Article

A Novel Algorithm for Histopathological Lung Cancer Detection

Nelson Faria [1], Sofia Campelos [2], and Vítor Carvalho [1,*]

[1] 2Ai - School of Technology, Polytechnic Institute of Cávado and Ave, Barcelos, Portugal
[2] Pathology Laboratory, Institute of Pathology and Molecular Immunology, University of Porto, Porto, Portugal
* Correspondence: vcarvalho@ipca.pt

Abstract: Lung cancer is the leading cause of cancer mortality worldwide, and it is urgently necessary to diagnose it as early as possible. Usually, the diagnostic process begins with a radiological examination which, when a possible tumour is present, is followed by a biopsy to extract tissue samples from the patient's lungs. Therefore, the purpose of this study is the development of an artificial intelligence algorithm that will analyse the Whole Slide Image (WSI) generated by the digitisation of the glass slides obtained from the extracted samples and detect if there is a tumour. The developed learning algorithms as well as the tested neural networks (NNs) were trained on a dataset composed of previously annotated WSI tiles, classified as Tumour or Non-Tumour. From these, four developed convolutional neural networks stood out and were selected to be compared with each other and with the tested NNs. When the best result of each of the developed architectures was compared to the highest result of the tested NNs, it was possible to denote that version 4 of CancerDetecNN achieved an average accuracy of 89.749 % and an average loss of 0.220. Furthermore, the results for the four selected versions are in agreement with the results reported in the literature, however, the limited size of the given dataset must be considered. Given the results obtained, the fourth version has the potential to optimise the lung cancer diagnosis process.

Keywords: lung cancer; digital pathology; whole slide imaging; artificial intelligence; deep learning; convolutional neural networks; computer-aided diagnosis

1. Introduction

Known as the type of neoplasia that causes the highest number of deaths worldwide, the lung cancer reached, approximately, 2.21 million of new cases in 2020 [1]. This type of cancer most commonly affects persons over the age of 50, and detecting lung cancer at an early stage is critical since the earlier it is diagnosed, the better the chances of effective treatment and survival [2,3]. The main cause of lung cancer is tobacco, but it can be originated by other risk factors, such as previous respiratory diseases, exposure to occupational carcinogens (arsenic, asbestos, chromium, nickel, and radon), polycyclic aromatic hydrocarbons, human immunodeficiency, virus infection, and alcohol consumption [3–5]. Symptoms of this neoplasia include chronic cough, dyspnea, and chest discomfort, and metastatic patients may also experience weakness, fatigue, weight loss, and cachexia [6].

Historically, lung cancer has been classified into two main subtypes: non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). However, developments in recent years, such as the discovery of specific mutations in different subtypes of NSCLC, have caused the group to be divided into adenocarcinoma (ADC), squamous cell carcinoma (SCC), large cell carcinoma (LCC), among others. The NSCLC group represents more than 85% of cases, with ADC and SCC being the most prevalent [3].

The process for the detection and classification of lung cancer begins with a radiological technique like chest X-ray, computed tomography, or magnetic resonance imaging. Then, in the possibility of malicious tissue, a biopsy is performed in order to extract material for histopathological examination. The obtained tissues from the histologic biopsy, denominated as formalin-fixed paraffin-embedded tissues, are processed to originate glass slides

 doi:10.20944/preprints202211.0205.v1

routinely stained with hematoxylin and eosin. After that, the pathologists can analyse the glass slides through the brightfield microscope or whole slide image (WSI), which is a high-resolution digital scan of a microscopic slide that allows counting, measuring the sizes and density of the objects, and applying image processing algorithms to detect lesions or cancer. Finally, the result provided by the exam is transmitted to the patient [3]. Even though it seems like a simple process, histological evaluation takes a long period of time and requires a high level of precision. Attending to the constraints described above, as whole slide imaging capacity emerges, more clinical data regarding WSIs will be generated, allowing the development of systems capable of executing operations such as cancer detection and classification [7,8].

Integration of software that uses technologies like artificial intelligence (AI) can accelerate advances in oncologic pathology by making cancer diagnosis simpler and more accurate, reducing the time required by pathologists to perform detection and classification tasks, which allows them to move faster to the final diagnosis and execute the required treatment [8]. Until today, there is no consensus on the definition of AI, but it can be categorised as *Thinking Humanly*, *Thinking Rationally*, *Acting Humanly*, and *Acting Rationally*. The *Thinking Humanly* category represents the idea that even if an algorithm gets a correct answer, it does not necessarily mean imitating human thinking because it is challenging to define the model of human thinking, while *Thinking Rationally* aims to solve issues and construct models of cognitive processes, but it encounters challenges such as the difficulty of expressing informal knowledge using logical notation and the distinction between solving a theoretical and practical problem. *Acting Humanly* attempts to operationalise intelligence and assure human-level performance in all cognitive activities, but it is limited by its inability to adapt to new circumstances or learn how to deal with them, as well as its concentration on behaviour. The category *Acting Rationally* optimises the likelihood of achieving desired goals based on available knowledge, and rational behaviour entails making the proper option with an implicit logical decision [9].

