LUNG CANCER DETECTION IN LOW-RESOLUTION IMAGES

Mostafa .k .abd alrahman aladamey [1]
AL Imam AL Aadham College
mustafaapple2020@gmail.com

Duha .d .salman
Computer Technical Engineering Dept.
AL Rasheed College
duhaalubady@gmail.com

Abstract
One of the most important prognostic factors for all lung cancer patients is the accurate detection of metastases. Pathologists, as we all know, examine the body and its tissues. On the existing clinical method, they have a tedious and manual task. Recent analysis has been inspired by these aspects. Deep Learning (DL) algorithms have been used to identify lung cancer. The developed cutting-edge technologies beat pathologists in terms of cancer identification and localization inside pathology images. These technologies, though, are not medically feasible because they need a massive amount of time or computing capabilities to perceive high-resolution images. Image processing techniques are primarily employed for lung cancer prediction and early identification and therapy to avoid lung cancer. This research aimed to assess lung cancer diagnosis by employing DL algorithms and low-resolution images. The goal would be to see if Machine Learning (ML) models might be created that generate higher confidence conclusions while consuming fractional resources by comparing low and high-resolution images. A DL pipeline has been built to a small enough size from compressing high-resolution images to be fed into an or before CNN (Convolutional Neural Network) for binary classification i.e. cancer or normal. Numerous enhancements have been done to increase overall performance, providing data augmentations, including augmenting training data and implementing tissue detection. Finally, the created low-resolution models are practically incapable of handling extremely low-resolution inputs i.e. 299 x 299 to 2048 x 2048 pixels. Considering the lack of classification ability, a substantial reduction in models' predictable times is only a marginal benefit. Due to an obvious drawback with the methodology, this is disheartening but predicted finding: very low resolutions, essentially expanding out on a slide, preserve only data about macro-cellular structures, which is usually insufficient to diagnose cancer by itself.

Index items: Convolutional Neural Networks, Deep Learning, Image Processing, Lung Cancer Detection, Machine Learning, Medical Image Analysis

I. INTRODUCTION

The phase of cancer describes the strength of a patient's cancer diagnosis. Cancer staging is indicated by characteristics such as tumour size and location, as well as, more significantly, whether cancer has spread. When lung cancer starts to spread, the survival rate drops to 26.2% [1]. However, England's people suffering from lung cancer have a relatively good prognosis, with an approximate 5-year absolute survival rate of 85% that can save patients from needless, perhaps harmful therapies as lung cancer is a type of cancer that can enhance prognosis and chances of survival, especially when metastasis is present. It is believed, that in the histopathology images, ML options are equipped to accurately diagnose lung cancer. These systems can automate major sections of traditional diagnostic methods which are used to detect lung cancer, increase diagnostic assistance through digital second opinion, and decrease cognitive burden by transferring jobs away from medical staff. Yet, these existing systems are complex. They completely use high-resolution images with width and height dimensions in the hundreds and thousands of pixels. These images show pathology slides at extremely high magnification. Such technologies are often resource-heavy due to the increased resolution of the images, needing either substantial time or computation power, restricting their clinical practicality. The researchers looked toward automated lung cancer diagnosis using DL algorithms and low-resolution images. This study aims to determine if ML models which produce high certainty findings with a minimal number of resources can be constructed by comparing high and low-resolution images. The problem objectives of the study are as follows:

- For the localization and automated classification of lung cancer using high and low-resolution images, this study should conduct a literature study on cutting-edge ML algorithms. This research review will look at existing models for identifying metastases in the lung, as well as models created in the greater perspective of automated cancer detection.
- Using methodologies (e.g., ResNet) guided by the literature analysis, this study should design and improve (e.g., via hyper-parameter tuning) an ML system for lung cancer classification in low-resolution histology slides.
- This study must compare the outcomes of the methodologies studied in the literature review with the model structure for low-resolution images. This

evaluation should analyse variations in detection confidence and resource needs across strategies that utilise higher vs lower resolution images, and connect this to the evaluation metrics of actual pathologists.

- This study must enhance the methods which are used to create the low-resolution models. Perhaps the proposed procedures might be improved beyond that to achieve better results. Furthermore, this study compares a new approach (also inspired by the literature research) to the previously established low-resolution image models. Perhaps the approach used influenced the ability to categorise at low image resolutions.
- To discover probable image resolution 'boundaries' regarding performance, this study needs to explore a spectrum of image resolutions. But maybe some minimum whole-slide image resolution is necessary for adequate result confidence, whereas exceeding some higher resolution does not appreciably increase result confidence and hence wasted resources.

This paper is as follows: the literature review will be seen in the next part from either a traditional or modern approach. Additionally, earlier research that was published using low-resolution images is also discussed. Section 3 of the study presents an empirical investigation using a whole slide image classification design, dataset description, and processing. The outcomes of the suggested strategy are reported together with some comparisons in section 4. The paper is concluded with some findings and suggestions for future research in section 5.

II. Literature Review

This section establishes the medical background for the diagnosis of lung cancer in pathology slides. The history of lung cancer detection is examined, from human traditional diagnosis to automated state-of-the-art DL algorithms. Other low-resolution image analysis approaches from the larger world of clinical image processing are also addressed. As a result, this section summarises the study within the framework of medical image analysis and gives design recommendations for the proposed produced low-resolution image models.

A) Traditional Approach

Traditionally, the glass slides are examined under a microscope by a pathologist. A pathologist is a doctor who identifies and analyzes illness involving cells and tissue samples to assess if any metastases are visible or not [2]. The process is locating and identifying minor lesions over the whole slide space. A pathologist aims to discover three unique types of lung cancer lesions as shown in Table 1.

Table 1. Types of metastases and their corresponding size information

Metastases Type	Size Information
Macro-metastases	More than 2 mm
Micro-metastases	Greater than 0.2mm in diameter and/or comprising over 200 cells, but just not exceeding 2mm in diameter.
Individual Tumour Cells (ITCs)	A single tumour cell or a group of tumour cells no larger than 0.2mm or containing no more than 200 cells.

Diagnostic accuracy in classical lung cancer diagnosis is based on a pathologist's expertise, which is accumulated through many years of observations of various patient tissue samples and verified diagnoses [3]. However, high diagnostic accuracy cannot be assured. Even the most skilled pathologists can detect metastases, especially micro-metastases and solitary cancer cells, which are hard to identify due to the huge amount of tissue inspected on each slide. Although if pathologists were 100% correct, the procedure is time intensive and labour expensive, demanding lengthy reading time through microscope examination [2]. Pathologists examine harmless tissue i.e. no metastases in the majority of cases; 60-70% of lungs have no metastases. As a consequence, a significant percentage of a pathologist's time is spent essentially confirming negative cases. Tissue samples are susceptible to misunderstanding and may need repeat examinations by several pathologists when metastases are discovered. A historical analysis found that reviewing slide diagnoses by pathology specialists modified the nodal position in 24% of cases [4].

B. Modern Approach

A subfield of digital pathology using histopathology slides is called the study of illness in tissue specimens using virtual microscopy. ML approaches were used for modern cancer diagnosis as medical data and computational power became more available for study and research work [2]. Human engineers generally worked with pathologists to establish hand-crafted aspects of the issue domain (e.g., mitoses, consistency of cell size, uniformity of cell shape) [2]. Although the characteristics are chosen by hand, their relationship to the forecasts is learned automatically. This represented a transition in lung cancer detection from manual human-based diagnostic methods to automated systems educated by computers with human-driven hand-crafted characteristics [5]. Such technologies have the potential to reduce variability and subjectivity in pathology diagnosis [6]. The usage of hand-crafted elements provides predictability and is thus natural to a pathologist or physician. Since decision responsibility becomes essential, and it is important in diagnosis [5]. Deriving such hand-crafted qualities, however, is difficult since it requires a fundamental understanding of the nature of the illness and its presentation inside the tissue. This problem motivates research into ML and DL approaches including automated feature discovery investigated in the following subsection.

