIs of one patient were either in the training, validation or testing group, but never in more than one at a time.

At first, 3D cubes of each target lymphoma lesion with the tumour's centre of mass at the centre of the cube were cropped from CT images. After evaluating the range of target sizes, we determined one cube measuring $20 \times 80 \times 70$ mm to fit each target lesion completely within the cube without overlapping a potential neighbouring target lesion.

The original dataset was preprocessed using the above-described methods to generate 3D patches of $20 \times 80 \times 70$ pixels focusing only on the ROI. The adaptive moment estimation (Adam) optimiser was used to minimise the cross-entropy loss between the model outputs and target classification labels. Adam optimiser was configured with a learning rate of 0.001 and a dropout of 30% to prevent overfitting of our model. The loss of the output class "relapse of disease" was double weighted. Due to the limited amount of training data in our data set, we used augmentation techniques for image processing "Random Flip", "Gaussian Blur" and "Gaussian Noise" with Torchvtk (<https://github.com/torchvtk/torchvtk>; last accessed 15 January 2022) to enhance the robustness and to prevent overfitting [78].

Additionally, this ensures the model focuses on the target lesions rather than various sources of noise [80]. We stopped the training after 100 epochs.

The duration of training was 2 h per run, with a batch size of 4 determined by GPU memory. After every training epoch, the network was evaluated with the validation dataset and the model from the epoch with the best validation score was used for classifying the test set. We tested the classification performance of all three network architectures with five runs with five different seeds for the training and validation dataset.

We evaluated the DL models by accuracy, precision, recall, F1-Score and the area under the curve (AUC) by calculating the mean and standard deviation over the five runs. To assess the clinical value of the predictive models, we performed a decision curve analysis.

The following software modules were installed: PyTorch 1.9, Monai 0.6, Python 3.9.5, JupyterLab 3.0.16, NumPy 1.21.0. All code, from data preprocessing to model building, was written in Python 3.9.5 based on the open-source deep-learning library "PyTorch", together with the open-source framework for deep learning in medical imaging "Monai" (<https://monai.io/>; last accessed 15 January 2022).

Neural Networks were trained with CT images on a single workstation running Ubuntu 20.04 LTS with 32 GB memory, AMD Ryzen 9 3900X (AMD, Sunnyvale, CA, USA) and NVidia RTX 3080 10 GB VRAM (Santa Clara, CA, USA).

2.3.2. Deep CNN Architecture

Following are the architectural design details of the model used:

3D DenseNet: We used the 3D DenseNet-121 [81] from Monai [Medical Open Network for Artificial Intelligence MONAI <https://monai.io/>; last accessed 15 January 2022]. Densely connected convolutional networks (short: DenseNets) consist of several dense blocks. Each layer in one dense block is connected to all other layers with feed-forward connections, thus requiring fewer parameters than traditional CNNs. Advantages are reduced overfitting for smaller training sets and improved vanishing gradient problems.

3D Squeeze-and-Excitation ResNet 50 (SEResNet50): We used the 3D Squeeze-and-Excitation ResNet SEResNet50 [82] from Monai [Medical Open Network for Artificial Intelligence MONAI <https://monai.io/>; last accessed 15 January 2022].

Residual neural networks (short: ResNets) [83] use "skip connections" to improve the vanishing gradient problem. In Wightman, Ross et al. [84], they are described as the "gold-standard architecture" for image classification tasks. Squeeze-and-Excitation Networks [82] add so-called "Squeeze-and-Excitation blocks" to existing convolution neural networks like the ResNets to improve channel interdependencies and outperform standard ResNets in image classification tasks. It was found that Squeeze and Excitation (SE) blocks combined with ResNet architecture performed outstandingly well in classifying brain tumours [83].

Own 3D CNN: Our developed network consists of a 3D CNN with two 3D convolutional layers as a backbone with two fully connected downstream layers. The first layer consists of one input channel for the 3D image, eight output channels (kernel size 3×3 , stride 1) and a Relu activation function. The second layer has 16 output channels (kernel

size 3×3 , stride 1, 3D Dropout) and a Relu activation. The first fully connected layer reduces the feature size to 60 and the second one to the two output classes, “complete remission of disease” and “relapse of disease”. The output is passed through a softmax activation function.

3. Results

3.1. Patient Characteristics

A total of 30 consecutive patients with biopsy-proven MCL (4 women and 26 men; mean age 62.2 ± 9.7 years, range 42–76) met the criteria for participation in the study. An average of 4.4 lymph nodes per MCL patient (range 3–6) were analysed.

In total, 134 target lesions were evaluated. In this cohort, 17 patients (57%) were responders and in complete remission. 13 patients (43%) responded to first-line treatment but relapsed within the five year observation period.

All patients’ baseline clinical characteristics are summarised in Table 1.

Table 1. Baseline demographic and clinical data.

Characteristic	Number
Sex	
Male	26 (86.7%)
Female	4 (13.3%)
Average age (range)	62.2 ± 9.7 years (42–76)
Ann Arbor Stage	
Stage I	0
Stage II	2 (6.7%)
Stage III	5 (16.7%)
Stage IV	23 (76.7%)
Patients’ status in 5-years follow up	
Complete remission (CR)	17 (57%)
Relapse of disease (RD)	13 (43%)

3.2. Radiomic Analysis and the Machine Learning Prediction Model

Several traditional measurements of model performance were used, including sensitivity, precision, F1-Score, accuracy, ROC curve and the area under the ROC curve (AUC), to assess and compare the ML models. We used a principal component analysis (PCA) for dimensionality reduction while maintaining most of the variation in data to improve the accuracy of the predictive model. We used the grid search optimisation algorithm for fine-tuning and selected five components in our PCA.

Accuracy measures the model’s robustness and is defined by the percentage of correctly identified labels out of the entire population. Precision—the positive predictive value—is the probability that a predicted true label is indeed true. Sensitivity—the true positive rate (TPR) or recall—is the percentage of correctly identified true class labels. The F1-score is the harmonic mean of the sensitivity and precision.

The ROC curve is computed by plotting the recall versus the false positive rate at different decision thresholds. The AUC constructed from the ROC is the ability of the model to separate classes, with the number 1 standing for perfect separation and the number 0.5 standing for random classification. All tests were two-sided; $p < 0.05$ was considered statistically significant.

The predictive performances of the radiomics-based ML models to predict relapse of MCL disease are summarised in Table 2. The different models’ overall accuracy of predicting relapse was 61% (range: 58–64%, AUC = 0.58–0.62). The PCA KNN showed the best prediction accuracy.

Figure 5 shows the receiver operating characteristic (ROC) curves for the mean over the five runs. PCA KNN and KNN outperform the RF algorithm.

Table 2. The outcome of the radiomics-based machine learning models.

Machine Learning Models	Accuracy	Precision	Sensitivity	F1-Score	AUC
KNN	0.63 ± 0.02	0.64 ± 0.01	0.63 ± 0.02	0.62 ± 0.02	0.62 ± 0.02
PCA KNN	0.64 ± 0.02	0.64 ± 0.01	0.63 ± 0.02	0.62 ± 0.02	0.62 ± 0.02
RF	0.47 ± 0.07	0.50 ± 0.09	0.47 ± 0.07	0.45 ± 0.08	0.49 ± 0.07
PCA RF	0.58 ± 0.02	0.61 ± 0.04	0.58 ± 0.02	0.58 ± 0.02	0.58 ± 0.02

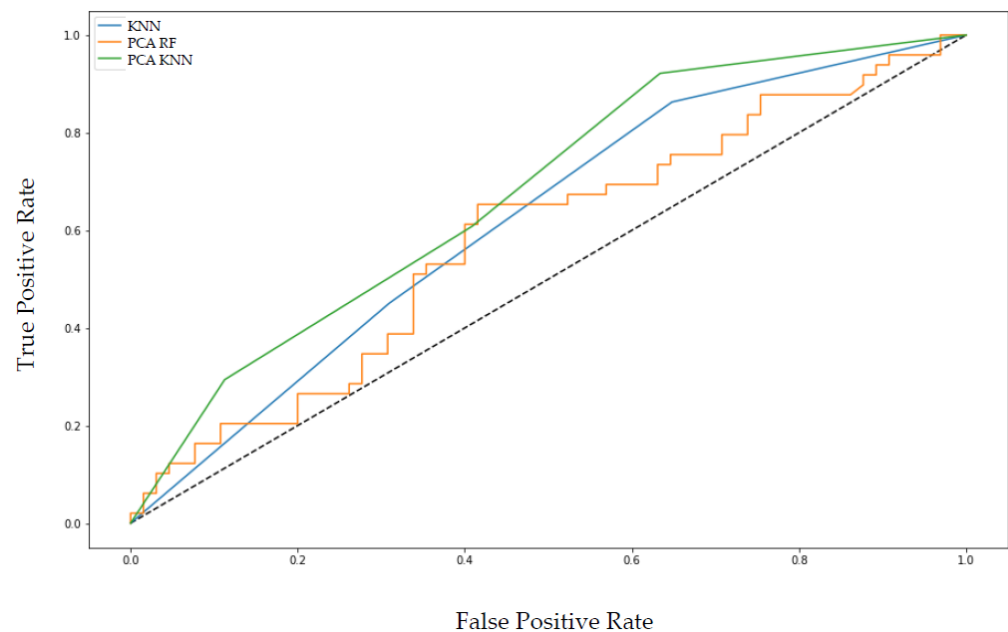


Figure 5. The receiver operating characteristics (ROC) curves of the KNN and PCA KNN algorithm yielded an area under the curve (AUC) of 0.62. In contrast, PCA RF yielded an AUC of 0.58 for classification performance of complete remission versus relapse of MCL.