Machine Learning (ML) and Deep Learning (DL) are subsets of AI that are made up of algorithms which are used in systems to generate predictions, rankings based on input data, image analysis, and decision making [9,10]. ML is built on a system that can autonomously acquire and integrate knowledge, allowing it to forecast or make judgements without being expressly programmed to do so. The system is made up of a learning algorithm that has been trained to adapt to the environment in which it is placed. DL is also a subset of ML, which includes sophisticated machine learning algorithms that perform better on tasks like image and speech recognition [11]. By employing artificial NNs with numerous processing layers and a vast quantity of training data, deep learning enables the learning of complex data representations with several degrees of abstraction [12].

An AI system in oncology might be combined to serve as the initial diagnostician, highlighting situations that require additional pathologist study, or to generate a second result to assess cases in which the pathologist and the system arrive at different conclusions. When these circumstances arise, the cases will be reexamined, and the pathologists will be given instructions by emphasising the output areas on which the algorithm's diagnosis was based [8].

Evaluating the approaches provided by the published studies and the procedure conducted by the pathologists, the lung cancer histopathological diagnosis is achieved in two steps: detection and classification. For the purpose of this study, only the step of detection will be analysed. Wang and colleagues released the first publication on the use of AI in lung cancer pathology where they describe the development of a convolutional neural network (CNN) based on the Inception V3 algorithm to analyse and classify ADC WSIs. The model has a sliding window mechanism that travels across the WSI in patches of 300 x 300 pixels, identifying each pixel as a nucleus centroid, non-nucleus, or nucleus boundary. Then, they retrieved the geographical distribution in the tumour microenvironment, nuclear morphology, and textural aspects from the tumour site and employed them as predictors in a recurrent prediction model. From a training and validation sets composed by 267 images

 doi:10.20944/preprints202211.0205.v1

and 457 images, respectively, they were able to achieve an accuracy of 89.8%. In 2018, Coudray and team followed the same architecture as Wang, but the size of patches was defined as 512 x 512 pixels. Besides the ADC samples, they also used non-malignant and SCC which resulted in a dataset with 1635 slides and an accuracy of 87%. Li and colleagues extracted samples from 33 lung patients to train the CNNs AlexNet, VGG, ResNet, and SqueezeNet. These NNs were tested in two different training schemas, from scratch and pre-trained, with the input being patches of 256 x 256 pixels that had been cropped with a stride of 196 pixels to provide enough overlapping between neighbouring patches. AlexNet obtained 97% accuracy when taught from scratch, whereas ResNet achieved 93% accuracy when pre-trained, however, the small dimension of the dataset must be taken into account. The next year, 2019, Yu and his team employed ImageNet's pre-trained CNNs, AlexNet, VGG-16, GoogLeNet, and ResNet-50, to identify ADC and SCC [3,8]. Using patches of 1000 x 1000 pixels with a 50% overlap, the NNs were able to achieve higher accuracy than 93.5% when identifying tumour regions from adjacent dense benign tissues, more than 87.7% when recapitulating expert pathologist's diagnosis, and a maximum accuracy of 86.4% in an independent cohort [3,13]. One limitation of these studies was the need to have pathologists annotating the WSIs to train the learning algorithms. So, Chen and his team described an architecture where the WSI is used as is as input for the NNs. Due to memory constraints, they proposed using the WSI with a magnification of 4 x initially to discover the critical regions and, for the final identification, the 40 x magnification pictures of those regions. This method yielded an accuracy for ADC and SCC of about 93% [3].

Based on the previous examined studies, the authors of the study presented in this paper suggested a method in which the WSI is divided into patches of 512 x 512 pixels to be used as input for the CNN. As result of the prediction, the system produces a heatmap and indicates the presence or absence of a tumour (Figure 1) [3,8]. Thus, the primary goals of this research are to evaluate current NNs and design new architectures of CNNs in order to identify an algorithm capable of obtaining clinically acceptable accuracy.

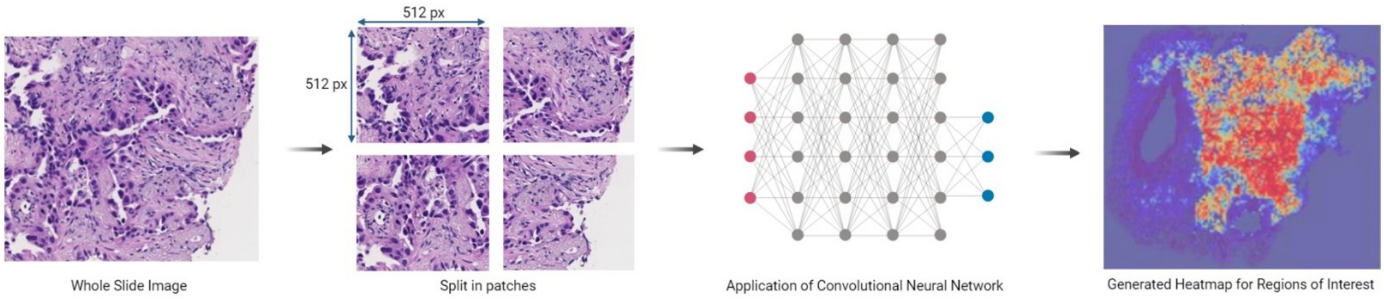

Figure 1. Proposed approach to detect Lung Cancer according to a given WSI (adapted from [3]).

This article is divided into three more sections: Section 2 describes the dataset's structure, as well as the tested and developed architectures of the NNs that will be used to fulfil the study's main goal. In Section 3, the characteristics of the training environment are illustrated, and the outcomes of training the NNs networks with varied numbers of epochs are discussed and analysed. Section 4 concludes with the main findings, limits, and future steps.