1) Machine Learning Approach

Medicine and healthcare have made significant advances in the previous forty years. The true causes of several diseases were discovered, new remedies were invented and newer diagnostic procedures were devised during this time. Even with all of our successes, illnesses like cancer continue to stalk

us because we are still vulnerable to them. Cancer is the world's second greatest cause of mortality. Masud et al. [7] have developed an ML approach to lung and colon cancer diagnosis. ML-based lung cancer prognosis models have been developed to help doctors to manage incidental or screen-detected indeterminate pulmonary nodules. Such methods may be able to minimise variability in nodule categorization, enhance decision-making, and, eventually, reduce the number of benign nodules that are followed or worked up ineffectively according to Kadir & Glees [8]. Since Cancer detection is difficult for medical personnel, lung cancer has the potential to be fatal. Cancer's real cause and complete therapy are still to be identified but it can be cured if detected early enough. Image processing procedures such as feature extraction, noise reduction, identifying damaged regions, and maybe a correlation with data on the medical history of lung cancer are utilised to discover cancer-affected areas of the lung. This study demonstrates accurate lung cancer categorization and prediction utilising technologies provided by ML using image processing. The geometric mean filter is utilised during image preprocessing with different patients and datasets. As a result, image quality improves. The images are then segmented using the *K*-means approach. This segmentation may be used to locate a portion of an image [9]. Mukherjee & Bohra [10] predicted that in the future, new imaging such as X-ray, Computed Tomography (CT), Magnetic Resonance Imaging (MRI), and Positron Emission Tomography (PET) would be examined, resulting in greater precision and allowing medical practitioners to provide fast prophylaxis at a cheap cost. Because this disease has a negative economic impact, more research can be conducted to uncover information gaps in disease control and detection methods, which can aid in the development of vaccines or other control strategies.

2) Deep Learning Approach

Several DL papers have also been proposed such as Wang et al. [11] proposed DL with weak supervision for whole-slide lung cancer image analysis. Lung cancer is the main cause of cancer mortality in both men and women, as we all know. Appropriate therapy for lung cancer patients is mostly determined by the kind of cancer, including small cell lung cancer (15%) or nonsmall cell lung cancer (85%). His technique initially employs a patch-based Fully Convolutional Network (FCN) to recover discriminative blocks and then efficiently offers representative deep features. Murugesan et al. [12] has investigated that a hybrid DL model is used for successful lung nodule segmentation and classification from CT images. DL models are employed to determine lung nodule irregularities early in their development. The fundamental purpose of this paper is to correctly detect lung cancer, which is crucial in saving someone's life. For decades, people all across the world have been concerned about lung cancer. Several researchers presented multiple difficulties and ideas for varying phases of a computer-aided system for early-stage lung cancer diagnosis, as well as lung cancer information. One of the fields of Artificial intelligence (AI) is computer vision, which is a better technique to identify and prevent lung cancer. This work focuses on the pre-processing, segmentation, and classification modelling steps involved in recognising lung tumour areas. In pre-processing, an adaptive median filter is used to identify the noise. The work's novelty aims to develop a simple yet effective model for the quick detection and segmentation of lung nodules using U-net architecture. This method focuses on recognising image normality and irregularities to identify and segment lung cancer. Tekade [13] has proposed DL for lung cancer classification and detection. They stated that with the help of image processing and DL approaches, early diagnosis of lung cancer has become highly crucial and also quite simple. This research presents a 3D multipath VGG-like network for 3D cubes derived from the Lung Image Database Consortium and the Image Database Resource Initiative (LIDC-IDRI).

C. Low-Resolution Images in Medical Image Analysis

Modern cancer detection methods have primarily relied on single-resolution models with high-resolution images [14]. There is one component that contributes to improved clinical viability is increasing efficiency by decreasing image input resolution, which is the subject of this study. Creating successful models that use low-resolution images has the potential to produce high-performance outcomes with minimal resource needs compared to high-resolution models. Low-resolution image processing medical image analysis research falls into one of two areas. The initial step is to create single-resolution models from low-resolution images. The second is the creation of multi-scale systems that mix low and high-resolution analyses. The second type of research is more abundant than the former; few attempts have been made to explore in depth ability of models employing strictly low-quality images for medical image processing tasks and we believe this study is among the first. There is an obvious reason for the disparity in research between the two categories: single-resolution models using low-resolution images forego the benefits of using high-resolution images i.e. detailed features based on greater spatial resolution, whereas multi-resolution schemes strike a balance between the two, prioritising both model performance and resource requirements.

A) Single-Scale Techniques

The following recent techniques are noteworthy in terms of single-resolution models employing low-resolution images: Patil et al. [15] introduced a CNN-based architecture for invasive breast cancer segmentation. The method employs very low-resolution images i.e. 320 x 320 pixels. The model is stated to have done well when the malignant patches occupied a considerable fraction of the extremely low-resolution images, but much poorer when the cancerous regions were smaller. It suggests that a created low-resolution paradigm may work well on macro-metastases, which can be rather big about total tissue but perform poorly on micro-metastases. Zormpas-Petridis et al. [16] established a computationally efficient methodology i.e 'SuperHistopath' for mapping global context characteristics by classifying and segmenting such in images, therefore representing the rich tumour morphological variability. The method includes segmenting low-resolution

images at 5x magnification into super-pixels, which are then classified using a CNN ('Xception' and a bespoke CNN). It was demonstrated that 'SuperHistopath' is efficient for training and inference (classifying images in 5 minutes and network training in as little as 30 minutes). 'SuperHistopath' classified lung cancer images into six predetermined tissue classifications of involvement i.e. stroma, tumour, necrosis, cluster of lymphocytes, fat, and lumen/ space with an overall accuracy of 93.1%, an average precision of 93.9%, and an average recall of 93.6% using 'Xception' and 91.7%, 92.5%, and 91.8% respectively using the custom CNN over 10,349 super-pixels in a test set. Lastly, super-resolution images from low-resolution images using a Generative Adversarial Network (GAN) have been developed by Shahidi [17]; a revolutionary technique. Because images have high hardware and storage requirements, this paper is motivated by the idea of being able to store images at low resolution to use super-resolution approaches to ensure classification results on the super-resolution images are comparable to the performance using the original high-resolution versions. Despite this, the major point of interest from this work is the performance comparison of controlled models i.e. ResNeXt-101 for tumour classification utilising low (64 x 64 pixel patches) and high-resolution (256 x 256-pixel patches) images. High-resolution images performed better (99.49% accuracy score), but not significantly better than low-resolution images (95.82% accuracy score).

B) Multi-Scale Techniques

In the literature, two types of multi-scale approaches have been observed. The first method is to employ CNN ensembles, in which images are analysed at numerous resolutions by various models and the results are combined. This method tries to improve model performance over single-resolution models by combining contextual data collected at lower resolutions (e.g., inter-tissue structures) with detailed characteristics gained at higher resolutions (e.g., cell morphology and intra-tissue structures) because inference uses several models incorporating higher quality images and efficiency increase is not an incentive for this technique. Tokunaga et al. [18], for example, have provided a segmentation approach that can adaptively employ image attributes from different resolutions to separate several cancer sub-type areas in an input image. By adaptively altering the weight of each expert based on the input images (through a 'weighting CNN'), the approach combines much segmentation CNN i.e. an ensemble that has been trained using different image magnifications (5x, 10x, and 20x magnification). The approach makes use of characteristics obtained from both broad and narrow field-of-view images, which may be valuable for identifying sub-types. As contrasted to state-of-the-art connections (UNet, SegNet, DeepLabV3+) taught at each of the magnification levels studied, the technique performed better for various segmentation tasks such as binary segmentation of malignant tissue and multi-class segmentation of tissue sub-types. The second type of multi-scale technique involves the employment of a mechanism that prioritises low-resolution image processing while allowing for additional high-resolution processing with low confidence outcomes. In obscure and challenging instances, this emulates pathologist methods by first analysing tissue at low magnification before zooming in on regions of interest. Traditional patch-based processing approaches do not represent the relatively efficient way pathologists explore slides in clinical microscopy. This sort of multi-scale strategy aims to achieve equivalent performance to ensemble or single-resolution systems employing high-resolution images while lowering average inference time. Maksoud et al. [19] provided a method for selectively using high-resolution processing at low resolutions based on prediction confidence. A decision process is used in their system to identify whether a low-resolution or high-resolution network should be used. When they used their technique to 'liver-kidney-stomach' images, they were able to enhance multi-class classification accuracy while lowering inference time by a factor of 7.74 (when compared to existing multi-scale systems). This technique is reported to be based on past work done by Dong et al. [20]. In this case, lung cancer segmentation is performed by images utilising low or high-resolution images as decided by a trained policy network through reinforcement learning, which determines if zooming in on regions of interest is necessary or not. Hering and Kybic [21] describe a novel multi-scale technique for detecting malignant tissue in histology images. In their technique, many tiny sections of interest, dubbed "glimpses", are analysed rather than the whole high-resolution images. The glances create a tree structure; low-resolution glimpses dictate the position of multiple higher-resolution glimpses, and so on until the highest resolution is attained. A whole-slide classification score is computed after classifying glances at the greatest resolution (the highest prediction score awarded to high-resolution glances). They reported a mean Area Under Curve (AUC) score ranging from 0.92 to 0.95 (based on model parameters), which is similar to the best AUC score. This proved that high confidence classification can be performed using only a tiny section of the high-resolution image, resulting in a significant speedup with only a minor performance degradation.