We performed a decision curve analysis to assess the clinical value of the predictive models in terms of net benefit against threshold probability. A decision curve analysis graph with threshold probability on the x-axis and net benefit on the y-axis illustrates the trade-offs between true positives and false positives, representing benefit and harm, respectively, as the threshold probability varies [13].

Figure 6 shows the decision curve analysis for the best run of the best performing machine learning model using PCA KNN.

3.3. The CNN-Based Neural Network Approach as Prediction Model

A total of 134 cubes were included for training. The results from five runs with five different seeds for the training and testing split were used for further comparative analysis. We evaluated the five scores' accuracy, precision, recall, F1-Score and the area under the curve (AUC) by calculating the mean and standard deviation over the five runs.

The DL models' overall accuracy of predicting relapse was 64% (range: 59–70%, AUC = 0.58–0.70). Based on the ROC analysis, our own 3D CNN significantly outperformed the 3D DenseNet and the 3D SEResNet50. Accuracy, Precision, Recall, F1-Score and AUC for the Own 3D Network is around 0.7 for the classification task.

The performances of the deep CNN models to predict the relapse of MCL disease are summarised in Table 3.

Figure 7 shows the receiver operating characteristics (ROC) curve for the mean over the five runs. For a False-Positive-Rate below 0.1 (= very high specificity), the 3D ResNet outperforms the Own 3D Network. For a False-Positive-Rate above 0.1, the own 3D network equals or mostly outperforms the 3D ResNet.

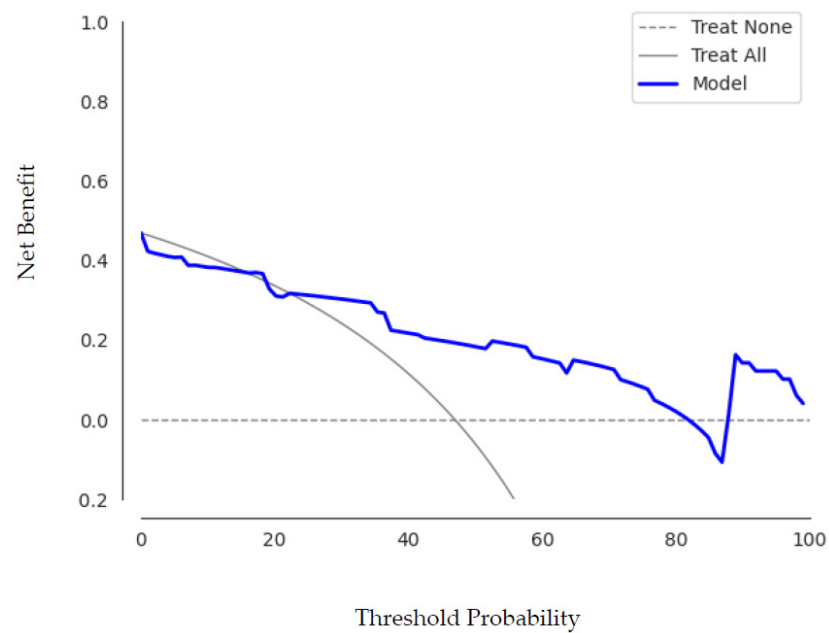


Figure 6. The decision curve analysis for the best run of PCA KNN shows that within a threshold probability from 25% to 55%, checking patients based on the classification model leads to a higher benefit than checking no or checking all lymphomas more frequently.

Table 3. The outcome of the deep learning models.

Deep Learning Models	Accuracy	Precision	Sensitivity	F1-Score	AUC
Own 3D Net	0.70 ± 0.02	0.71 ± 0.02	0.70 ± 0.02	0.69 ± 0.01	0.70 ± 0.04
3D DenseNet	0.59 ± 0.05	0.64 ± 0.07	0.59 ± 0.05	0.57 ± 0.06	0.58 ± 0.13
SEResNet50	0.62 ± 0.04	0.65 ± 0.07	0.62 ± 0.05	0.60 ± 0.04	0.62 ± 0.06

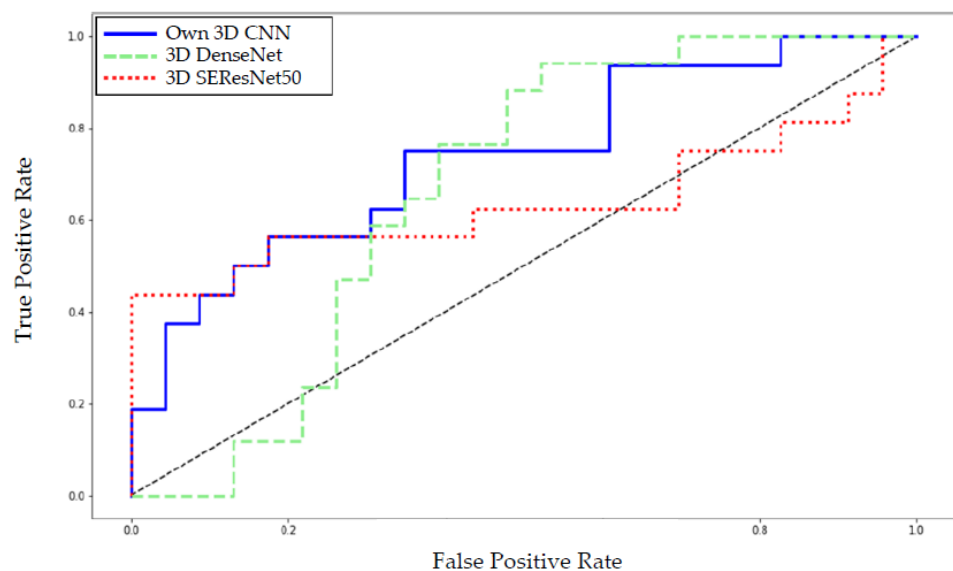


Figure 7. The deep learning models’ receiver operating characteristics (ROC) curve shows our 3D network mostly outperforming the 3D DenseNet and 3D SEResNet50.

Analogous to the machine learning approach, we performed a decision curve analysis to assess the clinical value of the predictive models. Figure 8 illustrates the trade-offs between true positives and false positives in the decision curve analysis for the best run of the best performing deep learning model using our own 3D CNN.

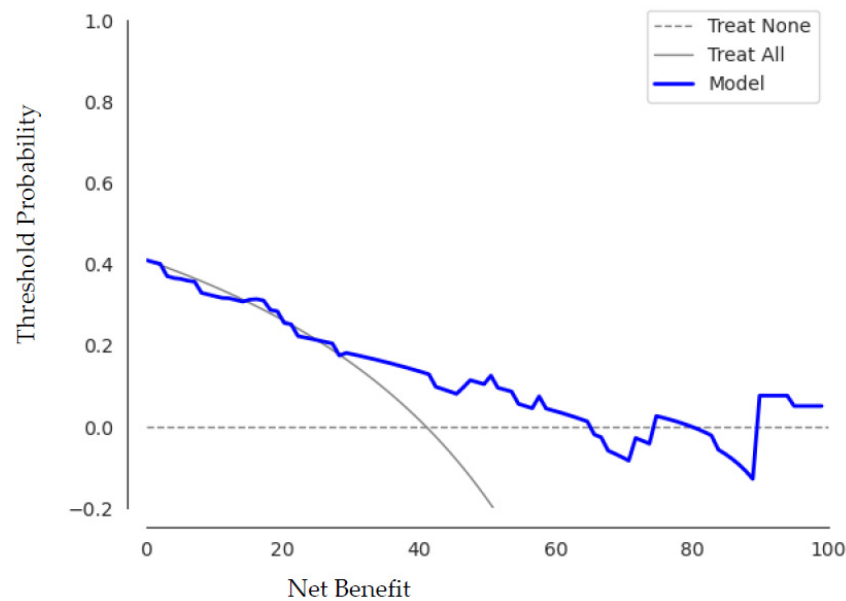


Figure 8. The decision curve analysis for the best run of the novel 3D CNN shows that within the threshold probability from 30% to 65% checking patients based on the classification model leads to a higher benefit than checking no or checking all lymphomas more frequently.

3.4. Comparing the Deep Neural Network with the Machine Learning Approach

Figure 9 illustrates the metrics grouped by each data set and presents them as grouped bar charts for clarity and ranking of the different deep neural networks and machine learning models.