2. Materials and Methods

The starting point for the development of a novel algorithm for the histopathological detection of lung cancer is the collection of WSIs that will be analysed, annotated and used to train and evaluate all the different models. Then, in order to have a comparison with the designed algorithms and taking into account the reviewed articles, some NNs already accepted in AI community were chosen. Simultaneously, as stated before, CNNs were developed from the most basic form to architectures with a reasonable number of layers.

 doi:10.20944/preprints202211.0205.v1

2.1. Dataset

The training of CNNs require the existence of a dataset with images according to the context of the problem. So, for this project, it was requested to National Lung Screening Trial (NLST) and The Cancer Genome Atlas (TCGA) datasets composed by lung WSIs malignant, especially ADC and SCC, and non-malignant. The information regarding the number of cases and respective WSIs retrieved from each repository is presented in Table 1.

Table 1. Image repository and respective numbers of cases and images associated with lung cancer.

Repository	Number of Cases	Number of WSIs
NLST	449	1221
TCGA	582	821

From the collected WSIs, the pathologist (S.C.) annotated 15 WSIs from the NLST repository, having obtained 97 tumour and 75 non-tumor regions of interest. These regions of interest were subjected to transformations, such as rotation and zoom, for training purposes.

2.2. Neural Networks

Based on the popular CNN architectures and the reviewed studies, the selected algorithms to be tested with the prepared dataset were AlexNext, GoogLeNet and ResNet-50. Besides that, the authors developed 14 architectures which were nominated as versions of CancerDetecNN.

2.2.1. Tested Neural Networks

AlexNet

Built by Alex Krizhevesky and his team, AlexNet was presented in 2012 during the ImageNet Large Scale Visual Recognition Challenge (ILSVRC). The AlexNet's architecture consists of eight layers where five are convolutional layers and the other three are fully-connected layers. The input image has a dimension of 224 x 224 x 3 and the output of the last layer is fed to a 1000-way softmax which produces a distribution over the 1000 class labels. When participated in the competition, AlexNet achieved the best results with 15.3% top-5 error rate in the test set [14].

GoogLeNet

Szegedy and his team demonstrated GoogLeNet which is constituted by 22 layers (27 layers if counting pooling) in which 9 of the layers are inception modules. These modules are composed by an average pool layer, a convolution layer, two fully connected layers, and a linear layer with a softmax activation function. Their functionality is to perform a classification based on the inputs with the aim of adding the loss calculated to the total loss of the network. In 2014, GoogLeNet was used in both challenges of ILSVRC, that is, in detection and classification. The best model was achieved in the seventh version with 144 as the number of crops and a 6.67% of top-5 error [15].

ResNet

One year later, 2015, He and team introduced ResNet with the concept of "shortcut connection". "Shortcut connection" can be defined as skipping one or more layers with the function of performing identity mapping. The initial ResNet consisted of 34 layers, however, there were more experiments with 50, 101, and 152 layers. Based on the architecture of VGG networks, the majority of the layers have 3 x 3 filters, after each convolution and before activation, it is adopted batch normalization and, if a convolution layer had a stride of 2, it is performed a downsampling. From the plain network, the shortcut connections were inserted when the input and output have the same dimensions. The ResNet demonstrated the best result of the classification challenge, being the result 3.57% in the top-5 error rate for the test dataset [16].

 doi:10.20944/preprints202211.0205.v1

2.2.2. CancerDetecNN

The initial version of CancerDetecNN (Version 0) begins with input dimensions of 512 x 512 x 3, a convolution layer with kernel size of 3 x 3, 32 filters, strides defined as 2 x 2, and the rectified linear unit (ReLU) function as the activation function. Following this layer is a global average pooling operation and a fully connected layer of two units employing the softmax activation function. For version 1, a new fully connected layer with 512 units and the ReLu activation function was introduced between the global average pooling and fully connected layers of Version 0. After the initial convolution layer, Version 2 added a max pooling layer with a pooling size of 2 x 2. Upon the addition of the new max pooling layer, a new set of convolution and max pooling layers were added, duplicating the identical dimensions and activation functions, but with 64 filters rather than 32 in the convolution layer. Version 3 included a pair of convolution and max pooling layers, with the convolution layer containing 128 filters, in addition to the layers added in Version 2. Version 4 expanded the number of convolution and max pooling layer sets by making the same addition as Version 3. The max pooling layer positioned before the global average pooling layer was removed in Version 5. In the subsequent version (Version 6), the final fully connected layer activation function was changed from softmax to sigmoid, and the two units were reduced to one. The number of filters for the final convolution layer was increased to 256 in version 7. Furthermore, before the global average pooling layer, Version 8 included a max pooling layer with a size of 2 x 2. Besides that, the next version (Version 9) also appended a set of convolution and max pooling layers to that location, one of which included 128 filters, and modified the first fully connected layer units from 512 to 128. Version 10 demonstrated a "U-architecture", i.e. it has pairs of convolution layers and max pooling where only the filters are changed and follows the order: 32 - 64 - 128 - 256 - 128 - 64 - 32. This version also reduced the number of first fully connected layer units from 128 to 32. With the same structure as Version 10, Version 11 just reverted to 512 the number of units of the first fully connected layer. Version 12 eliminated the previously added sets in version 10 and changed the convolution layer from 256 to 128 filters. As a result, convolutional layers of 32, 64, and 128 filters were retained, culminating in one convolution layer of 32 filters, another of 64 filters, and three of 128 filters. The last version (Version 13) deleted the last set of convolutional with 128 filters and maximum pooling layers, as well as the fully connected layer with 512 units.