III Empirical Study

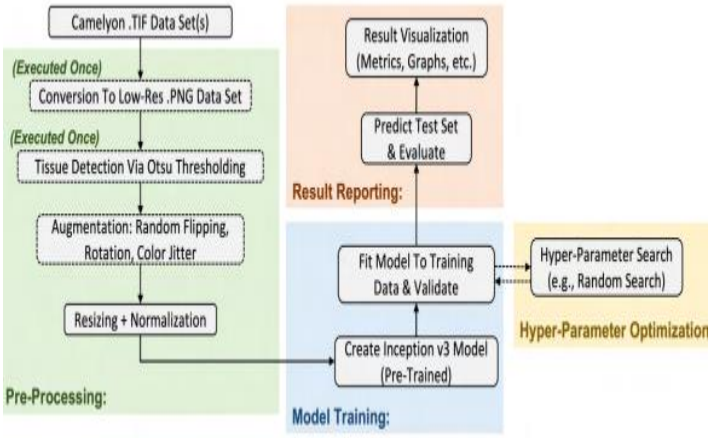
Single-resolution models based on low-resolution images have experienced some success in medical image processing tasks in summary. This does not, however, imply that the low-resolution model built in this effort will likewise provide high-confidence predictions; variances in medical activities performed are a major component. A low-resolution model, for example, will be better at segmenting huge tissue subtypes than smaller homogenous lesions. Meanwhile, research is centred on multi-scale techniques that, by default, employ low-resolution images but can switch to high-resolution processing when necessary. Because such techniques are intended to minimise inference time on average while preserving high network capabilities. The requirement for absolutely low-resolution image models may be questioned.

However, for many medical activities, purely low-resolution models may be acceptable, reducing the hardware and storage needs involved with maintaining high-resolution images. Even though the performance of a low-resolution model is lower than that of a multi-scale method, the low-resolution model may be exceptional for a certain class, justifying its adoption. A low-resolution model that consistently diagnoses benign tissue, for example, can filter the majority of negative instances, allowing pathologists to focus their efforts on difficult cases that require attention.

High-level design decisions are described and justified in this section, with a focus on the DL pipelines established for whole-slide classification and cancer localization. The study conducted in the literature review strongly influenced these conclusions. Where necessary, implementation details that are unusual given the design requirements are detailed. A Graphics Processing Unit (GPU) was used to give additional computing capacity because of the large amount of memory required by the datasets used, as well as the intricacy of the deep CNN models generated. To capitalise on the GPU's processing power, a DL framework with Compute Unified Device Architecture (CUDA) support (for GPU acceleration) and CNN-based implementations was chosen [22]. PyTorch was chosen above similar options, including the more prominent TensorFlow with Keras framework, due to the parties participating in this studies expertise with PyTorch. TorchVision/SKImage for image processing, RayTune for automatic hyper-parameter optimisation, and ASAP/OpenSlide for manipulating high-resolution images have been included in this framework.

A. Whole Slide Classification Design Pipeline

Image analysis in histopathology is the gold standard for cancer diagnosis. Whole slide image classification is a difficult task. In histopathology, it is becoming a regular clinical practice due to the fast advancement of medical imaging technologies. The extraordinarily high spatial resolution of whole slide images makes DL for digital pathology difficult. Numerous DL algorithms for the diagnosis of lung cancer have indeed been suggested in order to aid pathologists and researchers in their job. This section of the whole slide classification design describes the DL pipeline architecture for the whole-slide classification. Pipelines, as we all know, are a method of simplifying the ML workflow by allowing data to be converted and correlated into a model, which can then be examined to provide outputs. For whole-slide classification, the DL pipeline is easily defined by reducing the high-resolution data set images to a sufficiently tiny size that they can be given as input into a pre-trained CNN for binary classification i.e. cancerous or normal. However, additional strategies were used to improve the overall performance of this simple strategy. Model training, pre-processing, hyper-parameter optimization, and result reporting are the four stages of the pipeline as shown in Figure 1.



Note: Dashed lines indicate pipeline steps added to create additional pipeline versions to improve general performance. Some pre-processing steps are executed once, and their results are used in the pipeline as opposed to running the steps every time a model is trained, which prevents wasted computation time.

Figure 1: Whole slide classification overview

B. Dataset and Dataset Split

In this study, we used the lung cancer dataset from The Cancer Genome Atlas (TCGA) [23]. Two subtypes of lung cancer, lung adenocarcinoma and lung squamous cell carcinoma, are represented in the WSIs, which can be obtained from the National Cancer Institute DataPortal in the form of 1054 diagnostic digital slides. The dataset is divided into subsets for training and testing. There are 216 images in the training subsets and 129 in the testing subsets as shown in Table 1. Utilizing stratified random sampling, the training data was further divided into training and validation partitions while maintaining the proportion of output classes. In automated lung cancer detection using machine learning, a split of 80% training to 20% validation was used. It was made sure that the stratified random sample kept both the intra-class proportion for the positive cancer class as well as the proportion of the binary output classes (i.e., cancer and normal). Without assuring the intra-class percentage, the validation set might have had an excessive amount of macro-metastases, which are probably simpler to detect given their size, leading the model to report doing better on the validation set than would be seen generally.

Table 1. A summary of the dataset's data distribution

	Train	Validate	Test	Total
Normal	127	32	80	239
Cancer	89	22	49	160
Total	216	54	129	399

A) Output Class Balance

A study of the output class distribution was critical in deciding if re-balancing was needed to avoid model bias towards one of the output classes since the model performs binary classification as shown in Figure 2. The training data contains a positive i.e. cancer class proportion of 38%, which corresponds to the real-world statistic that 60-70% of retrieved images do not feature metastases. While the observed output class imbalance is not significant, re-balancing prevents any

modest class bias. Several solutions are being examined to rebalance the data. Under the sampling scheme, normal slides would be dropped to match the number of cancer slides, but this would reduce the restricted amount of majority class data, making it feasible to eliminate important characteristics. Oversampling the minority class, for example, with the Synthetic Minority Oversampling Technique (SMOTE) [24], will indeed balance the output classes and raise the quantity of information available to the model, but it is more computationally expensive than the simpler technique used: class weights. When constructing the loss function, class weighting prioritises the under-represented class and does not require any further modification of the training data. Given its capacity to substantially expand the size of the training dataset, the application of SMOTE for image data might be examined in future studies. The percentage of normal to cancer training samples is used to calculate the class weight assigned to the positive cancer class:

$$\frac{382\ normal\ slides}{236\ positive\ slides} = = \sim 1.62$$

As a result of this part, we know that the cancer class is weighted 62% greater in the loss function since there are 62% more normal samples than cancer samples.