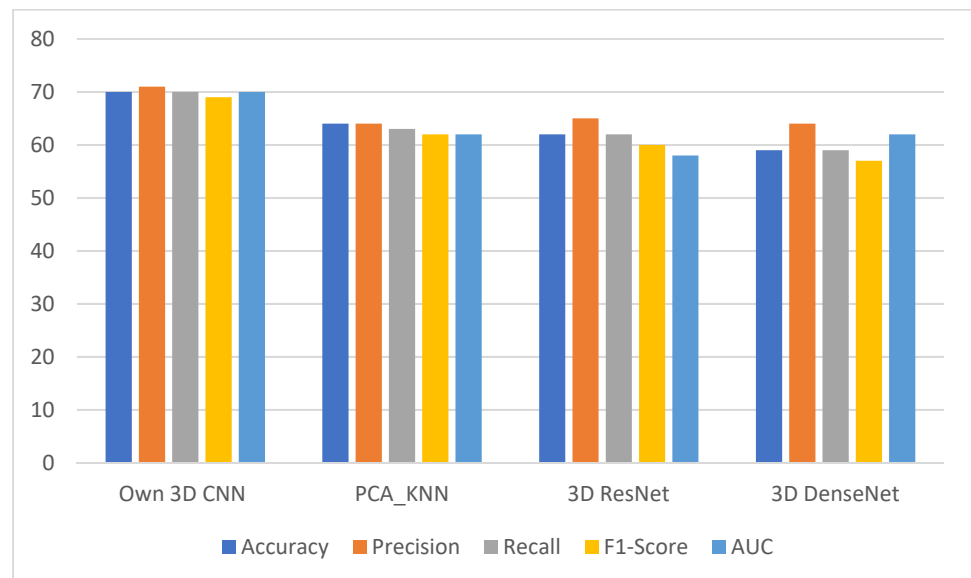


Figure 9. The grouped bar chart shows that our own 3D CNN has the highest performance among all prediction models.

Our own 3D CNN provided the best performance (AUC, accuracy, precision, sensitivity and F1-score were 0.70 ± 0.04 , 0.70 ± 0.02 , 0.71 ± 0.02 , 0.70 ± 0.02 and 0.69 ± 0.01 , respectively) among all deep and machine learning prediction models.

The deep learning model 3D SEResNet50 (AUC, accuracy, precision, sensitivity and F1-score were 0.62 ± 0.06 , 0.62 ± 0.04 , 0.65 ± 0.07 , 0.62 ± 0.05 , 0.60 ± 0.04 , respectively) and the 3D DenseNet model performed worse in the comparison (AUC, accuracy, preci-

sion, sensitivity and F1-score were 0.58 ± 0.13 , 0.59 ± 0.05 , 0.64 ± 0.07 , 0.59 ± 0.05 and 0.57 ± 0.06 , respectively).

The machine learning model using KNN/PCA KNN showed the second-best result of all models (AUC, accuracy, precision, sensitivity and F1-score were 0.62 ± 0.02 , $0.63 \pm 0.02/0.64 \pm 0.02$, 0.64 ± 0.01 , 0.63 ± 0.02 and 0.62 ± 0.02 , respectively).

The PCA RF/RF-based ML models scored the worst (AUC, accuracy, precision, sensitivity and F1-score were 0.58 ± 0.02 , 0.61 ± 0.04 , 0.58 ± 0.02 , 0.58 ± 0.02 and 0.58 ± 0.02 , and 0.49 ± 0.07 , 0.47 ± 0.07 , 0.50 ± 0.09 , 0.47 ± 0.07 and 0.45 ± 0.08 , respectively).

4. Discussion

Mantle cell lymphoma is an aggressive disease associated with early relapse, chemorefractory disease and poor long-term survival in most patients [13]. Standard of care emphasises aggressive treatment approaches, demonstrating the most prolonged durable remissions [7]. However, MCL has increasingly been considered a highly heterogeneous disease with subtypes with favourable disease features in which observation can be considered [1,85,86]. MCL is a “genomically unstable” malignancy associated with (sub)clonal heterogeneity and modulation of the genetic profile over time [15]. No routine biomarkers are currently established to predict short-term clinical outcomes. Radiomics refers to a fast-growing research area that converts standard-of-care imaging into minable high-dimensional data and builds subsequent predictive models to stratify patients into different risk groups based on the risk of occurrence of clinical endpoints, such as relapse of disease and personalise treatment [11,12,87,88].

As radiomic data convey information about tumour biology, such as temporal and spatial heterogeneity, they potentially reflect tumour behaviour and aggressiveness [26,88–91]. As AI techniques can analyse the tumour lesion as a whole and can be applied at multiple sites without the risk of a sampling error and biopsy-related complications, there is great potential for being used as a “virtual biopsy”.

One of the challenges in the classic machine learning algorithms relies on the several preprocessing steps, such as tumour segmentation, which regularly requires manual correction of the tumour boundaries computed by the (semi)automated algorithms which increases cost, time and the risk of inter-observer variation [92]. Deep learning can automatically perform these steps on the raw data and is considered a powerful analytical tool for different predictive data mining applications, especially in complex processes like biological systems [93–96]. Machine and deep learning-based artificial intelligence (AI) imaging techniques have the potential to support the decision making in clinical oncology for precise imaging. Of great clinical interest is the potential additive value of radiomics at baseline staging for early identification of high-risk patients to select the optimal therapeutic management and improve prognosis initially using the standard of care test, a simple contrast-enhanced CT. Moreover, radiomics holds the promising potential for disease monitoring, offering information related to disease evolution.

However, the current black-box approach makes it less acceptable for clinical application, particularly deep learning. To enhance the translation of radiomics into clinical practice, the reproducibility and stability of the features against technical perturbation are essential elements [97]. Unfortunately, there is often not one best-performing model from a statistical standpoint. Different classifiers can lead to models performing only slightly worse than or statistically similar to the best-performing model because radiomic features depend highly on acquisition parameters, hindering generalizability [98].

At best, the results should be validated using independent data sets, preferably using data from a different institution [26,99]. On a multi-institutional scale, a large amount of available medical imaging data exists, which constitutes a great opportunity for complex model training. But access for research purposes is highly restricted by law and ethics. And although the imaging biomarker standardisation initiative and quantitative imaging biomarkers alliance for standardisation give recommendations including postprocedural harmonisation and standardisation of the acquisition protocol for improved reproducibil-

ity [38,100], these methods are difficult to implement, particularly across multiple sites and concerning the constant innovations in medical imaging techniques [38,101,102]. Federated or distributed learning might be a potential solution in the future to perform local model development and training and to share the weights for collaboration [103]. According to the literature, internal validation techniques like ours can be used in small-scale pilot studies [104].

Most radiomics related studies focus on lymphoma classification, are PET/CT based and centre around Hodgkin lymphoma or diffuse large B cell lymphoma (DLBCL) [41–45,47–49,53,105]. However, there is no standard way to extract radiomics features, meaning reproducibility is a key challenge in this field [106].

Deep learning algorithms can detect or classify lymphoma; in line with radiomics studies, most rely on PET/CT imaging: Weisman et al. used 3D neural networks to automate lymph node detection and lymphoma assessment in PET/CT [107]. Li et al. used PET/CT for lymphoma characterisation [79] and Sibille et al. used DL to classify positive FDG uptake regions on PET/CT images as lymphoma or lung cancer [81]. Several studies address automated segmentation of lymphoma, especially in diffuse large B-cell lymphoma (DLBCL), which typically consists of larger lesion sizes and more intense signals on PET images than MCL [82,83]. Prognostication modelling could potentially be exploited to determine new and appropriate personalised medicine treatments [12]. However, only a few studies compared radiomics-based versus deep learning methods, for example, Bibault et al. regarding outcome prediction of rectal cancer [84].

There is little data regarding (AI)-based image analysis on MCL, which all used PET/CT imaging. Mayerhoefer et al. showed that PET/CT-derived radiomic features combined with clinical parameters lead to improved PFS prognostication [29]. Mayerhoefer et al. also deployed a machine learning approach for detecting bone marrow involvement of MCL on pelvic PET/CT images [108]. Zhou et al. used CNN architecture [109] to segment mantle cell lymphoma (MCL) in PET/CT images.

Also, the role of baseline semiquantitative 18F-FDG PET/CT parameters such as SUVmax for prognostication of MCL is not yet clear with controversial results [110–114]. Limitations are the heterogeneity between the study cohorts, such as therapy and baseline features. In the recent LyMa-PET Project, Bailly et al. demonstrated that SUVmax was the only independent prognostic factor for PFS and OS and suggested a scoring system combining MIPI and SUVmax for outcome prediction [112]. Albano et al. studied other baseline metabolic parameters like SUVbody surface area and demonstrated no correlation with PFS and OS, while for OS, only the MIPI score resulted as an independent prognostic factor in contrast to other common variables, including the Ki-67 score without a significant association with outcome survival [110]. So far, no strong evidence about the potential prognostic impact of 18F-FDG PET/CT on survival outcomes is available and more studies are needed to confirm or contradict these results.

To our knowledge, this is the first study that used two machine learning and three deep learning algorithms to investigate their potential as predictive models for early prediction of relapse of MCL. In this study, we compared the performance of RF and KNN—two traditional machine learning classifiers with three deep learning convolutional neural networks (3D Dense Net, 3D SEResNet50 and our own 3D CNN) to predict future relapse of MCL.

Our own 3D Net (AUC, accuracy, precision, sensitivity and F1-score were 0.70 ± 0.04 , 0.70 ± 0.02 , 0.71 ± 0.02 , 0.70 ± 0.02 and 0.69 ± 0.01 , respectively) provided the best performance among all deep and machine learning prediction models with a low variance in the five runs speaking for the stability of our model.