3. Results and Discussion

After collecting and analyzing all data, it was necessary to train and evaluate the selected and developed learning algorithms. The NLST dataset was used to produce a subset composed by 75 non-tumour and 97 tumour patches. Furthermore, the training dataset comprised 80% of the parts while the validation dataset had 20%, and each NN was trained with six different numbers of epochs, these being 50, 100, 150, 200, 250, and 300.

3.1. Training Environment

The training of each NN was performed in a HPC (High Performance Computer) with the following specifications:

- **Central Process Unit(s):** 4 x AMD EPYC 7452 32-Core Processor
- **Random Access Memory:** 512 Gigabytes
- **Graphical Process Unit(s):** 4 x NVIDIA A100-PCI 40 Gigabytes
- **Physical Memory**: 2 x KCD61LUL3T84 3.5 Terabytes

3.2. Tested Neural Networks

AlexNet, GoogLeNet, and ResNet were the existing learning algorithms selected to evaluate the performance when trained with the available subsets and with the different epochs. The ResNet NN was trained from the scratch and from weights obtained through the ImageNet dataset. The training results are shown in Table 2.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 November 2022 doi:10.20944/preprints202211.0205.v1

Table 2. Accuracies and Losses for Tested Neural Networks.

Epochs	Neural Networks	Dataset Type	Accuracy	Loss
50	AlexNet	Training	90.580 %	0.434
		Validation	73.530 %	3.674
	GoogLeNet	Training	81.884 %	1.604
		Validation	82.353 %	1.554
	ResNet	Training	83.333 %	0.551
		Validation	76.470 %	0.596
	ResNet Pre-Trained	Training	78.986 %	0.566
		Validation	44.118 %	0.728
100	AlexNet	Training	92.750 %	0.348
		Validation	79.410 %	1.453
	GoogLeNet	Training	86.232 %	0.912
		Validation	73.529 %	1.447
	ResNet	Training	86.957 %	0.464
		Validation	70.590 %	0.604
	ResNet Pre-Trained	Training	86.957 %	0.438
		Validation	73.529 %	0.577
150	AlexNet	Training	87.681 %	0.667
		Validation	76.470 %	11.545
	GoogLeNet	Training	56.522 %	2.057
		Validation	55.882 %	2.059
	ResNet	Training	85.507 %	0.451
		Validation	76.470 %	0.536
	ResNet Pre-Trained	Training	76.812 %	0.538
		Validation	58.824 %	0.817
200	AlexNet	Training	89.855 %	0.711
		Validation	55.880 %	189.040
	GoogLeNet	Training	93.478 %	0.633
		Validation	82.353 %	1.420
	ResNet	Training	84.783 %	55.880
		Validation	70.590 %	0.803
	ResNet Pre-Trained	Training	78.986 %	0.490
		Validation	73.530 %	0.579
250	AlexNet	Training	79.710 %	0.913
		Validation	64.706 %	7.000
	GoogLeNet	Training	87.681 %	1.010
		Validation	76.471 %	2.358
	ResNet	Training	88.406 %	0.381
		Validation	64.706 %	0.717
	ResNet Pre-Trained	Training	95.652 %	0.214
		Validation	44.118 %	1.135
300	AlexNet	Training	90.580 %	0.211
		Validation	85.294 %	5.037
	GoogLeNet	Training	56.522 %	2.060
		Validation	55.882 %	2.060
	ResNet	Training	87.681 %	0.384
		Validation	82.353 %	0.456
	ResNet Pre-Trained	Training	92.029 %	0.266
		Validation	79.412 %	0.575

When training the learning algorithms with 50 epochs and despite the loss values, it is possible to note that GoogLeNet was the most consistent when analysing the set of accuracy values for training and validation for each of the NNs and the one that demonstrated the highest validation accuracy. Although AlexNet provided the best result for training

 doi:10.20944/preprints202211.0205.v1

accuracy, it presented a lower value and the biggest loss among the NNs in validation. The ResNet models showed losses mainly between 0.551 and 0.596, except for the validation loss of the non-tuned ResNet which achieved 0.728. The ResNet model trained from scratch recorded the second-best accuracies, 83.333 % for training and 76.470 % for validation, while the pre-trained model registered the worst accuracy results, 78.986 % for training accuracy and 44.118 % for validation accuracy.

Changing the number of epochs to 100 globally indicates lower losses and less discrepancy in precision values than in the previous number of epochs. The GoogLeNet had the lowest training accuracy among the models. It is possible to observe that ResNet models shared the same training accuracy value (86.957 %), but the pre-trained model had a slightly better validation accuracy. In the set of train and validation accuracy, and considering the loss irrelevant, AlexNet can be classified as the best model in these circumstances. On the other hand, if the loss values are included, ResNet's pre-trained model showed the most consistent results.

The results of the NNs trained with 150 epochs allow to verify modifications in the accuracies and losses. The AlexNet and GoogLeNet models demonstrated higher losses and lower training and validation accuracies than in the preceding number of epochs. The ResNet model trained from scratch reduced the losses and training accuracy but improved the validation accuracy, while the pre-trained model had its accuracies decreased and losses incremented. If the losses are ignored, AlexNet obtained the best accuracies, however, if not, ResNet presented the superior set of values.