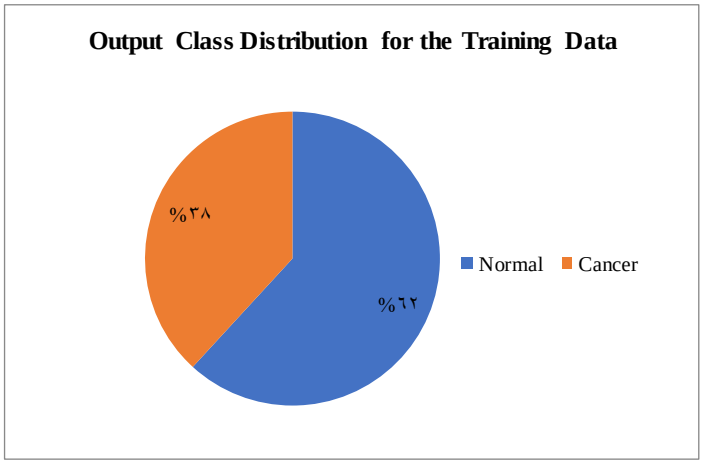


Figure 2: Pie chart displaying the output class distribution between the training data's positive cancer class and negative normal class

B. Label Encoding

The output class labels are given in categorical text format, however, they were transformed into a numerical format for model training and inference. Because the model performs binary classification, the target classes have been encoded in a simple binary format as shown in Table 2.

Table 2. Table showing binary encoding of the output classes.

Category Level	Binary Encoding
Normal	0
Cancer	1

C. Pre-Processing

This section describes the design of the pre-processing performed on the images before feeding the images as input to the model as shown in Figure 3.

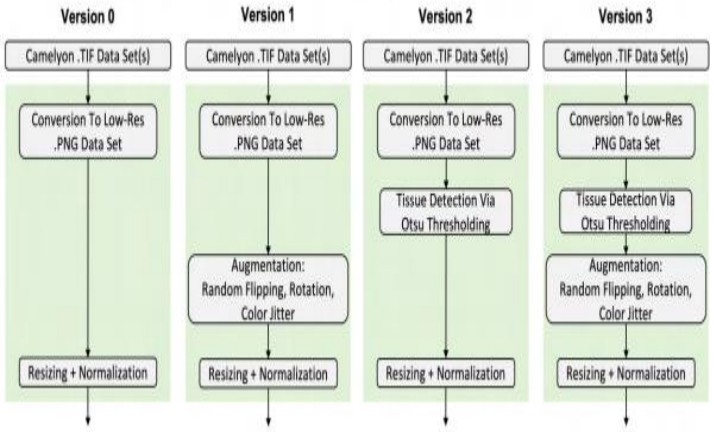


Figure 3: Whole slide classification pre-processing overview

A) Dataset Conversion to Low-Resolution Images

In this sub-section, we will see dataset conversion to low-resolution images. The high-resolution (.TIF) images have been reduced to low resolution. As a one-time pre-processing step, (.PNG) images are used i.e. not always done out through inference and the model training. This resulted in speedier model training (the very high-resolution images only needed to be down-sampled once) and simpler model debugging (at low resolution, the input images are more controllable and inspectable.). The high-resolution images were initially compressed to assess or investigate multiple model versions utilizing 299 x 299-pixel images into resolution level 8. Each pixel in such images represents 62.2 μm^2. Because of dimension discrepancies between high-resolution images from various medical centres, not all images supported resolution level 8 image downsampling. In these circumstances, the level with the next lowest resolution was chosen. As a consequence, the low-resolution (.PNG) images can only be 1000 x 1000 pixels in size. As a result, these (.PNG) images may be supplied as input to the deep CNN model in their entirety.

The high-resolution images have to be down-sampled to 6000 pixels in width and height for subsequent tests of other input image dimensions up to 2048 x 2048 pixels i.e. resolution level 6. As a result, tissue detection would provide tissue images of about 2048 × 2048 pixels, as required. From such tissue images, input resolutions less than 2048 × 2048 pixels might be studied. Using the baseline dataset downsampled to resolution level 8 would have been pointless; the images would have been upsampled for all requested input resolutions greater than roughly 1000 x 1000 pixels, resulting in little information gain i.e. these images' upscaling is not the same as 'zooming out' in magnification.

B) Data Augmentation

To artificially increase the image collection, an elastic deformation data augmentation approach is presented. So, data augmentation approaches were iteratively incorporated based on intuition and literature survey research to develop several

model versions, each aiming to improve on the previous one. There are four different model variations as shown in Figure 3. Augmenting the training images acts as a type of regularisation, preventing over-fitting and hence improving overall model performance.

- **Version 0:** There was no data augmentation used in Version 0. This model version serves as a baseline against which subsequent versions may be evaluated to confirm that the augmentation approaches enhance the model.
- **Version 1:** This model version includes common image adjustments to improve overall model performance. Because the slides have no canonical orientation. The image output labels are orientation-invariant, hence random vertical and horizontal flipping of the images is used. In addition, the image is rotated randomly at 90-degree intervals i.e. 0, 90, 180, and 270. While image rotations may be done in a continuous range e.g. [0, 360], this resulted in low-resolution images with either a considerable proportion of pixels cut out by the rotation, or low-resolution images with a significant proportion of 'empty' pixels. Both scenarios are extremely damaging to a low-resolution model with limited pixel content.
- **Version 2:** To improve model input, this model version includes tissue detection but without the version 1 transformation mechanisms. Previous versions sent whole low-resolution slide images into the deep CNN. However, the majority of the pixels in these slide images are background; only a small fraction of the pixels belong to the significant tissue region that informs the outcome. As a result, tissue detection was employed to identify the tissue area inside low-resolution slide images, which was then removed and fed into the model. The RGB images are first transformed to the Hue-Saturation-Value (HSV) colour scheme. The H and S channels are then Otsu thresholded to automatically identify ideal foreground threshold values. The H and S channels are combined into a binary image using their best thresholds, yielding a mask of the identified tissue patches in the image. To remove spurious regions, median filtering is performed on the tissue mask; however, filtering is kept to a minimum to avoid over-filtering, which may remove tiny tissue regions, which might represent micro-metastases. Finally, the identified tissue's bounding box coordinates are derived from the mask and utilised to extract the relevant region from the original RGB images. Later, immediately before calculating bounding box coordinates, modest dilatation was introduced to the identified tissue in the tissue masks so that the resulting cropping did not confine the tissue region as firmly. This allowed the identified tissue's margins, which define the bounding box's boundaries, to be kept in cases where a little (but considerable at the cellular level) cropping out of tissue content was seen. Initially, an examination of the generated tissue images indicated that visual information, such as text and black crosses, resulted in undesired tissue segmentation.
- **Version 3:** The tissue identification mechanism from version 2 is combined with the processing methods or transformation techniques from version 1. As a result, for the regularization of low-resolution tissue-identified images, this model leverages data augmentation.

C) Data Preparation

Data preparation is conducted as a pre-processing step on all images (train, validate, and test) immediately before feeding the images as model input. Image scaling and normalisation are two aspects of data preparation. Image scaling is the process of changing low-resolution (.PNG) images to the necessary low-resolution dimensions for study. Initially, image sizes of 299 × 299 pixels were used. This is the lowest input size that the selected InceptionV3 architecture allows. Investigating the input dimensions of 299 x 299 pixels will set a lower bound for model performance based on the premise that the model would likely perform worse when the image resolution is reduced. We have investigated following image dimensions: 512 x 512, 1024 x 1024, and 2048 x 2048 pixels. Investigating higher input resolutions (4096 x 4096 pixels, for example) is also acceptable because these resolutions are still low-resolution, being thousands of times smaller than the original dataset images. However, given the memory limits they impose, such resolutions are becoming increasingly infeasible for exploration, requiring extensive training time. An accumulated gradient technique was devised for usage with bigger image resolutions that would otherwise cause out-of-memory concerns, to manage increased input image resolutions (above the initial 299 x 299 pixels) while retaining the same batch size (at 16). The cumulative gradient method simply sets the batch size to one, allowing individual images to be stored in memory. The network, however, is only updated when the required batch size number of images i.e. 16 have had their losses computed. This is equivalent to just utilising the appropriate batch size, however, memory management is used for practical considerations. One noticeable issue with image scaling is that images are warped to a square (1:1) aspect ratio. As a result, the model versions strive to uncover generic traits to diagnose cancer in 'squashed' images. Tissue morphing has a detrimental impact on categorization abilities since any information on the

morphology of the cell and multicellular structures is lost. The problem is more prevalent in model versions 0 and 1 when the original rectangular low-resolution images are utilised, but it is alleviated in versions 2 and 3 because tissue-identified areas typically adhere to a roughly square area. However, the problem has not been entirely resolved and is an intrinsic flaw of the pipeline technique.