The sensitivity is a significant parameter for models' evaluation as it reports how well the predictive model can detect MCL patients at high risk of relapse of disease.

The deep learning model 3D SEResNet50 (AUC, accuracy, precision, sensitivity and F1-score were 0.62 ± 0.06 , 0.62 ± 0.04 , 0.65 ± 0.07 , 0.62 ± 0.05 and 0.60 ± 0.04 , respectively) and the 3D DenseNet model performed worse in the comparison (AUC, accuracy, preci-

sion, sensitivity, and F1-score were 0.58 ± 0.13 , 0.59 ± 0.05 , 0.64 ± 0.07 , 0.59 ± 0.05 and 0.57 ± 0.06 , respectively). An explanation could be that the SEResNet50 and DenseNet have more learned variables, potentially leading to faster overfitting amid the input cubes of $20 \times 80 \times 70$ pixels.

Interestingly, the machine learning model using KNN/PCA KNN showed the second-best result of all models (AUC, accuracy, precision, sensitivity and F1-score were 0.62 ± 0.02 , 0.63 ± 0.02 / 0.64 ± 0.02 , 0.64 ± 0.01 , 0.63 ± 0.02 and 0.62 ± 0.02 , respectively).

The PCA RF/RF-based ML models scored the worst (AUC, accuracy, precision, sensitivity and F1-score were 0.58 ± 0.02 , 0.61 ± 0.04 , 0.58 ± 0.02 , 0.58 ± 0.02 and 0.58 ± 0.02 , and 0.49 ± 0.07 , 0.47 ± 0.07 , 0.50 ± 0.09 , 0.47 ± 0.07 and 0.45 ± 0.08 , respectively) although random forests (RF) have displayed high predictive performance in several other biomedical applications.

The AUC results focus only on predictive accuracy and thus cannot prove a model is worth using. Metrics that concern accuracy does not incorporate information on consequences. Decision analysis incorporates consequences and thus can tell whether a model is worth using to identify the optimal prognostic model that maximises the outcome of interest [115].

In our case, a false-negative result—a patient being put in a low-risk group resulting in fewer follow-ups or missed maintenance therapy- is more harmful than a false-positive result resulting in closer follow-up. A model with much greater specificity but lower sensitivity than another would have a higher AUC yet yield a smaller net benefit and be a poorer choice for clinical use.

Prognostication modelling could potentially be exploited to determine new and appropriate personalised medicine treatments [12]. With the exponential increase in biomarkers and information technology advancements, prognostic and diagnostic models will likely increase. Reliable, comparable and easily understandable methods to determine the value of such models are required [116].

The number of studies on diagnostic and prognostic markers using accuracy metrics dwarfs those using decision analysis methods in the medical literature. Decision analytics is a great statistical tool without requiring further information, such as the costs or effectiveness of treatment. It can be a beneficial method directly applied to a model validation dataset.

The decision curve for our best-performing 3D CNN proved its value. It shows that between a threshold probability of 30% and 65%, checking patients based on our proposed classification model leads to a higher benefit than checking no or checking all lymphomas more frequently.

Although seemingly promising, this study is affected by several limitations that need further discussion. We note that the small sample size ($n = 30$ pts) did not allow for multiple testing corrections for the large number of radiomic features tested. However, statistical analyses were performed lesion-based ($n = 134$) as 3D-volumetric ROIs; To address the issue of the limited sample size, we applied recommended used techniques such as the bootstrap technique or data augmentation. Although our results showed sufficient statistical power, our proof-of-concept study remains hypothesis-generating, requiring further prospective validation as with other similar published studies before a wider adaptation into a clinical decision process.

Secondly, the nature of the study cohort: MCL is a less common lymphoma subtype than Hodgkin lymphoma or diffuse large B cell lymphoma (DLBCL) with smaller patient populations and more difficulties obtaining a sufficient number of cases with the event of interest. Furthermore, our cohort consisted of treatment-naïve MCL patients with in-house imaging and tissue confirmation at baseline and a documented five year follow up. MCL patients with progression were excluded for getting a distinct signature of relapse or remission. Including MCL patients with disease progression or a reference group with benign lymphadenopathy could be part of further analyses. In addition, this study was retrospective, raising potential patient selection biases as treatment was not randomly

assigned to participants. Keeping this in mind, all extracted study data were randomly split into training and testing sets but stratified to ensure consistent distribution and mimic the real-world distribution of remission and relapse of disease. Thirdly, study data were collected from a single centre without external validation. Radiomic features depend highly on acquisition parameters, hindering generalizability [98]. But the reproducibility and stability of the features against technical perturbation are a prerequisite for developing quantitative imaging biomarkers used in clinical routine. Although there exist recommendations by the imaging biomarker standardisation initiative and quantitative imaging biomarkers alliance for standardisation [38,100], including postprocedural harmonisation and standardisation of the acquisition protocol for improved reproducibility, these methods are difficult to implement, particularly across multiple sites and concerning the constant innovations in medical imaging techniques [38,101,102].

Developing a prognostic model is a key tool for individualised therapy of lymphoma patients. Our pilot study compared the most popular classic machine learning techniques with advanced deep learning methods. While our proposed deep learning 3D CNN prediction model requires further improvement, we consider our preliminary results as a promising methodological basis for more extensive studies, ideally prospective and multicentric, that involve comparing and combining AI-based imaging models with the established MIPI score for improved staging and prognostication at diagnosis and as real-time surveillance for improved clinical decision-making amid innovative treatment strategies to pave the way for precise imaging, and thus improving the standard of care for mantle cell lymphoma.

5. Conclusions

This study demonstrates the application of deep neural networks and radiomics-based machine learning models to predict relapse in MCL on baseline CT studies.

The results revealed that (1) the own 3D CNN showed the best predictive performance compared to all deep and machine learning models (2) the machine learning model using KNN/PCA KNN showed the second-best results, superior to the 3D SEResNet50 and the 3D DenseNet and (3) the decision curve analysis proved our optimised 3D CNN a valuable predictive model. Further application of these predictive models might supplement future approaches for refined MCL diagnosis and assist workflows for personalised cancer medicine.

Author Contributions: Conceptualisation, C.S.L. and C.G.L.; methodology, C.S.L., C.G.L. and M.G.; software, M.F.M., D.W. and M.G.; validation, M.G.; formal analysis, C.S.L. and C.G.L.; investigation, C.S.L., C.G.L. and S.A.S.; resources, C.S.L. and S.A.S.; data curation, C.S.L. and C.G.L., writing—original draft preparation, C.S.L., C.G.L. and M.G.; writing—review and editing, E.T., W.M.T., A.J.B., S.S. and S.A.S.; visualisation, C.S.L. and C.G.L., supervision, A.J.B. and M.B.; project administration, C.S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the ethics committee of the medical faculty of the University of Ulm (protocol code 155/18, 25 April 2018).

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

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy restrictions.

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

Appendix A

Table A1. Settings of the feature extraction.

Setting	Determination
Bin Method	FBS
Bin Amount	1
LoG Filter	0
LoG Sigma	1
Matrix Aggregation Method	3D Average Directions
Resample Filter	0
Resample Spacing X	1
Resample Spacing Y	1
Resample Spacing Z	0
Second-Order Distance	1
Threshold Filter	0
Threshold Filter Min	−1000
Threshold Filter Max	3000

Table A2. Radiomic features used for model development.

Radiomic Features of First Order: Histogram	Radiomic Features of Second Order: Gray Level Co-Occurrence Matrix (GLCM)
Skewness	Angular Second Moment
Uniformity	Autocorrelation
Entropy	Correlation
Kurtosis	Contrast
Minimum Histogram Gradient	Energy
Maximum Histogram Gradient	Joint Average
Coefficient Variation	Sum Average
Quartile Coefficient Dispersion	Joint Maximum
P 10th (10th percentile)	Joint Entropy
P 25th (25th percentile)	Sum Entropy
P 50th (50th percentile)	Difference Entropy
P 90th (90th percentile)	Cluster Prominence
Interquartile Range	Cluster Shade
Minimum	Cluster Tendency
Maximum	Information Correlation
Mean	Information Correlation Difference
Mean Absolute Deviation	Inverse Difference
Median Absolute Deviation	Inverse Difference Normalised
Range	Inverse Difference Moment
Root Mean Square	Inverse Difference Moment Normalised
Standard Deviation	Difference Average
Variance	Difference Variance
	Dissimilarity
	Inverse Variance
	Joint Variance
	Sum Variance

References

1. Epperla, N.; Hamadani, M.; Fenske, T.S.; Costa, L.J. Incidence and survival trends in mantle cell lymphoma. *Br. J. Haematol.* **2017**, *181*, 703–706. [[CrossRef](#)] [[PubMed](#)]
2. 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. [[CrossRef](#)] [[PubMed](#)]