Observing the results for 200 epochs, even though AlexNet improved its training accuracy, the validation accuracy registered was lower and the loss incremented substantially. The GoogLeNet achieved the best results in terms of accuracy among the NNs, and the loss values decreased. Non-tuned ResNet had worse scores for accuracy and losses, but when pre-trained observed the opposite.

Evaluating the results with 250 epochs, AlexNet improved its validation results, but worsened the training values. Comparing to the previous values for 150 epochs, GoogLeNet demonstrated lower precision and higher losses both in training and validation, however remaining the NN with the most consistent accuracy values. The non-tuned model of ResNet improved its accuracies and losses, while the pre-trained only refined the training results. Furthermore, the ResNet pre-trained decreased its validation accuracy by 29.412 % and increased the validation loss by more than double.

The use of 300 epochs in NNs made possible to improve, mainly, their validation accuracies and losses, except for GoogLeNet where the training and validation accuracies reduced considerably. The ResNet models had their training values slightly decreased. In the set of training and validation accuracies, AlexNet demonstrated to be the best model if the losses are ignored.

3.2.1. Evaluation of Best Tested Neural Networks Results

After analysing the results for each of the tested NNs individually by epoch, the best results of each of the learning algorithms were taken (Table 3). Every NN presented results close to the clinically acceptable value (>= 90 %) in training accuracy, being the non-tuned ResNet model the one with the lower value (87.681 %). However, when looking at the validation accuracy, none of them can be used in a clinic environment since the best result was 82.353 % (obtained by GoogLeNet and ResNet trained from scratch), as it does not satisfy the requirements. When the loss values are not accounted for, GoogLeNet was the most accurate model. On the other hand, if considered, ResNet's tuned model stood out.

 doi:10.20944/preprints202211.0205.v1

Table 3. Best tested neural networks results.

Neural Networks	Epoch	Dataset Type	Accuracy	Loss	Accuracy Mean	Loss Mean
AlexNet	100	Training Validation	92.750 % 79.410 %	0.348 1.453	86.080 %	0.901
GoogLeNet	200	Training Validation	93.478 % 82.353 %	0.633 1.420	87.916 %	1.027
ResNet	300	Training Validation	87.681 % 82.353 %	0.384 0.456	85.017 %	0.420
ResNet Pre-Trained	300	Training Validation	92.029 % 79.412 %	0.266 0.575	85.721 %	0.421

Validating against the values observed in the literature and considering that the dataset size was smaller than the datasets available to the authors, it is notable that the accuracy of the NNs tested with the provided dataset reached lower values than the two authors [13,17] who also tested the same networks (Table 4).

Table 4. Comparison of the average of precision obtained with the results of the literature.

Neural Networks	Accuracy Mean	[13]	[17]
AlexNet	86.080 %	90.000 %	97.040 %
GoogLeNet	87.916 %	91.100 %	-
ResNet	85.017 %	-	95.420 %
ResNet Pre-Trained	85.721 %	89.000 %	93.070 %

3.3. CancerDetecNN

Following the evaluation of the tested NNs, the same tests were run on each of the CancerDetecNN architectures. From the results of all versions, the four CNNs with the greatest outcomes were chosen to be compared (Table 5).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 November 2022 doi:10.20944/preprints202211.0205.v1

Table 5. Accuracies and Losses for CancerDetecNN Versions.

Epochs	CancerDetecNN Version	Dataset Type	Accuracy	Loss
50	2	Training	86.232 %	0.392
		Validation	79.412 %	0.457
	4	Training	86.232 %	0.334
		Validation	79.412 %	0.436
	5	Training	89.130 %	0.269
		Validation	79.412 %	0.442
	13	Training	86.232 %	0.316
		Validation	76.471 %	0.484
100	2	Training	86.232 %	0.284
		Validation	67.647 %	0.538
	4	Training	86.957 %	0.291
		Validation	76.471 %	0.386
	5	Training	94.928 %	0.152
		Validation	58.824 %	0.931
	13	Training	85.507 %	0.281
		Validation	91.176 %	0.428
150	2	Training	90.580 %	0.216
		Validation	67.647 %	0.570
	4	Training	90.580 %	0.241
		Validation	55.882 %	2.457
	5	Training	93.478 %	0.146
		Validation	82.353 %	0.341
	13	Training	89.130 %	0.206
		Validation	70.588 %	0.483
200	2	Training	87.681 %	0.217
		Validation	85.294 %	0.349
	4	Training	88.406 %	0.233
		Validation	82.353 %	0.352
	5	Training	95.652 %	0.121
		Validation	82.353 %	0.379
	13	Training	93.478 %	0.160
		Validation	61.765 %	2.035
250	2	Training	86.957 %	0.239
		Validation	64.706 %	0.725
	4	Training	94.203 %	0.136
		Validation	85.294 %	0.303
	5	Training	94.928 %	0.154
		Validation	79.412 %	0.319
	13	Training	94.928 %	0.144
		Validation	61.765 %	1.847
300	2	Training	93.478 %	0.175
		Validation	67.647 %	0.522
	4	Training	95.652 %	0.122
		Validation	67.647 %	0.893
	5	Training	96.377 %	0.099
		Validation	73.529 %	1.335
	13	Training	96.377 %	0.132
		Validation	58.824 %	2.372

The results of training the different CancerDetecNN versions with 50 epochs showed a lot of similarities because three of them had as training accuracy the value 86.232 % and 79.412 % of validation accuracy. The maximum precision for the train dataset type

 doi:10.20944/preprints202211.0205.v1

was 89.130 % and, for the validation, it was 79.412 %. In terms of loss values, they are concentrated between 0.269 and 0.484. Comparing the different NNs, the architecture with higher accuracy values was version 5.