However, in this case, the images themselves are utilised as input, and they do not all have identical dimensions (as needed by the CNN design), therefore scaling is required. The issue of slide content morphing can be lessened in versions 0 and 1 by utilising a rectangular aspect ratio that corresponds better to the low-resolution slides, although the issue will still be present in a lesser form. For versions 2 and 3, extraction of the tissue areas i.e. cropping of the slides can be done to ensure a square aspect ratio, but this involves the inclusion of additional background pixels to varying degrees, which negates the benefit of using tissue detection in the first place; tissue detection seeks to minimise the amount of background present that does not inform the output.

Finally, the training images are normalised, which is a common pre-processing step to improve training by increasing model numerical stability and/or speeding up the training process. The model was initially trained without normalisation, then with the normalisation values recommended for the PyTorch pre-trained InceptionV3 model, and finally with normalisation values calculated for the low-resolution training images. The normalisation parameters were chosen since they are more appropriate than the values supplied for the pre-trained InceptionV3 model. Because the images (which feature monotonous stained tissue) vary fundamentally from the 'ImageNet' images used to train InceptionV3 (1000 classes including goldfish, umbrella, and so on), finding the mean and standard deviation for the dataset is more suitable. Furthermore, because the input data to the model is drastically modified when utilising tissue detection i.e. versions 2 and 3 of the model, the mean and standard deviation derived for the tissue-detected image are utilised for normalisation instead.

D. Model Training

This section describes the model training design that was carried out utilising pre-processed images as shown in Figure 4.

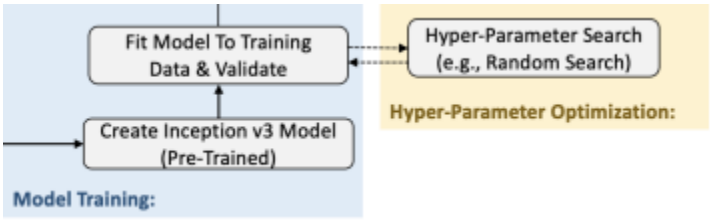


Figure 4: Model training overview

1) Deep-CNN Architecture

This section will go over deep CNN architecture. The bulk of successful DL architectures, are typically GoogLeNet, AlexNet, or VGGNet. While each of these architectures is a feasible option for this study's low-resolution models. However, the InceptionV3 architecture was chosen specifically because it delivers multiple iterations of improvement over the initial GoogLeNet design. To identify which architecture is most suitable for lung cancer diagnosis using low-resolution images, the InceptionV3 architecture might have been substituted for the other feasible designs (e.g., AlexNet and VGGNet). Moreover, the predictions of these many models might have been pooled to form an ensemble, to assess whether the overall performance could be improved even further. Assuming that the ensemble models can infer in parallel and that there is no major variance in their inference durations, using an ensemble might increase prediction performance while reducing overall average inference time.

2) Transfer Learning

The InceptionV3 architecture, which was first trained on the 1000-class 'ImageNet' dataset, was used [25]. This makes use of transfer learning, which allows the categorization of information from a broad area to be used. This is especially advantageous in this case because the short size of the training dataset restricts learning power; the limited training data is best utilised by fine-tuning the weights of an existing network rather than training a whole new one. A fully-connected layer with a single output neuron has been used at the end of the InceptionV3 architecture. Because the output neuron is activated by Sigmoid, the model delivers a single value between 0 and 1, representing the chance that the input images belong to the positive cancer class. Because the initial output classes for the pre-trained network i.e. 1000, 'ImageNet' outputs are considerably different from the output classes in this medical application. The fully connected layer described replaces the original pre-trained fully-connected layer (rather than merely adding to it). Pre-trained feature extraction layers are beneficial, while pre-trained classification layers are not. Overall, this switches the architecture for the whole-slide classification job from a 1000-class classification to a binary classifier.

3) Loss Function

Binary cross entropy loss is also known as log loss was chosen as the loss function, which is a common ML choice for binary classification problems [2]. Each anticipated output probability is compared to its matching actual output via binary cross entropy. An overall score is computed, which penalises estimated probabilities depending on their deviation from their expected output. Loss grows exponentially; the model is discouraged from forecasting low probability for positive class samples due to the high penalties.

4) Optimiser

Because of the complexities connected with the deep nature of the CNN model, it is critical to keep the number of hyper-parameters to a minimum. When compared to traditional optimisation methods such as Stochastic Gradient Descent (SGD), adaptive learning rate optimisation techniques give out-of-the-box performance due to quick convergence and less parameter fine-tuning. Adaptive moment estimation (Adam), which combines momentum for greater steps in the direction of the sharpest gradient and root means square propagation, is the most often used adaptive optimiser (for higher acceleration

on steep slopes). Adam was chosen as a good optimiser for these attributes, even though many alternative optimisers and versions of Adam exist. An enhancement would be to do an empirical inquiry to determine which optimiser produces the best results for the dataset. This was not prioritised because of the complicated interaction of model variables, where a change to one variable may affect the optimum decision for all other variables, requiring significant work for what would amount to a marginal benefit in terms of study objectives.

5) Early Stopping

As previously stated, the original training data was divided into training and validation partitions with an 80/20 ratio. Model predictions on the validation set are done at the concerning epoch to produce an epoch validation loss. To avoid over-fitting on the training data, training is stopped if the validation loss has not dropped substantially after a certain number of epochs. The number of epochs allowed without improvement was set to 10% of the maximum number of epochs for training. This percentage strikes a compromise between the necessity to avoid over-fitting and the patience required to avoid prematurely discontinuing training owing to local minima. This design could be improved by using *K*-fold cross-validation, which allows for training using all of the available training data, avoidance of validation set over-fitting, and more accurate reporting of model performance on unseen data. *K*-fold cross-validation, on the other hand, splits the training data into *K* subsets and trains *K* separate models. Due to study time restrictions, the needed time to train a model (in the order of hours), and inexperience with its implementation in PyTorch, the usage of *K*-fold cross-validation was not prioritised, and a basic validation set was decided instead.

6) Hyper-Parameter Optimization

Tuning model hyper-parameters is a critical step in ensuring the DL pipeline's capabilities. Batch size, learning rate, and a maximum number of epochs are the hyper-parameters that are adjusted. An automated hyperparameter optimisation strategy was applied. First, a broad hyper-parameter space search is done on the basic model (i.e., version 0 with no data augmentation) to identify reasonable hyper-parameter values with which each of the model versions may be trained for version comparison. The best model version is then subjected to a search in a tighter hyper-parameter space to good the hyper-parameter values and optimize the whole-slide performance of the classifier. When employing 299 x 299-pixel input images, none of the model variants showed much promise at first. Thus, optimisation was conducted on model version 3 because it is theoretically the best, however, this could not be practically guaranteed. By randomly selecting from the hyper-parameter search space, random search provides a good compromise between execution time and obtaining adequate hyper-parameter values. Grid search, as an alternative, ensures that the hyper-parameter values supplied are optimal by exhaustively traversing the search space, but it takes a long time, especially for larger search spaces. Despite this, the automatic hyper-parameter optimisation system allowed for both types of search to be employed; grid search could be used for hyper-parameters with few potential values (i.e., batch size), while random search could be used for remaining hyper-parameters with a wider range of values (i.e., learning rate and a maximum number of epochs).

IV. Result Analysis

The metrics used to evaluate low-resolution models are established in this section, and the associated results for the whole-slide classification job are provided and analyzed.

A) Evaluation Metric

The key parameter used to evaluate the low-resolution models in this study allows for relevant performance comparisons [14]. A slide-based assessment for the entire-slide categorization problem has been identified. Additional metrics are utilised to compare resource requirements or to aid in further investigation of the produced models. Secondary metrics (precision, recall, etc.) are derived from the confusion matrices of the binary classifiers, for example, to analyse the limits of the whole-slide classification models when suitable.