3. Kienle, D.; Katzenberger, T.; Ott, G.; Saupe, D.; Benner, A.; Kohlhammer, H.; Barth, T.F.; Höller, S.; Kalla, J.; Rosenwald, A.; et al. Quantitative gene expression deregulation in mantle-cell lymphoma: Correlation with clinical and biologic factors. *J. Clin. Oncol.* **2007**, *25*, 2770–2777. [[CrossRef](#)] [[PubMed](#)]
4. Rosenwald, A.; Wright, G.; Wiestner, A.; Chan, W.C.; Connors, J.M.; Campo, E.; Gascoyne, R.D.; Grogan, T.M.; Muller-Hermelink, H.K.; Smeland, E.B.; et al. The proliferation gene expression signature is a quantitative integrator of oncogenic events that predicts survival in mantle cell lymphoma. *Cancer Cell* **2003**, *3*, 185–197. [[CrossRef](#)]
5. Salaverria, I.; Zettl, A.; Beà, S.; Moreno, V.; Valls, J.; Hartmann, E.; Ott, G.; Wright, G.; Lopez-Guillermo, A.; Chan, W.C.; et al. Specific secondary genetic alterations in mantle cell lymphoma provide prognostic information independent of the gene expression-based proliferation signature. *J. Clin. Oncol.* **2007**, *25*, 1216–1222. [[CrossRef](#)]
6. Tiemann, M.; Schrader, C.; Klapper, W.; Dreyling, M.H.; Campo, E.; Norton, A.; Berger, F.; Kluin, P.; Ott, G.; Pileri, S.; et al. Histopathology, cell proliferation indices and clinical outcome in 304 patients with mantle cell lymphoma (MCL): A clinicopathological study from the European MCL Network. *Br. J. Haematol.* **2005**, *131*, 29–38. [[CrossRef](#)]
7. Dreyling, M.; Campo, E.; Hermine, O.; Jerkeman, M.; le Gouill, S.; Rule, S.; Shpilberg, O.; Walewski, J.; Ladetto, M. Newly diagnosed and relapsed mantle cell lymphoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.* **2017**, *28*, iv62–iv71. [[CrossRef](#)]
8. Swerdlow, S.H.; Campo, E.; Harris, N.L.; Jaffe, E.S.; Pileri, S.A.; Stein, H.; Thiele, J.; Vardiman, J.W. *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues*. WHO Classification of Tumours, 4th ed.; WHO Press: Geneva, Switzerland, 2017; Volume 2.
9. Geisler, C.H.; Kolstad, A.; Laurell, A.; Andersen, N.S.; Pedersen, L.B.; Jerkeman, M.; Eriksson, M.; Nordström, M.; Kimby, E.; Boesen, A.M. Long-term progression-free survival of mantle cell lymphoma after intensive front-line immuno-chemotherapy with in vivo-purged stem cell rescue: A nonrandomized phase 2 multicenter study by the Nordic Lymphoma Group. *Blood. J. Am. Soc. Hematol.* **2008**, *112*, 2687–2693.
10. Cook, G.J.R.; Goh, V. What can artificial intelligence teach us about the molecular mechanisms underlying disease? *Eur. J. Pediatr.* **2019**, *46*, 2715–2721. [[CrossRef](#)]
11. Lambin, P.; Rios-Velazquez, E.; Leijenaar, R.; Carvalho, S.; van Stiphout, R.G.P.M.; Granton, P.; Zegers, C.M.L.; Gillies, R.; Boellard, R.; Dekker, A.; et al. Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur. J. Cancer* **2012**, *48*, 441–446. [[CrossRef](#)]
12. Lambin, P.; Leijenaar, R.T.H.; Deist, T.M.; Peerlings, J.; de Jong, E.E.C.; van Timmeren, J.; Sanduleanu, S.; Larue, R.T.H.M.; Even, A.J.G.; Jochems, A.; et al. Radiomics: The bridge between medical imaging and personalized medicine. *Nat. Rev. Clin. Oncol.* **2017**, *14*, 749–762. [[CrossRef](#)] [[PubMed](#)]
13. Schieber, M.; Gordon, L.I.; Karmali, R. Current overview and treatment of mantle cell lymphoma. *F1000Research* **2018**, *7*, 1136. [[CrossRef](#)] [[PubMed](#)]
14. Hill, H.A.; Qi, X.; Jain, P.; Nomie, K.; Wang, Y.; Zhou, S.; Wang, M.L. Genetic mutations and features of mantle cell lymphoma: A systematic review and meta-analysis. *Blood Adv.* **2020**, *4*, 2927–2938. [[CrossRef](#)] [[PubMed](#)]
15. Nadeu, F.; Martin-Garcia, D.; Clot, G.; Díaz-Navarro, A.; Duran-Ferrer, M.; Navarro, A.; Vilarrasa-Blasi, R.; Kulis, M.; Royo, R.; Gutiérrez-Abril, J.; et al. Genomic and epigenomic insights into the origin, pathogenesis, and clinical behavior of mantle cell lymphoma subtypes. *Blood* **2020**, *136*, 1419–1432. [[CrossRef](#)] [[PubMed](#)]
16. Hoster, E.; Dreyling, M.; Klapper, W.; Gisselbrecht, C.; Van Hoof, A.; Kluin-Nelemans, J.C.; Pfreundschuh, M.; Reiser, M.; Metzner, B.; Einsele, H.; et al. A new prognostic index (MIPI) for patients with advanced-stage mantle cell lymphoma. *Blood* **2008**, *111*, 558–565. [[CrossRef](#)] [[PubMed](#)]
17. Hoster, E.; Rosenwald, A.; Berger, F.; Bernd, H.-W.; Hartmann, S.; Loddenkemper, C.; Barth, T.F.; Brousse, N.; Pileri, S.; Rymkiewicz, G.; et al. Prognostic value of Ki-67 index, cytology, and growth pattern in mantle-cell lymphoma: Results from randomized trials of the European Mantle Cell Lymphoma Network. *J. Clin. Oncol.* **2016**, *34*, 1386–1394. [[CrossRef](#)] [[PubMed](#)]
18. Ladetto, M.; Magni, M.; Pagliano, G.; De Marco, F.; Drandi, D.; Ricca, I.; Astolfi, M.; Matteucci, P.; Guidetti, A.; Mantoan, B.; et al. Rituximab induces effective clearance of minimal residual disease in molecular relapses of mantle cell lymphoma. *Biol. Blood Marrow Transplant.* **2006**, *12*, 1270–1276. [[CrossRef](#)]
19. Pott, C.; Hoster, E.; Delfau-Larue, M.-H.; Beldjord, K.; Böttcher, S.; Asnafi, V.; Plonquet, A.; Siebert, R.; Callet-Bauchu, E.; Andersen, N.; et al. Molecular remission is an independent predictor of clinical outcome in patients with mantle cell lymphoma after combined immunochemotherapy: A European MCL intergroup study. *Blood* **2010**, *115*, 3215–3223. [[CrossRef](#)]
20. Martin, P.; Chadburn, A.; Christos, P.; Weil, K.; Furman, R.R.; Ruan, J.; Elstrom, R.; Niesvizky, R.; Ely, S.; DiLiberto, M. Outcome of deferred initial therapy in mantle-cell lymphoma. *J. Clin. Oncol.* **2009**, *27*, 1209–1213. [[CrossRef](#)]
21. Steinbuss, G.; Kriegsmann, M.; Zgorzelski, C.; Brobeil, A.; Goepfert, B.; Dietrich, S.; Mechttersheimer, G.; Kriegsmann, K. Deep Learning for the Classification of Non-Hodgkin Lymphoma on Histopathological Images. *Cancers* **2021**, *13*, 2419. [[CrossRef](#)]
22. Davnall, F.; Yip, C.S.P.; Ljungqvist, G.; Selmi, M.; Ng, F.; Sanghera, B.; Ganeshan, B.; Miles, K.A.; Cook, G.J.; Goh, V. Assessment of tumor heterogeneity: An emerging imaging tool for clinical practice? *Insights Imaging* **2012**, *3*, 573–589. [[CrossRef](#)] [[PubMed](#)]
23. Burrell, R.A.; McGranahan, N.; Bartek, J.; Swanton, C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* **2013**, *501*, 338–345. [[CrossRef](#)] [[PubMed](#)]
24. Stanta, G.; Bonin, S. Overview on clinical relevance of intra-tumor heterogeneity. *Front. Med.* **2018**, *5*, 85. [[CrossRef](#)] [[PubMed](#)]
25. Schürch, C.M.; Federmann, B.; Quintanilla-Martinez, L.; Fend, F. Tumor heterogeneity in lymphomas: A different breed. *Pathobiology* **2018**, *85*, 130–145. [[CrossRef](#)]