Changes in the accuracy and reductions in the training losses are seen when the number of epochs is increased to 100. For the training dataset type, Version 2 presented the same accuracy as before with 50 epochs, but a lower loss. However, during validation, the accuracy dropped by 11.765 %, and the loss rised. The fourth and fifth versions improved their training results but had worsened the validation accuracies. Regarding the validation loss of these two architectures, while version 4 reduced its loss, version 5 obtained results superior to twice the value received with 50 epochs. Version 13 had slightly lower training accuracy but refined all other metrics, making it the top model for 100 epochs among the other versions.

When the epoch count was increased to 150, the four variants produced acceptable results in terms of training accuracy and loss. Comparing with the results obtained with 100 epochs, version 2 maintained the accuracy but increased the loss during validation, while, for training dataset type, both results were improved. Although, versions 4 and 13 also improved during training, the validation loss had its value increased and validation accuracy reduced noticeably. Analysing the version 5, which gave the most consistent results with 150 epochs, even with a minimal decrease in training accuracy, the loss had a slightly better value. For the validation, significant improvements are seen since accuracy was increased from 58.824 % to 82.353 % and the loss lowered from 0.931 to 0.341.

Despite of the validation results for the thirteenth version, the accuracies obtained by the different architectures were between 82.353 % and 95.652 %, and the losses maintained from 0.121 to 0.379. Evaluating the values for each NN, versions 2 and 4 lowered their training accuracies, but had major increases in the validation values. When compared to the previous number of epochs, the fifth and thirteenth training values were improved, but their validation results remained the same, like accuracy of version 5, or were worsened. Version 2 produced the highest results in the training and validation set when using 200 epochs.

Evaluating the results using 250 epochs, it is possible to see that, when using the training dataset, versions 4, 5 and 13 showed high precision values, ranging from 94 % to 95 %, and losses of 0.136, 0.156 and 0.144, respectively, whereas version 2 reduced its accuracy and increased its loss. Regarding the results of the validation dataset, version 2 suffered a high reduction in the accuracy which also resulted in a higher loss. The fourth version achieved the best accuracy and loss results, outperforming the values from 100 epochs. Despite the good results in the training dataset and the improvement in the level of losses, version 5 reduced its accuracy in the validation. The accuracy of version 13 remained unchanged and, even with the loss lowered, was the highest loss among the learning algorithms. Version 4 of the architectures proved to be the most accurate in the set of training and validation accuracies.

The existence of high accuracies throughout training, ranging from 93.478 % to 96.377 %, and minimal losses, from 0.099 to 0.175, stands out among the results of 300 epochs. However, when the validation data is evaluated, there are considerable accuracy declines and severe losses, with accuracies between from 58.824 % to 73.529 % and losses reaching 2.372. None of them provided results clinically acceptable, but, checking the best model for 300 epochs, version 5 acquired the best accuracies; if the losses are taken into account, version 2 produced the best set of values.

3.3.1. Evaluation of best CancerDetecNN architectures

The analysis of the best results given by each CancerDetecNN architecture shows that the majority of the versions needed 200 or more epochs to achieve their best performances (versions 2, 4, and 5) (Table 6). However, version 13 was able to get competitive values when trained with 100 epochs. Calculating the mean of the accuracies and losses between the training and validation values, version 4 is the one which provided the closest clinical

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 November 2022 doi:10.20944/preprints202211.0205.v1

acceptable values, being followed by versions 5, 13, and 2, respectively. In addition, when checking the highest result for accuracy mean (Version 4) and the results of the NNs tested by the different authors in the literature, this CNN architecture is in concordance.

Table 6. Best CancerDetecNN architectures results.

CancerDetecNN Version	Epoch	Dataset Type	Accuracy	Loss	Accuracy Mean	Loss Mean
2	200	Training	87.681 %	0.216	86.488 %	0.283
		Validation	85.294 %	0.349		
4	250	Training	94.203 %	0.136	89.749 %	0.220
		Validation	85.294 %	0.303		
5	200	Training	95.652 %	0.121	89.003 %	0.250
		Validation	82.353 %	0.379		
13	100	Training	85.507 %	0.281	88.342 %	0.355
		Validation	91.176 %	0.428		

3.4. CancerDetecNN Version 4 and ResNet Pre-trained: Comparison of the NNs with the highest results

Based on the accuracy and loss values, ResNet pre-trained was chosen as the best model for the NNs tested and, for CancerDetecNN, version 4 (Table 7). In a general overview, the CancerDetecNN V4 achieved better results than the pre-trained model of ResNet with 50 fewer epochs, and a possible explanation lies in the fact that the architecture of the fourth version is smaller than ResNet's, which results in a lower number of parameters to be trained. The analysis of the results shows an increase of 4.028 % in the average of precision and 0.201 less in the average of losses for version 4 in relation to the tuned model of ResNet.

Table 7. Comparison of CancerDetecNN Version 4 and ResNet Pre-trained.