1) Slide-Based Evaluation

The AUC- Receiver Operating Characteristic Curve (ROC) is the measure. This will be known as the AUC score. For all classification threshold values, the ROC curve is a plot of the true positive rate i.e. sensitivity vs the false positive rate.

$$True\ Positive\ Rate = \frac{True\ Positives}{True\ Positives + False\ Negetaives}$$

$$False\ Positive\ Rate = \frac{False\ Positives}{True\ Positives + False\ Negetaives}$$

Furthermore, given the reported AUC scores, percentile bootstrapping was utilised to produce 95% confidence intervals. Providing such confidence intervals tries to evaluate variations in performance found while training and inferring models with different seeds (the established pipeline employs seeding to assure repeatability of findings, although seed selection has an impact on model performance).

2) Average Inference Time

This research places a premium on the average inference time of the created models. The goal is to increase the clinical feasibility of automated lung cancer detection systems. The time required to transform the original high-resolution images into low-resolution model inputs, as well as the time required for the model to use the input to create a prediction, is included in the average inference time for the developed low-resolution models. Average inference times are recorded in all scenarios where an 'Nvidia GeForce GTX 1060 (6GB)' GPU was employed for execution.

3) Model Version Comparison Results

To produce a reasonable starting set of hyper-parameter values, hyper-parameter optimisation was done on model version 0. Then, using these values as a reference, the

following version models were trained. All hyper-parameter values were maintained constant, except for subsequent model versions that included data augmentations i.e. flipping, rotation, and colour jitter to allow for adequate utilisation of the enhanced images. When additional epochs are allowed, the implementation does not expand the dataset size, but rather performs the alterations at random, which is functionally similar. In all cases, the average (per image) inference time includes the average conversion time (360 milliseconds) from a high-resolution (.TIF) image to a low-resolution (.PNG) image with black image content filtering applied, as well as the average time required for the model to infer a low-resolution image. When tissue detection is used, the average (per image) inference time includes the time it takes to transform a low-resolution (.PNG) image to a low-resolution tissue-detected image (190 milliseconds). The results for each model version while predicting the test dataset at the initial input resolution of 299 × 299 pixels are shown in Table 3.

Table 3. AUC score and average inference time comparison between the various model iterations utilising 299 x 299 pixel input images.

Model Version	AUC Score (With 95% Confidence Level)	Average Per Image Inference Time (Seconds)
0 - No Data Augmentation, Original Low-Res Images	0.527 [0.429 - 0.638]	0.427
1 - Data Augmentation, Original Low-Res Images	0.363 [0.274 - 0.474]	0.427
2 - No Data Augmentation, Tissue Detected Images	0.427 [0.339 - 0.543]	0.637
3 - Data Augmentation, Tissue Detected Images	0.540 [0.500 - 0.500	0.637

In the above table, the overall performance for the various model versions, and the reported AUC scores are confusing and deceptive. The AUC score should be between 0.5 and 1. Further investigation of the developed low-resolution models revealed that none of these models has any significant skill in determining whether metastases exist within the 299 x 299-pixel images. The developed low-resolution models are incapable of distinguishing between metastatic and normal slides as shown in Figure 5 and Figure 6. There are various explanations for the models' low capabilities, the most common of which is a faulty training method or insufficient training data. However, given the high confidence in the pipeline implementation's correctness, and the fact that these models are observed to converge to a validation loss minimum, the following conclusion can be drawn: The built pipeline's very low input resolution of 299 x 299 pixels is just insufficient for effective categorization.

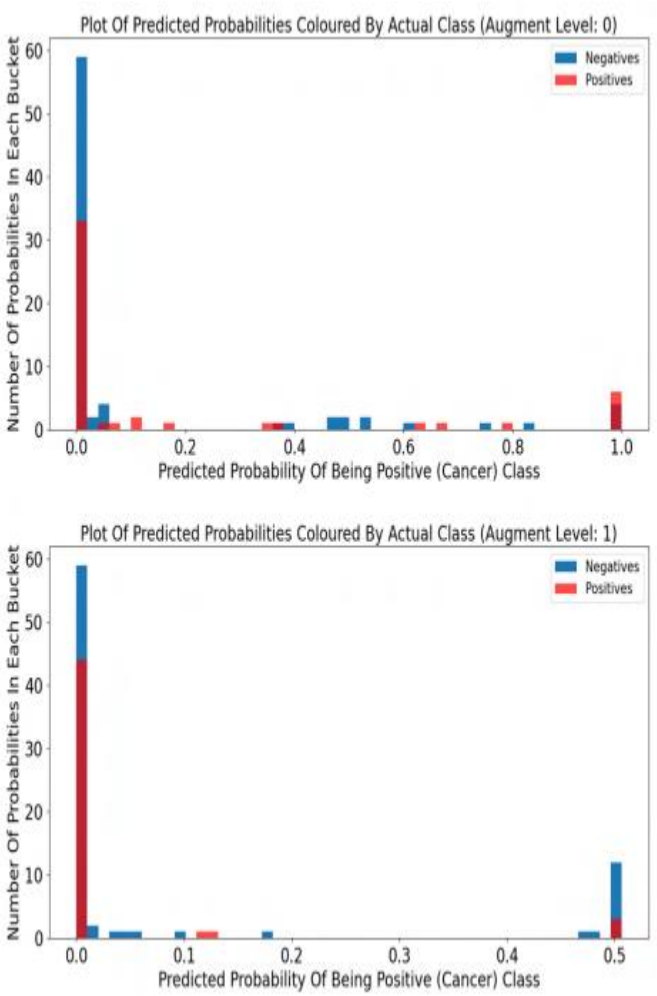


Figure 5: Figures for model versions 0 and 1 that display the expected probabilities for the test data set images (at 299 by 299 pixels) coloured by their actual output class

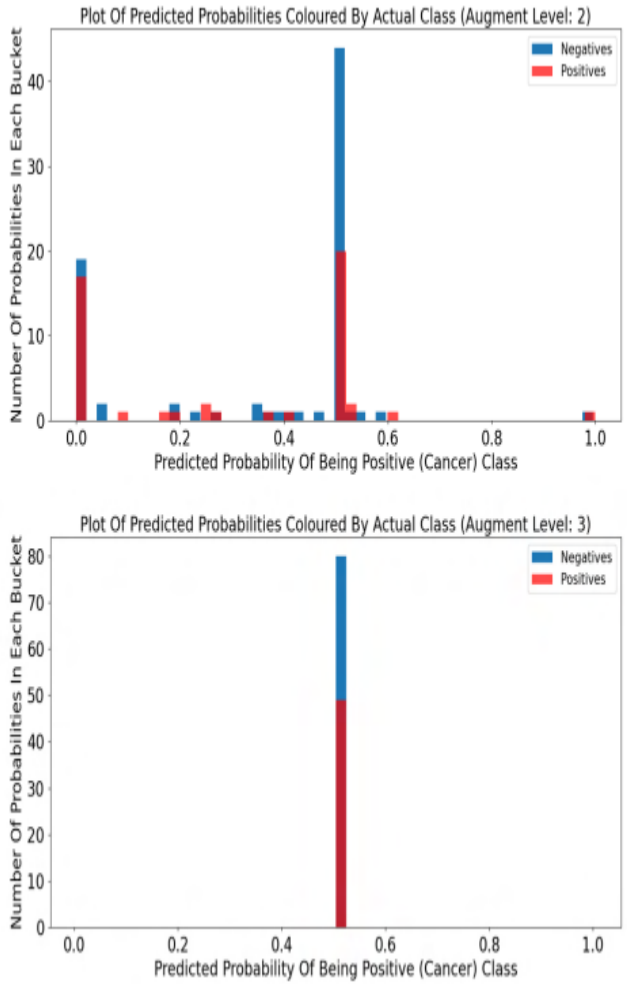


Figure 6: Figures for model versions 2 and 3 that display the expected probabilities for the test data set images (at 299 by 299 pixels) coloured by their actual output class

Figures 5 and 6 show that model versions 0 and 2 (the ones that do not employ the flips, rotation, and colour fluctuation augmentations) have a wider range of projected probabilities. The broader dispersion implies that these models are attempting to make educated selections rather than merely predicting one of the classes (yet, all models predict some likelihood). This might imply that the data augmentation technique is insufficient. However, the inverse might also be true. Model versions 1 and 3 may have learned orientation and colour-invariant characteristics by using data augmentations, increasing the confidence of predictions. Model version 3, which uses data augmentations to recognise tissue, predicts that both groups are about equally probable with low fluctuation as shown in Figure 6. This, on the other hand, suggests that the model for version 3 has either learned nothing from training or has learned that it cannot forecast with any accuracy given the low pixel information available in the 299 x 299 images. As an aside, several result analysis approaches were incorporated to enhance comprehension of the created models, most notably the development of confusion matrices with the class distribution of false negative data (i.e., macro-metastasis, micro-metastasis, and ITCs). These were meant to analyse the sorts of slides and cancer the models had poor classification ability. However, because the models do not separate the output classes, an appropriate threshold for generating the confusion matrices could not be chosen (so the analysis described could not be performed).