26. Gillies, R.J.; Kinahan, P.E.; Hricak, H. Radiomics: Images are more than pictures, they are data. *Radiology* **2016**, *278*, 563–577. [CrossRef]
27. Ibrahim, A.; Primakov, S.; Beuque, M.; Woodruff, H.; Halilaj, I.; Wu, G.; Refaee, T.; Granzier, R.; Widaatalla, Y.; Hustinx, R.; et al. Radiomics for precision medicine: Current challenges, future prospects, and the proposal of a new framework. *Methods* **2021**, *188*, 20–29. [CrossRef]
28. Sollini, M.; Antunovic, L.; Chiti, A.; Kirienko, M. Towards clinical application of image mining: A systematic review on artificial intelligence and radiomics. *Eur. J. Nucl. Med. Mol. Imaging* **2019**, *46*, 2656–2672. Available online: <https://link.springer.com/article/10.1007/s00259-019-04372-x> (accessed on 17 December 2021). [CrossRef]
29. Mayerhoefer, M.E.; Riedl, C.C.; Kumar, A.; Gibbs, P.; Weber, M.; Tal, I.; Schilksy, J.; Schöder, H. Radiomic features of glucose metabolism enable prediction of outcome in mantle cell lymphoma. *Eur. J. Nucl. Med. Mol. Imaging* **2019**, *46*, 2760–2769. [CrossRef]
30. Limkin, E.J.; Reuzé, S.; Carré, A.; Sun, R.; Schernberg, A.; Alexis, A.; Deutsch, E.; Ferté, C.; Robert, C. The complexity of tumor shape, spiculatedness, correlates with tumor radiomic shape features. *Sci. Rep.* **2019**, *9*, 4329. [CrossRef]
31. Jeong, W.K.; Jamshidi, N.; Felker, E.R.; Raman, S.S.; Lu, D.S. Radiomics and radiogenomics of primary liver cancers. *Clin. Mol. Hepatol.* **2019**, *25*, 21–29. [CrossRef]
32. Horvat, N.; Bates, D.D.B.; Petkowska, I. Novel imaging techniques of rectal cancer: What do radiomics and radiogenomics have to offer? A literature review. *Abdom. Radiol.* **2019**, *44*, 3764–3774. [CrossRef] [PubMed]
33. Lohmann, P.; Kocher, M.; Cecon, G.; Bauer, E.K.; Stoffels, G.; Viswanathan, S.; Ruge, M.I.; Neumaier, B.; Shah, N.J.; Fink, G.R.; et al. Combined FET PET/MRI radiomics differentiates radiation injury from recurrent brain metastasis. *NeuroImage Clin.* **2018**, *20*, 537–542. [CrossRef] [PubMed]
34. Huang, S.-Y.; Franc, B.L.; Harnish, R.J.; Liu, G.; Mitra, D.; Copeland, T.P.; Arasu, V.A.; Kornak, J.; Jones, E.F.; Behr, S.C.; et al. Exploration of PET and MRI radiomic features for decoding breast cancer phenotypes and prognosis. *Npj Breast Cancer* **2018**, *4*, 24. [CrossRef] [PubMed]
35. Acharya, U.R.; Hagiwara, Y.; Sudarshan, V.K.; Chan, W.Y.; Ng, K.H. Towards precision medicine: From quantitative imaging to radiomics. *J. Zhejiang Univ. Sci. B* **2018**, *19*, 6–24. [CrossRef]
36. Coroller, T.P.; Agrawal, V.; Narayan, V.; Hou, Y.; Grossmann, P.; Lee, S.W.; Mak, R.H.; Aerts, H.J. Radiomic phenotype features predict pathological response in non-small cell lung cancer. *Radiother. Oncol.* **2016**, *119*, 480–486. [CrossRef]
37. Antunes, J.; Viswanath, S.; Rusu, M.; Valls, L.; Hoimes, C.; Avril, N.; Madabhushi, A. Radiomics Analysis on FLT-PET/MRI for Characterization of Early Treatment Response in Renal Cell Carcinoma: A Proof-of-Concept Study. *Transl. Oncol.* **2016**, *9*, 155–162. [CrossRef]
38. Zwanenburg, A.; Vallières, M.; Abdalah, M.A.; Aerts, H.J.W.L.; Andrearczyk, V.; Apte, A.; Ashrafinia, S.; Bakas, S.; Beukinga, R.J.; Boellaard, R.; et al. The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* **2020**, *295*, 328–338. [CrossRef]
39. Götz, M.; Maier-Hein, K.H. Optimal statistical incorporation of independent feature stability information into radiomics studies. *Sci. Rep.* **2020**, *10*, 737. [CrossRef]
40. Zwanenburg, A. Radiomics in nuclear medicine: Robustness, reproducibility, standardization, and how to avoid data analysis traps and replication crisis. *Eur. J. Nucl. Med. Mol. Imaging* **2019**, *46*, 2638–2655. [CrossRef]
41. Lippi, M.; Gianotti, S.; Fama, A.; Casali, M.; Barbolini, E.; Ferrari, A.; Fioroni, F.; Iori, M.; Luminari, S.; Menga, M.; et al. Texture analysis and multiple-instance learning for the classification of malignant lymphomas. *Comput. Methods Programs Biomed.* **2019**, *185*, 105153. [CrossRef]
42. Xu, H.; Guo, W.; Cui, X.; Zhuo, H.; Xiao, Y.; Ou, X.; Zhao, Y.; Zhang, T.; Ma, X. Three-dimensional texture analysis based on PET/CT images to distinguish hepatocellular carcinoma and hepatic lymphoma. *Front. Oncol.* **2019**, *9*, 844. [CrossRef] [PubMed]
43. Tatsumi, M.; Isohashi, K.; Matsunaga, K.; Watabe, T.; Kato, H.; Kanakura, Y.; Hatazawa, J. Volumetric and texture analysis on FDG PET in evaluating and predicting treatment response and recurrence after chemotherapy in follicular lymphoma. *Int. J. Clin. Oncol.* **2019**, *24*, 1292–1300. [CrossRef] [PubMed]
44. Ganeshan, B.; Miles, K.A.; Babikir, S.; Shortman, R.; Afaq, A.; Ardeschna, K.M.; Groves, A.M.; Kayani, I. CT-based texture analysis potentially provides prognostic information complementary to interim fdg-pet for patients with hodgkin’s and aggressive non-hodgkin’s lymphomas. *Eur. Radiol.* **2017**, *27*, 1012–1020. [CrossRef]
45. Santiago, R.; Jimenez, J.O.; Forghani, R.; Muthukrishnan, N.; Del Corpo, O.; Karthigesu, S.; Haider, M.Y.; Reinhold, C.; Assouline, S. CT-based radiomics model with machine learning for predicting primary treatment failure in diffuse large B-cell Lymphoma. *Transl. Oncol.* **2021**, *14*, 101188. [CrossRef]
46. Wang, H.; Zhou, Y.; Li, L.; Hou, W.; Ma, X.; Tian, R. Current status and quality of radiomics studies in lymphoma: A systematic review. *Eur. Radiol.* **2020**, *30*, 6228–6240. [CrossRef]
47. Zhu, S.; Xu, H.; Shen, C.; Wang, Y.; Xu, W.; Duan, S.; Chen, H.; Ou, X.; Chen, L.; Ma, X. Differential diagnostic ability of 18F-FDG PET/CT radiomics features between renal cell carcinoma and renal lymphoma. *Q. J. Nucl. Med. Mol. Imaging* **2019**, *65*, 72–78. [CrossRef]
48. Milgrom, S.A.; Elhalawani, H.; Lee, J.; Wang, Q.; Mohamed, A.S.R.; Dabaja, B.S.; Pinnix, C.C.; Gunther, J.R.; Court, L.; Rao, A.; et al. A PET radiomics model to predict refractory mediastinal hodgkin lymphoma. *Sci. Rep.* **2019**, *9*, 1322. [CrossRef]
49. Suh, H.B.; Choi, Y.S.; Bae, S.; Ahn, S.S.; Chang, J.H.; Kang, S.-G.; Kim, E.H.; Kim, S.H.; Lee, S.-K. Primary central nervous system lymphoma and atypical glioblastoma: Differentiation using radiomics approach. *Eur. Radiol.* **2018**, *28*, 3832–3839. [CrossRef]