Neural Networks	Epoch	Dataset Type	Accuracy	Loss	Accuracy Mean	Loss Mean
CancerDetecNN V4	250	Training	94.203 %	0.136	89.749 %	0.220
		Validation	85.294 %	0.303		
ResNet Pre-Trained	300	Training	92.029 %	0.266	85.721 %	0.421
		Validation	79.412 %	0.575		

3.5. Detection using CancerDetecNN Version 4

The heatmaps generated by the best-performing model (CancerDetecNN Version 4) were examined in order to validate tumour or non-tumour regions incorrectly classified by the CNN (Figure 2). Following the identification of these areas, they are annotated to be used in a subsequent training of the model to increase the accuracy.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 November 2022 doi:10.20944/preprints202211.0205.v1

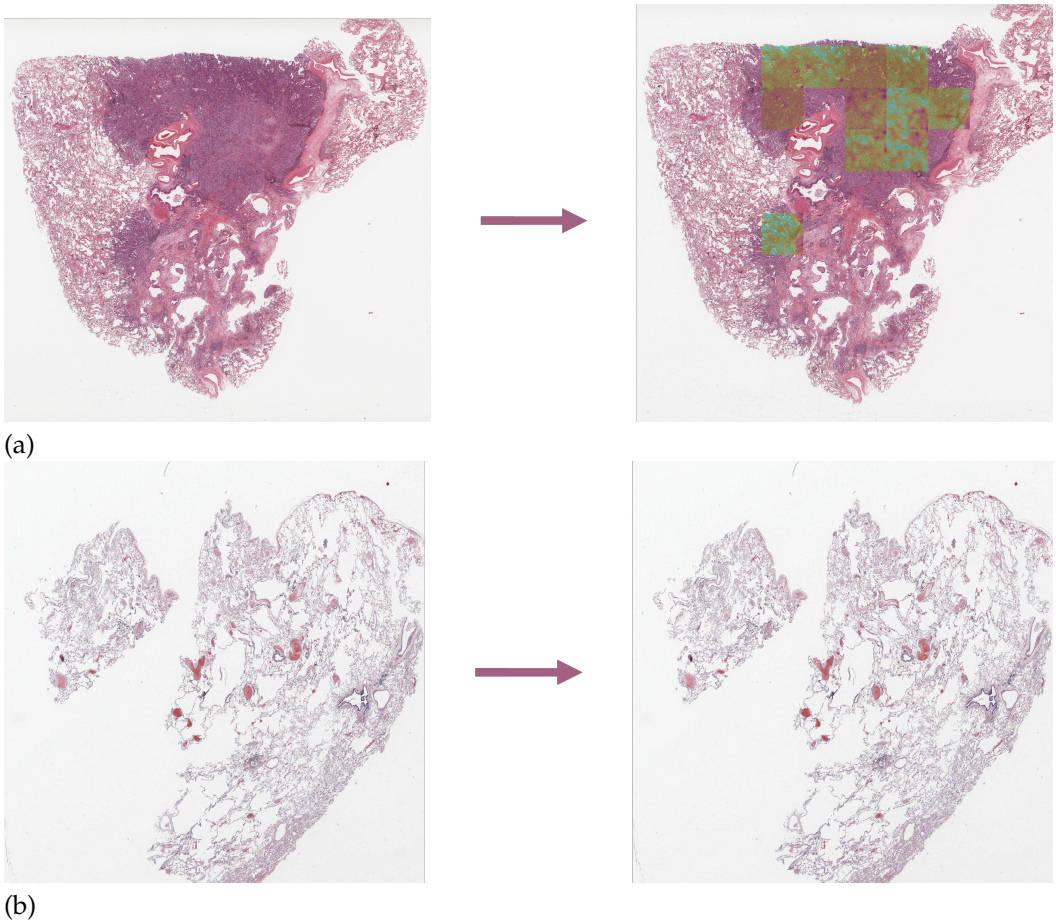

Figure 2. Examples of heatmaps generated for tumoural **(a)** and non-tumoural **(b)** WSIs. When presented to a tumour region, a colour gradation is displayed from lighter to darker to give a more intuitive perspective of the tumour location. Otherwise, when there is no tumour area, the heatmap is not applied, showing only the lung tissue.

4. Conclusion and Future Work

The study presented in this paper aims to provide the development of an algorithm capable of achieving clinically acceptable accuracy when evaluating the presence or absence of tumour tissue in lung WSIs in order to improve the lung cancer diagnosis accuracy and reduce the time spent by the pathologists. Following this idea, the authors collected datasets of tumour and non-tumour lung WSIs to train the developed and tested CNNs. Evaluating the training results given by all networks, it is possible to highlight Version 4 of CancerDetecNN since it delivered satisfactory results for accuracy and loss means (89.749 % for accuracy and 0.220 for loss).

The fundamental issue is that there are no annotated datasets with adequate size and range of situations to design algorithms with high enough performance to be validated for clinical application.

Further phases of this project include the application of the best model in a medical system for the analysis of lung WSIs. In addition, when more WSI annotations become available, the neural networks will be retrained and their results analysed in order to obtain a more accurate model. This will be a permanent learning algorithm. Another point to be implemented in the future is the testing of neural networks in the area of image segmentation, such as U-Net and Mask R-CNN.