D. Image Resolution Comparison Results

In the above subsection, it was discovered that the model versions generated in this study were practically incapable of identifying the presence of cancer in low-resolution images of 299 x 299 pixels. This is most likely due to an intrinsic constraint of the methodology: there are too few training images to allow for adequate training, the images themselves lack sufficient information for classification, or both aspects are to blame. There is nothing that can be done to explore the former issue because more training samples are just unavailable for usage. As a result, this section examines model performance at various input resolutions, all of which are still deemed low-resolution when compared to the original high-resolution (.TIF) images., Additional image dimensions studied include 512 x 512 pixels, 1024 x 1024 pixels, and 2048 x 2048 pixels (299 x 299 pixels were initially used). Pipeline version 3 was chosen for use across all of these input image resolutions. Version 3 should theoretically deliver the best overall performance (however this could not be empirically shown in part 4.2.1) by utilizing regularisation through data augmentations and higher quality input images through tissue identification. Furthermore, the hyper-parameter values utilised were the same as those used in version 3 with 299 x 299-pixel inputs to confirm that any improvement in model performance is due to the improved resolution of the input images.

As predicted, bigger input resolutions need more time to infer the original high-resolution images. There will be a significant rise in average inference time between 299 x 299 pixels (0.627 seconds) and 512 x 512 pixels (12.070 seconds), but only a little increase between 512 x 512 and 1024 x 1024 pixels (12.100 seconds). The original images are downsampled to resolution level 8 for the 299 x 299-pixel model. The dark content filtering and tissue detection processes are performed once on images of about 1000 × 1000 pixels, which are subsequently reduced to tissue images at the appropriate input resolution. The 512 x 512 and 1024 x 1024 models, on the other hand, obtain their inputs from tissue identification performed on the original images rather than being down-sampled to about resolution level 6. As a result, the one-time pre-processing is performed on images of about 6000 × 6000 pixels (which takes substantially longer) to obtain the requisite tissue images of the required input resolution. Given that image up-scaling does not equal an increase in magnification, it was vital to ensure that input images were produced via down-sampling and pre-processing from higher-than-required resolutions. Initial examination of the obtained AUC values in Table 4 leads to the conclusion that model performance increases with increasing input resolution; the AUC score has

grown marginally from 0.500 at 299 x 299 pixels to 0.540 at 1024 x 1024 pixels. This lends credence to the idea that as input resolution grows, the model should become more capable of distinguishing malignant tissue due to the additional spatial.

Table 4. For model version 3 across various input image resolutions, the AUC score and average inference time are shown.

Image Resolutions (Pixels)	AUC Score (With 95% Confidence Level)	Average Per Image Inference Time (Seconds)
299 x 299	0.510 [0.500 - 0.500]	0.627
512 x 512	0.520 [0.429 - 0.628]	12.070
1024 x 1024	0.550[0.452 - 0.629]	12.100
2048 x 2048	Timed Out[1]	Timed Out

However, the increase in average inference time is only a surface improvement because additional analysis reveals that these models are equally incapable of differentiating the output classes in the test dataset as shown in Figure 7, as previously discovered when analysing the different model versions. As the AUC score increases, models with greater input resolutions can be considered to be somewhat less unable of differentiating the output classes. Regardless, these models are unsuitable for usage since they effectively forecast at random; the models predict scores near 0.5 at all input resolutions studied. As previously stated, this might indicate that the models had inadequate training and were unable to learn the broad characteristics necessary for confident classification i.e. not enough training samples. Alternatively, the models may have learned that no certainty can be predicted given the limited information within the input images; even though 1024 x 1024 inputs provide more than ten times the amount of pixel content as 299 x 299 inputs i.e. a magnified view of the slide, critical information that informs on cancer presence is still indiscernible. The number of cells in a specific location, for example, and the regularity of cell shapes are markers of cancer, particularly in terms of micro-metastasis and individual cancer cells, but these characteristics are obscured by blurring.

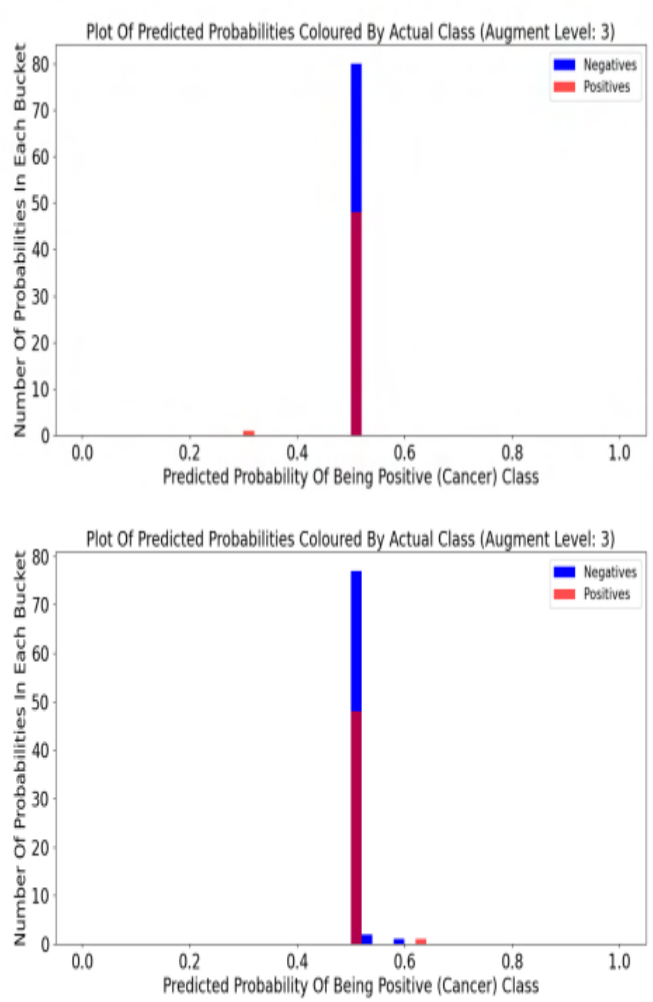


Figure 7: Figures depicting the model version 3's predicted probabilities for the test data set images coloured according to their actual output class when utilising 512 x 512 and 1024 x 1024 input images, respectively

5. Conclusion

The study's overarching outcome is that the established DL pipeline for binary cancer classification is virtually skillless at the explored low input resolutions, even when additional strategies to boost general capacity were included. The DL pipeline may gain competency at higher input resolutions than those explored, perhaps revealing an ideal resolution that captures adequate prediction performance and average inference time. However, this is implausible given the apparent trade-off between the two required features, necessitating an unpleasant decision to prioritise either model performance or inference speed to determine the input resolution to utilise. This is not the case with the multi-resolution techniques presented. Currently, the models in this research suffer from the 'garbage-in, garbage-out problem associated with DL models' black box nature. Even at 1024 x 1024 pixels, the low-resolution input images are greatly blurred at the cellular level, making the identification of critical markers for cancer detection (cell density, shape irregularity, discolouration, etc.) doubtful even by qualified

[1] The model could not be evaluated since the training process utilising 2048 x 2048 image inputs failed to attain convergence in the allocated 72 hours. This was mostly caused by the training of this model being noticeably slowed down when memory could only hold small batches of higher definition photos.

pathologists. It's also possible that the pipeline is intrinsically bottlenecked at all input resolutions because of the scarcity of training data. Simply put, there aren't enough labelled training data for the model to learn how to distinguish metastases inside homogenous-looking tissue, especially considering that the machine is seeking to combine numerous forms of cancer for classification (i.e., micro-metastases, macro-metastases, ITCs). This is an inherent problem with utilising images as inputs in their totality since the size of the training dataset becomes tied to the number of images that can be reliably annotated comprehensively by skilled medical staff.