50. Kang, D.; Park, J.E.; Kim, Y.-H.; Kim, J.H.; Oh, J.Y.; Kim, J.; Kim, Y.; Kim, S.T.; Kim, H.S. Diffusion radiomics as a di-agnostic model for atypical manifestation of primary central nervous system lymphoma: Development and multicenter external validation. *Neuro-Oncology* **2018**, *20*, 1251–1261. [[CrossRef](#)]
51. Reinert, C.P.; Krieg, E.-M.; Bösmüller, H.; Horger, M. Mid-term response assessment in multiple myeloma using a texture analysis approach on dual energy-CT-derived bone marrow images—A proof of principle study. *Eur. J. Radiol.* **2020**, *131*, 109214. [[CrossRef](#)]
52. Reinert, C.P.; Federmann, B.; Hofmann, J.; Bösmüller, H.; Wirths, S.; Fritz, J.; Horger, M. Computed tomography textural analysis for the differentiation of chronic lymphocytic leukemia and diffuse large B cell lymphoma of Richter syndrome. *Eur. Radiol.* **2019**, *29*, 6911–6921. [[CrossRef](#)] [[PubMed](#)]
53. Reinert, C.P.; Kloth, C.; Fritz, J.; Nikolaou, K.; Horger, M. Discriminatory CT-textural features in splenic infiltration of lymphoma versus splenomegaly in liver cirrhosis versus normal spleens in controls and evaluation of their role for longitudinal lymphoma monitoring. *Eur. J. Radiol.* **2018**, *104*, 129–135. [[CrossRef](#)] [[PubMed](#)]
54. Nicolasilwan, M.; Hu, Y.; Yan, C.; Meerzaman, D.; Holder, C.; Gutman, D.; Jain, R.; Colen, R.; Rubin, D.; Zinn, P. TCGA Glioma Phenotype Research Group Addition of MR imaging features and genetic biomarkers strengthens glioblastoma survival prediction in TCGA patients. *J. Neuroradiol.* **2015**, *42*, 212–221. [[CrossRef](#)] [[PubMed](#)]
55. Mohri, M.; Rostamizadeh, A.; Talwalkar, A. *Foundations of Machine Learning*; MIT Press: Cambridge, MA, USA, 2012; Chapter 1, pp. 1–3.
56. Koza, J.R.; Bennett, F.H.; Andre, D.; Keane, M.A. Automated Design of Both the Topology and Sizing of Analog Electrical Circuits Using Genetic Programming. In *Artificial Intelligence in Design' 96*; Gero, J.S., Sudweeks, F., Eds.; Springer: Dordrecht, The Netherlands, 1996; pp. 151–170.
57. Korotcov, A.; Tkachenko, V.; Russo, D.P.; Ekins, S. Comparison of deep learning with multiple machine learning methods and metrics using diverse drug discovery data sets. *Mol. Pharm.* **2017**, *14*, 4462–4475. [[CrossRef](#)] [[PubMed](#)]
58. Greenspan, H.; Van Ginneken, B.; Summers, R.M. Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. *IEEE Trans. Med. Imaging* **2016**, *35*, 1153–1159. [[CrossRef](#)]
59. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Netw.* **2015**, *61*, 85–117. [[CrossRef](#)]
60. Rajkomar, A.; Lingam, S.; Taylor, A.G.; Blum, M.; Mongan, J. High-throughput classification of radiographs using deep convolutional neural networks. *J. Digit. Imaging* **2017**, *30*, 95–101. [[CrossRef](#)]
61. Haralick, R.M.; Shanmugam, K.; Dinstein, I.H. Textural Features for Image Classification. *IEEE Trans. Syst. Man Cybern.* **1973**, *SMC-3*, 610–621. [[CrossRef](#)]
62. Duin, R.P.; Pekalska, E. *Dissimilarity Representation for Pattern Recognition, The: Foundations and Applications*; World Scientific: Singapore, 2005; Volume 64.
63. Sánchez-Maróño, N.; Alonso-Betanzos, A.; Tombilla-Sanromán, M. Filter Methods for Feature Selection—A Comparative Study. In *International Conference on Intelligent Data Engineering and Automated Learning*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 178–187.
64. Alhaj, T.A.; Siraj, M.M.; Zainal, A.; Elshoush, H.T.; Elhaj, F. Feature selection using information gain for improved structural-based alert correlation. *PLoS ONE* **2016**, *11*, e0166017. [[CrossRef](#)]
65. Kira, K.; Rendell, L.A. A Practical Approach to Feature Selection. In *Machine Learning Proceedings 1992*; Elsevier: Amsterdam, The Netherlands, 1992; pp. 249–256. [[CrossRef](#)]
66. Urbanowicz, R.J.; Meeker, M.; La Cava, W.; Olson, R.S.; Moore, J.H. Relief-based feature selection: Introduction and review. *J. Biomed. Inform.* **2018**, *85*, 189–203. [[CrossRef](#)]
67. Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I.H. The WEKA data mining software: An update. *ACM SIGKDD Explor. Newsl.* **2009**, *11*, 10–18. [[CrossRef](#)]
68. Kononenko, I. Estimating attributes: Analysis and extensions of RELIEF. In *Proceedings of the European Conference on Machine Learning, Catania, Italy, 6–8 April 1994*; pp. 171–182.
69. Cui, X.; Li, Y.; Fan, J.; Wang, T. A novel filter feature selection algorithm based on relief. *Appl. Intell.* **2021**, *52*, 5063–5081. [[CrossRef](#)]
70. Kononenko, I.; Šimec, E.; Robnik-Šikonja, M. Overcoming the myopia of inductive learning algorithms with RELIEFF. *Appl. Intell.* **1997**, *7*, 39–55. [[CrossRef](#)]
71. Sidey-Gibbons, J.A.M.; Sidey-Gibbons, C.J. Machine learning in medicine: A practical introduction. *BMC Med. Res. Methodol.* **2019**, *19*, 64. [[CrossRef](#)] [[PubMed](#)]
72. Rajkomar, A.; Dean, J.; Kohane, I. Machine learning in medicine. *N. Engl. J. Med.* **2019**, *380*, 1347–1358. [[CrossRef](#)]
73. Parmar, C.; Grossmann, P.; Bussink, J.; Lambin, P.; Aerts, H.J.W.L. Machine Learning methods for Quantitative Radiomic Biomarkers. *Sci. Rep.* **2015**, *5*, 13087. [[CrossRef](#)]
74. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
75. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
76. Liaw, A.; Wiener, M. Classification and regression by randomForest. *R News* **2002**, *2*, 18–22.
77. Laaksonen, J.; Oja, E. Classification with learning k-nearest neighbors. In *Proceedings of the International Conference on Neural Networks (ICNN'96)*, Washington, DC, USA, 3–6 June 1996; pp. 1480–1483.

78. Kayalibay, B.; Jensen, G.; van der Smagt, P. CNN-based segmentation of medical imaging data. *arXiv* **2017**, arXiv:1701.03056. [[CrossRef](#)]
79. Li, K.; Zhang, R.; Cai, W. Deep learning convolutional neural network (DLCNN): Unleashing the potential of (18)F-FDG PET/CT in lymphoma. *Am. J. Nucl. Med. Mol. Imaging* **2021**, *11*, 327–331. [[PubMed](#)]
80. Roth, H.R.; Lu, L.; Liu, J.; Yao, J.; Seff, A.; Cherry, K.; Kim, L.; Summers, R.M. Improving computer-aided detection using convolutional neural networks and random view aggregation. *IEEE Trans. Med. Imaging* **2015**, *35*, 1170–1181. [[CrossRef](#)] [[PubMed](#)]
81. Sibille, L.; Seifert, R.; Avramovic, N.; Vehren, T.; Spottiswoode, B.; Zuehlsdorff, S.; Schäfers, M. 18F-FDG PET/CT uptake classification in lymphoma and lung cancer by using deep convolutional neural networks. *Radiology* **2020**, *294*, 445–452. [[CrossRef](#)] [[PubMed](#)]
82. Yuan, C.; Zhang, M.; Huang, X.; Xie, W.; Lin, X.; Zhao, W.; Li, B.; Qian, D. Diffuse large B-cell lymphoma segmentation in PET-CT images via hybrid learning for feature fusion. *Med. Phys.* **2021**, *48*, 3665–3678. [[CrossRef](#)] [[PubMed](#)]
83. Blanc-Durand, P.; Jégou, S.; Kanoun, S.; Berriolo-Riedinger, A.; Bodet-Milin, C.; Kraeber-Bodéré, F.; Carlier, T.; Le Gouill, S.; Casasnovas, R.-O.; Meignan, M.; et al. Fully automatic segmentation of diffuse large B cell lymphoma lesions on 3D FDG-PET/CT for total metabolic tumour volume prediction using a convolutional neural network. *Eur. J. Nucl. Med. Mol. Radiol.* **2020**, *48*, 1362–1370. [[CrossRef](#)]
84. Bibault, J.-E.; Giraud, P.; Housset, M.; Durdux, C.; Taieb, J.; Berger, A.; Coriat, R.; Chaussade, S.; Dousset, B.; Nordlinger, B.; et al. Deep Learning and Radiomics predict complete response after neo-adjuvant chemoradiation for locally advanced rectal cancer. *Sci. Rep.* **2018**, *8*, 12611. [[CrossRef](#)]
85. Jain, P.; Wang, M. Mantle cell lymphoma: 2019 update on the diagnosis, pathogenesis, prognostication, and management. *Am. J. Hematol.* **2019**, *94*, 710–725. [[CrossRef](#)]
86. Abrisqueta, P.; Scott, D.; Slack, G.; Steidl, C.; Mottok, A.; Gascoyne, R.; Connors, J.; Sehn, L.; Savage, K.; Gerrie, A. Observation as the initial management strategy in patients with mantle cell lymphoma. *Ann. Oncol.* **2017**, *28*, 2489–2495. [[CrossRef](#)]
87. Rogers, W.; Seetha, S.T.; Refaee, T.A.G.; Lieverse, R.I.Y.; Granzier, R.W.Y.; Ibrahim, A.; Keek, S.A.; Sanduleanu, S.; Primakov, S.P.; Beuque, M.P.L.; et al. Radiomics: From qualitative to quantitative imaging. *Br. J. Radiol.* **2020**, *93*, 20190948. [[CrossRef](#)]
88. Aerts, H.J.W.L.; Velazquez, E.R.; Leijenaar, R.T.H.; Parmar, C.; Grossmann, P.; Carvalho, S.; Bussink, J.; Monshouwer, R.; Haibe-Kains, B.; Rietveld, D.; et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat. Commun.* **2014**, *5*, 4006. [[CrossRef](#)]
89. Liu, Y.; Kim, J.; Balagurunathan, Y.; Li, Q.; Garcia, A.L.; Stringfield, O.; Ye, Z.; Gillies, R.J. Radiomic features are associated with EGFR mutation status in lung adenocarcinomas. *Clin. Lung Cancer* **2016**, *17*, 441–448.e6. [[CrossRef](#)] [[PubMed](#)]
90. Yoon, H.J.; Sohn, I.; Cho, J.H.; Lee, H.Y.; Kim, J.-H.; Choi, Y.-L.; Kim, H.; Lee, G.; Lee, K.S.; Kim, J. Decoding tumor phenotypes for ALK, ROS1, and RET fusions in lung adenocarcinoma using a radiomics approach. *Medicine* **2015**, *94*, e1753. [[CrossRef](#)] [[PubMed](#)]
91. Dagogo-Jack, I.; Shaw, A.T. Tumour heterogeneity and resistance to cancer therapies. *Nat. Rev. Clin. Oncol.* **2018**, *15*, 81–94. [[CrossRef](#)] [[PubMed](#)]
92. Korfiatis, P.; Kline, T.L.; Lachance, D.H.; Parney, I.F.; Buckner, J.C.; Erickson, B.J. Residual Deep Convolutional Neural Network Predicts MGMT Methylation Status. *J. Digit. Imaging* **2017**, *30*, 622–628. [[CrossRef](#)]
93. Lin, H.; Zheng, W.; Peng, X. Orientation-Encoding CNN for Point Cloud Classification and Segmentation. *Mach. Learn. Knowl. Extr.* **2021**, *3*, 601–614. [[CrossRef](#)]
94. Pickens, A.; Sengupta, S. Benchmarking Studies Aimed at Clustering and Classification Tasks Using K-Means, Fuzzy C-Means and Evolutionary Neural Networks. *Mach. Learn. Knowl. Extr.* **2021**, *3*, 695–719. [[CrossRef](#)]
95. Araújo, V.J.S.; Guimarães, A.J.; de Campos Souza, P.V.; Rezende, T.S.; Araújo, V.S. Using resistin, glucose, age and BMI and pruning fuzzy neural network for the construction of expert systems in the prediction of breast cancer. *Mach. Learn. Knowl. Extr.* **2019**, *1*, 466–482. [[CrossRef](#)]
96. Škrlić, B.; Kralj, J.; Lavrač, N.; Pollak, S. Towards robust text classification with semantics-aware recurrent neural architecture. *Mach. Learn. Knowl. Extr.* **2019**, *1*, 575–589. [[CrossRef](#)]
97. Vallières, M.; Zwanenburg, A.; Badic, B.; Cheze, C.; Rest, L.; Visvikis, D.; Hatt, M. Responsible Radiomics Research for Faster Clinical Translation. *J. Nucl. Med.* **2018**, *59*, 189–193. [[CrossRef](#)]
98. Berenguer, R.; del Rosario Pastor-Juan, M.; Canales-Vazquez, J.; Castro-García, M.; Villas, M.V.; Masilla Legorburo, F.; Sabater, S. Radiomics of CT Features May Be Nonreproducible and Redundant: Influence of CT Acquisition Parameters. *Radiology* **2018**, *288*, 407–415. [[CrossRef](#)]
99. Hayes, D.F. Biomarker validation and testing. *Mol. Oncol.* **2014**, *9*, 960–966. [[CrossRef](#)] [[PubMed](#)]
100. Shukla-Dave, A.; Obuchowski, N.A.; Chenevert, T.L.; Jambawalikar, S.; Schwartz, L.H.; Malyarenko, D.; Huang, W.; Noworolski, S.M.; Young, R.J.; Shiroishi, M.S.; et al. Quantitative imaging biomarkers alliance (QIBA) recommendations for improved precision of DWI and DCE-MRI derived biomarkers in multicenter oncology trials. *J. Magn. Reson. Imaging* **2019**, *49*, e1–e301. [[CrossRef](#)] [[PubMed](#)]
101. Ger, R.B.; Zhou, S.; Chi, P.-C.M.; Lee, H.J.; Layman, R.R.; Jones, A.K.; Goff, D.L.; Fuller, C.D.; Howell, R.M.; Li, H.; et al. Comprehensive investigation on controlling for CT imaging variabilities in radiomics studies. *Sci. Rep.* **2018**, *8*, 13047. [[CrossRef](#)] [[PubMed](#)]