Author Contributions: Conceptualization, N.F.; methodology, N.F.; software, N.F.; validation, N.F., S.C. and V.C.; formal analysis, N.F.; investigation, N.F.; resources, N.F., S.C. and V.C.; data curation,

N.F. and S.C.; writing—original draft preparation, N.F.; writing—review and editing, N.F., S.C. and V.C.; visualization, S.C. and V.C.; supervision, S.C. and V.C.; project administration, V.C.; funding acquisition, V.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by FCT/MCTES grant number UIDB/05549/2020.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: https://portal.gdc.cancer.gov. The data presented in this study are available on request from the corresponding author. The data are not publicly available due to DTAs are required for each research project to protect the confidentiality of the identity of study participants.

Acknowledgments: The authors would like to thank to the National Lung Screening Trial, The Cancer Genome Atlas and the Genomic Data Commons Data Portal for the lung cancer datasets availability and to FCT – Fundação para a Ciência e Tecnologia and FCT/MCTES in the scope of the project UIDB/05549/2020 for funding.

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

Abbreviations

The following abbreviations are used in this manuscript:

ADC	Adenocarcinoma
AI	Artificial Intelligence
CNN	Convolution Neural Network
DL	Deep Learning
LCC	Large Cell Carcinoma
ML	Machine Learning
NLST	National Lung Screening Trial
NN	Neural Network
NSCLC	Non-small cell lung cancer
ReLU	Rectified Linear Unit
SCC	Squamous Cell Carcinoma
SCLC	Small Cell Lung Cancer
TCGA	The Cancer Genome Atlas
WSI	Whole Slide Image

References

1. Cancer. Available online: https://www.who.int/news-room/fact-sheets/detail/cancer (accessed on 03 October 2022).
2. Nasim, F.; Sabath, B.F.; Eapen, G.A. Lung Cancer. *Medical Clinics of North America* **2019**, *103*, 463–473. Pulmonary Disease, https://doi.org/10.1016/j.mcna.2018.12.006.
3. Faria, N.; Campelos, S.; Carvalho, V. Cancer Detec-Lung Cancer Diagnosis Support System: First Insights. In Proceedings of the Proceedings of the 15th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 3: BIOINFORMATICS,. INSTICC, SciTePress, 2022, pp. 81–88. https://doi.org/10.5220/0010767800003123.
4. Bade, B.C.; Cruz, C.S.D. Lung cancer 2020: epidemiology, etiology, and prevention. *Clinics in chest medicine* **2020**, *41*, 1–24. https://doi.org/10.1016/j.ccm.2019.10.001.
5. Duma, N.; Santana-Davila, R.; Molina, J.R. Non-small cell lung cancer: epidemiology, screening, diagnosis, and treatment. In Proceedings of the Mayo Clinic Proceedings. Elsevier, 2019, Vol. 94(8), pp. 1623–1640. https://doi.org/10.1016/j.mayocp.2019.01.013.
6. Polanski, J.; Jankowska-Polanska, B.; Rosinczuk, J.; Chabowski, M.; Szymanska-Chabowska, A. Quality of life of patients with lung cancer. *OncoTargets and therapy* **2016**, *9*, 1023. https://doi.org/10.2147/OTT.S100685.
7. Bera, K.; Schalper, K.A.; Rimm, D.L.; Velcheti, V.; Madabhushi, A. Artificial intelligence in digital pathology—new tools for diagnosis and precision oncology. *Nature reviews Clinical oncology* **2019**, *16*, 703–715. https://doi.org/10.1038/s41571-019-0252-y.
8. Faria, N.; Campelos, S.; Carvalho, V. Development of a Lung Cancer Diagnosis Support System. In Proceedings of the ALLSENSORS 2022. IARIA, 2022, pp. 30–32.
9. Russell, S.; Norvig, P. *Artificial intelligence: a modern approach*, 4 ed.; Prentice Hall, 2021.

 doi:10.20944/preprints202211.0205.v1

10. Greenfield, D. Artificial Intelligence in Medicine: Applications, implications, and limitations - Science in the News, 2019. Available online: https://sitn.hms.harvard.edu/flash/2019/artificial-intelligence-in-medicine-applications-implications-and-limitations (accessed on 04 October 2022).
11. Zhang, X.D. *A Matrix Algebra Approach to Artificial Intelligence*, 1 ed.; Springer Singapore, 2020. https://doi.org/10.1007/978-981-15-2770-8.
12. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. https://doi.org/10.1038/nature14539.
13. Yu, K.H.; Wang, F.; Berry, G.J.; Ré, C.; Altman, R.B.; Snyder, M.; Kohane, I.S. Classifying non-small cell lung cancer types and transcriptomic subtypes using convolutional neural networks. *Journal of the American Medical Informatics Association* **2020**, *27*, 757–769. https://doi.org/10.1093/jamia/ocz230.
14. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Advances in neural information processing systems* **2012**, *25*. https://doi.org/10.1145/3065386.
15. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1–9. https://doi.org/10.1109/CVPR.2015.7298594.
16. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778. https://doi.org/10.1109/CVPR.2016.90.
17. Li, Z.; Hu, Z.; Xu, J.; Tan, T.; Chen, H.; Duan, Z.; Liu, P.; Tang, J.; Cai, G.; Ouyang, Q.; et al. Computer-aided diagnosis of lung carcinoma using deep learning - a pilot study. *arXiv* **2018**. https://doi.org/10.48550/arxiv.1803.05471.