In this concluding section, we concluded that after considering and using the input resolutions to be evaluated, the proposed DL pipeline is inappropriate as a medically feasible cancer detection method. In a nutshell, our proposed DL pipeline is inadequate as a preliminary filter to swiftly and rapidly categorise simple pathology cases, validating the usage of slower, high-performance submission of data for tough but unusual or infrequent situations. However, this study did not scientifically examine any kind of ideal input resolutions which balance resource requirements and prediction performance. On the other hand, a therapeutically practical strategy would more properly combine the speed of low-resolution slide processing with the assurance and certainty of high-resolution slide processing. Although the efficacy of such multi-resolution techniques has been proven in the research itself and they have yet to be used to the identification of metastasis in pathology slides.

References

[1] O. for N. Statistics, “Cancer survival in England - adults diagnosed - Office for National Statistics,” *Office for National Statistics*. 2019, Accessed: Sep. 24, 2022. [Online]. Available: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/datasets/cancersurvivalratescancersurvivalinenglandadultsdiagnosed.

[2] G. Litjens *et al.*, “A survey on deep learning in medical image analysis,” *Medical Image Analysis*, vol. 42. Med Image Anal, pp. 60–88, Dec. 01, 2017, doi: 10.1016/j.media.2017.07.005.

[3] W. Yue, Z. Wang, H. Chen, A. Payne, and X. Liu, “Machine learning with applications in breast cancer diagnosis and prognosis,” *Designs*, vol. 2, no. 2. Multidisciplinary Digital Publishing Institute, pp. 1–17, May 09, 2018, doi: 10.3390/designs2020013.

[4] J. H. M. J. Vestjens *et al.*, “Relevant impact of central pathology review on nodal classification in individual breast cancer patients,” *Annals of Oncology*, vol. 23, no. 10, pp. 2561–2566, Oct. 2012, doi: 10.1093/annonc/mds072.

[5] A. Madabhushi and G. Lee, “Image analysis and machine learning in digital pathology: Challenges and opportunities,” *Medical Image Analysis*, vol. 33. Elsevier, pp. 170–175, Oct. 01, 2016, doi: 10.1016/j.media.2016.06.037.

[6] J. Griffin and D. Treanor, “Digital pathology in clinical use: Where are we now and what is holding us back?,” *Histopathology*, vol. 70, no. 1. John Wiley & Sons, Ltd, pp. 134–145, Jan. 01, 2017, doi: 10.1111/his.12993.

[7] M. Masud, N. Sikder, A. Al Nahid, A. K. Bairagi, and M. A. Alzain, “A machine learning approach to diagnosing lung and colon cancer using a deep learning-based classification framework,” *Sensors (Switzerland)*, vol. 21, no. 3, pp. 1–21, Jan. 2021, doi: 10.3390/s21030748.

[8] T. Kadir and F. Gleeson, “Lung cancer prediction using machine learning and advanced imaging techniques,” *Translational Lung Cancer Research*, vol. 7, no. 3. AME Publishing Company, pp. 304–312, Jun. 01, 2018, doi: 10.21037/tlcr.2018.05.15.

[9] S. Nageswaran *et al.*, “Lung Cancer Classification and Prediction Using Machine Learning and Image Processing,” *BioMed Research International*, vol. 2022, pp. 1–8, Aug. 2022, doi: 10.1155/2022/1755460.

[10] S. Mukherjee and S. U. Bohra, “Lung cancer disease diagnosis using machine learning approach,” in *Proceedings of the 3rd International Conference on Intelligent Sustainable Systems, ICISS 2020*, Dec. 2020, pp. 207–211, doi: 10.1109/ICISS49785.2020.9315909.

[11] X. Wang *et al.*, “Weakly Supervised Deep Learning for Whole Slide Lung Cancer Image Analysis,” *IEEE Transactions on Cybernetics*, vol. 50, no. 9, pp. 3950–3962, Sep. 2020, doi: 10.1109/TCYB.2019.2935141.

[12] M. Murugesan, K. Kaliannan, S. Balraj, K. Singaram, T. Kaliannan, and J. R. Albert, “A Hybrid deep learning model for effective segmentation and classification of lung nodules from CT images,” *Journal of Intelligent and Fuzzy Systems*, vol. 42, no. 3, pp. 2667–2679, Jan. 2022, doi: 10.3233/JIFS-212189.

[13] R. Tekade and K. Rajeswari, “Lung Cancer Detection and Classification Using Deep Learning,” Jul. 2018, doi: 10.1109/ICCUBEA.2018.8697352.

[14] B. E. Bejnordi *et al.*, “Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer,” *JAMA - Journal of the American Medical Association*, vol. 318, no. 22, pp. 2199–2210, Dec. 2017, doi: 10.1001/jama.2017.14585.

[15] S. M. Patil, L. Tong, and M. D. Wang, “Generating Region of Interests for Invasive Breast Cancer in Histopathological Whole-Slide-Image,” in *Proceedings - 2020 IEEE 44th Annual Computers, Software, and Applications Conference, COMPSAC 2020*, Jul. 2020, pp. 723–728, doi: 10.1109/COMPSAC48688.2020.0-174.

[16] K. Zormpas-Petridis, R. Noguera, D. K. Ivankovic, I. Roxanis, Y. Jamin, and Y. Yuan, “SuperHistopath: A Deep Learning Pipeline for Mapping Tumor

Heterogeneity on Low-Resolution Whole-Slide Digital Histopathology Images," *Frontiers in Oncology*, vol. 10, p. 3052, Jan. 2021, doi: 10.3389/fonc.2020.586292.

[17] F. Shahidi, "Breast Cancer Histopathology Image Super-Resolution Using Wide-Attention GAN with Improved Wasserstein Gradient Penalty and Perceptual Loss," *IEEE Access*, vol. 9, pp. 32795–32809, 2021, doi: 10.1109/ACCESS.2021.3057497.

[18] H. Tokunaga, Y. Teramoto, A. Yoshizawa, and R. Bise, "Adaptive weighting multi-field-of-view CNN for semantic segmentation in pathology," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2019, vol. 2019-June, pp. 12589–12598, doi: 10.1109/CVPR.2019.01288.

[19] S. Maksoud, K. Zhao, P. Hobson, A. Jennings, and B. C. Lovell, "SOS: Selective objective switch for rapid immunofluorescence whole slide image classification," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2020, pp. 3861–3870, doi: 10.1109/CVPR42600.2020.00392.

[20] N. Dong, M. Kampffmeyer, X. Liang, Z. Wang, W. Dai, and E. Xing, "Reinforced auto-zoom net: Towards accurate and fast breast cancer segmentation in whole-slide images," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018, vol. 11045 LNCS, pp. 317–325, doi: 10.1007/978-3-030-00889-5_36.

[21] J. Hering and J. Kybic, "Multiple Instance Learning Via Deep Hierarchical Exploration for Histology Image Classification," in *Proceedings - International Symposium on Biomedical Imaging*, Apr. 2020, vol. 2020-April, pp. 235–238, doi: 10.1109/ISBI45749.2020.9098616.

[22] A. Paszke *et al.*, "PyTorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems*, 2019, vol. 32.

[23] NIH, "The Cancer Genome Atlas Program - NCI," 2022. https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga (accessed Sep. 25, 2022).

[24] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.

[25] O. Russakovsky *et al.*, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, Dec. 2015, doi: 10.1007/s11263-015-0816-y.