102. Hagiwara, A.; Fujita, S.; Ohno, Y.; Aoki, S. Variability and standardization of quantitative imaging: Monoparametric to multiparametric quantification, radiomics, and artificial intelligence. *Investig. Radiol.* **2020**, *55*, 601. [[CrossRef](#)]
103. Montagnon, E.; Cerny, M.; Cadrin-Chênevert, A.; Hamilton, V.; Derennes, T.; Ilinca, A.; Vandenbroucke-Menu, F.; Turcotte, S.; Kadoury, S.; Tang, A. Deep learning workflow in radiology: A primer. *Insights Imaging* **2020**, *11*, 1–15. [[CrossRef](#)]
104. Kocak, B.; Durmaz, E.S.; Ates, E.; Kilickesmez, O. Radiomics with artificial intelligence: A practical guide for beginners. *Diagn. Interv. Radiol.* **2019**, *25*, 485–495. [[CrossRef](#)]
105. Rogasch, J.M.M.; Hundsdoerfer, P.; Hofheinz, F.; Wedel, F.; Schatka, I.; Amthauer, H.; Furth, C. Pretherapeutic FDG-PET total metabolic tumor volume predicts response to induction therapy in pediatric Hodgkin’s lymphoma. *BMC Cancer* **2018**, *18*, 521. [[CrossRef](#)]
106. van Timmeren, J.E.; Cester, D.; Tanadini-Lang, S.; Alkadhi, H.; Baessler, B. Radiomics in medical imaging—“How-to” guide and critical reflection. *Insights Imaging* **2020**, *11*, 1–16. [[CrossRef](#)]
107. Weisman, A.J.; Kieler, M.W.; Perlman, S.B.; Hutchings, M.; Jeraj, R.; Kostakoglu, L.; Bradshaw, T.J. Convolutional neural networks for automated PET/CT detection of diseased lymph node burden in patients with lymphoma. *Radiol. Artif. Intell.* **2020**, *2*, e200016. [[CrossRef](#)]
108. Mayerhoefer, M.E.; Riedl, C.C.; Kumar, A.; Dogan, A.; Gibbs, P.; Weber, M.; Staber, P.B.; Huicochea Castellanos, S.; Schöder, H. [18F] FDG-PET/CT radiomics for prediction of bone marrow involvement in mantle cell lymphoma: A retrospective study in 97 patients. *Cancers* **2020**, *12*, 1138. [[CrossRef](#)]
109. Zhou, Z.; Jain, P.; Lu, Y.; Macapinlac, H.; Wang, M.L.; Son, J.B.; Pagel, M.D.; Xu, G.; Ma, J. Computer-aided detection of mantle cell lymphoma on 18F-FDG PET/CT using a deep learning convolutional neural network. *Am. J. Nucl. Med. Mol. Imaging* **2021**, *11*, 260. [[PubMed](#)]
110. Albano, D.; Bosio, G.; Bianchetti, N.; Pagani, C.; Re, A.; Tucci, A.; Giubbini, R.; Bertagna, F. Prognostic role of baseline 18F-FDG PET/CT metabolic parameters in mantle cell lymphoma. *Ann. Nucl. Med.* **2019**, *33*, 449–458. [[CrossRef](#)] [[PubMed](#)]
111. Hosein, P.J.; Pastorini, V.H.; Paes, F.M.; Eber, D.; Chapman, J.R.; Serafini, A.N.; Alizadeh, A.A.; Lossos, I.S. Utility of positron emission tomography scans in mantle cell lymphoma. *Am. J. Hematol.* **2011**, *86*, 841–845. [[CrossRef](#)] [[PubMed](#)]
112. Bailly, C.; Carlier, T.; Berriolo-Riedinger, A.; Casasnovas, O.; Gyan, E.; Meignan, M.; Moreau, A.; Burrioni, B.; Djaileb, L.; Gressin, R. Prognostic value of FDG-PET in patients with mantle cell lymphoma: Results from the LyMa-PET Project. *Haematologica* **2020**, *105*, e33. [[CrossRef](#)]
113. Bodet-Milin, C.; Touzeau, C.; Leux, C.; Sahin, M.; Moreau, A.; Maisonneuve, H.; Morineau, N.; Jardel, H.; Moreau, P.; Gallazini-Crépin, C.; et al. Prognostic impact of 18F-fluoro-deoxyglucose positron emission tomography in untreated mantle cell lymphoma: A retrospective study from the GOELAMS group. *Eur. J. Nucl. Med. Mol. Pediatr.* **2010**, *37*, 1633–1642. [[CrossRef](#)]
114. Karam, M.; Ata, A.; Irish, K.; Feustel, P.J.; Mottaghy, F.M.; Stroobants, S.G.; Verhoef, G.E.; Chundru, S.; Douglas-Nikitin, V.; Wong, C.-Y.O.; et al. FDG positron emission tomography/computed tomography scan may identify mantle cell lymphoma patients with unusually favorable outcome. *Nucl. Med. Commun.* **2009**, *30*, 770–778. [[CrossRef](#)]
115. Fernández-Delgado, M.; Cernadas, E.; Barro, S.; Amorim, D. Do we need hundreds of classifiers to solve real world classification problems? *J. Mach. Learn. Res.* **2014**, *15*, 3133–3181.
116. Freedman, A.N.; Seminara, D.; Gail, M.H.; Hartge, P.; Colditz, G.A.; Ballard-Barbash, R.; Pfeiffer, R.M. Cancer risk prediction models: A workshop on development, evaluation, and application. *J. Natl. Cancer Inst.* **2005**, *97*, 715–723. [[CrossRef](#)]