LU-TP 23-07
June 2023

A Machine Learning Approach to Skin Cancer Delineation on Photoacoustic Imaging

Alice Fracchia

Computational Biology and Biological Physics Division, Lund University, Lund, Sweden

FYTM05

Master thesis supervised by Victor Olariu and Emil Andersson

LUND
UNIVERSITY

Abstract

Skin cancer is a growing public health concern due to its prevalence among the population. Current clinical procedures require high invasiveness and multiple surgeries, which are responsible for patient discomfort and high medical expenses. Photoacoustic imaging offers an alternative to standard skin carcinoma diagnosis by exploiting the low scattering rate of ultrasound, which enables deep tissue penetration, and high imaging resolution. This thesis focuses on the application of machine learning models to systematically identify and delineate tumour regions according to their photoacoustic spectra. This is achieved through the implementation of a multilayer perceptron and a convolutional neural network that binary classify the spectral inputs. The same procedure is repeated on data that were dimensionally reduced by an autoencoder. The model predictions are further refined through post-processing contouring techniques. We apply our approach to six samples of the most common skin tumour types and successfully estimate the carcinoma extension. Specifically, the convolutional network accurately estimated the tumour extension, this way consistently decreasing the size of excision margins. This model combined with the contouring technique constitutes a safe approach to skin cancer diagnosis. Our method significantly reduces the number of required surgeries and once automated will decrease the current medical staff workload.

Keywords: Skin cancer, carcinoma, photoacoustic imaging, ultrasound imaging, machine learning, dimensionality reduction, sandpiles algorithm, active contour, multilayer perception (MLP), convolutional neural networks (CNN), autoencoder

Popular science introduction

Skin carcinoma is one of the most common types of cancer and is on the rise among the Caucasian population. When not treated in time, this tumour is fatal. Even though several methods to diagnose skin carcinoma exist, most are limited by invasiveness or poor resolution at increasing depths, which is essential when applied to our bodies. Removal of the suspected tissue and microscopic examination is the most widely used technique. The tumour area is extracted with almost a 1-centimetre margin to ensure that all the carcinogenic regions are removed. This is a lot when the average size of a tumour is around a few millimetres. Unfortunately, sometimes this margin is still not enough, and another excision needs to be made.

Newly emergent imaging methods based on the propagation of high-frequency sound in our body offer a great resolution for skin cancer detection without the need for excision. One of these techniques, called photoacoustic imaging, uses the natural optical properties of some of our skin's components, such as melanin, collagen, blood, or water. These skin's elements emit a high-frequency acoustic signal when excited with light. This emission reveals the concentration of biological components in the sample analysed, just like a fingerprint. This can allow for the classification of healthy and tumorous tissue based on their characteristic composition.

However, due to the multiple tumour types and variations between patients, these techniques can be time-consuming and require human intervention for analysis and assessment. Introducing the use of artificial intelligence (AI) coupled with photoacoustic imaging can reduce the number of medical procedures and offer a less invasive and painful experience to the patient. Each AI model is trained on the singular tumour sample in order to eliminate the variance between different tumour types and patient variations. These models can classify between healthy and tumorous pixels by inspecting their photoacoustic fingerprints and indicating the in-depth extension of the tumour.

In this thesis, we demonstrate the efficacy of combining AI with photoacoustic imaging for skin cancer size prediction. Detecting skin tumours could become more accessible, meaning an earlier diagnosis, which is important for the effectiveness of the treatment and the prognosis of the disease. Such intelligent models could potentially offer unprecedented levels of accuracy in cancer extraction without the need for follow-up visits or the unnecessary removal of skin in case of wrongly diagnosed tumours. Our technique can help the healthcare system to shift its focus to other crucial problems that still need human assessment and intervention.

Acknowledgements

I would like to convey my genuine gratitude to my supervisor, Victor Olariu, whose enthusiasm and guidance have been essential throughout my master's thesis journey. Your approach to problems and ability to integrate multidisciplinary matters have deepened my appreciation for interdisciplinary research. Moreover, your sense of humour and always professional lightheartedness have created a cheerful and inspiring working environment.

I am also extremely grateful to my cosupervisor, Emil Andersson, for his priceless support and unwavering availability. Despite my countless questions, you patiently and positively assisted me every step of the way.

In addition, I would like to extend my sincere appreciation to my fellow office mates, Alexander and Paulina. Your encouragement and positive outlook have been a daily source of inspiration. Our discussions not only broadened my perspectives with respect to this field but also contributed to my personal and athletic growth.

Ultimately, I must thank my family for always pushing me to be the best version of myself. Your belief in my abilities and unconditional support have been a constant source of determination and perseverance. Without you, I would not be here today, and neither would this thesis.

Contents

1. Introduction

Skin carcinoma is the most common type of cancer, with an increasing incidence on a global scale. The dominant types of epidermal tumours are squamous cell carcinoma (SCC), basal cell carcinoma (BCC), and malignant melanoma (MM), the latter being the deadliest one when not treated promptly. Skin cancer is predominant amongst the Caucasian population and the variation between tumour types originates from the cell type subjected to abnormal growth [1–3]. Standard diagnosis and therapy are at present highly invasive and time-consuming. Current clinical methods lack the ability to directly assess the in-depth extension of the tumour. Along with multiple surgeries, these procedures can become a burden on both healthcare expenditure and patient well-being. In addition, late detection can lead to the further spreading of carcinogenic cells and ultimately metastasis [1, 2].

Dermoscopy, also called epiluminescence microscopy (ELM), and histopathological examination are nowadays the golden standard methods to identify skin carcinomas. The first step consists of a visual and sub-superficial assessment through a dermatoscope, a lens equipped with light-emitting linearly polarised diodes [4]. This enables an expert to closely examine the tissue and determine whether to pursue extraction. If needed, the evaluation is followed by an excision biopsy, which is the surgical removal of the affected portion of tissue with a significant safe margin. The specimen is later fixed in formalin and a few sampled vertical sections are examined under the microscope by a pathologist. This analysis is called histopathology and permits confirmation of the complete eradication of the tumour.

When the tissue surrounding the carcinoma is structurally or functionally relevant, the Mohs micrographic surgery (MMS) can instead be adopted. Thin consecutive horizontal layers of skin are removed and histopathologically analysed until the tumour has been entirely extracted. Wound reconstruction is only performed when the margins are assessed to be tumour-free, which might require a separate operation. This technique allows for inspecting the full extension of the margins without relying on the interval of sampled slices like histopathology does. For this reason, MMS yields the highest cure rates for treating skin carcinoma while preserving the largest amount of healthy tissue. Despite its success, this technique remains tedious and time-consuming. Besides representing a high cost for the healthcare system, the delay in wound reconstruction can easily increase patient discomfort during the procedure [5].

Research on numerous optical techniques is currently conducted to non-invasively examine the tissue prior to excision. Some of these methods include confocal and multiphoton microscopy, optical coherent tomography, Raman, fluorescence, and diffuse reflectance spectroscopy. However, these methods are limited by optical scattering and only allow for shallow penetration, with a maximum of 1 mm for optical coherence tomography [6]. Photoacoustic (PA) and ultrasound (US) techniques recently gained interest as imaging techniques with submillimetre resolution up to several centimetres of depth into the tissue [7]. The signal can be used to characterise tissue types according to their wavelength-dependent optical properties. Photon propagation and energy absorption into the medium can give information about the presence and concentration of light-absorbing molecules, also called chromophores [8]. However, due to the complexity and heterogeneity of the tissue structure, an analytical investigation of light propagation is challenging. Computational methods, for instance, Monte Carlo simulations, can be

applied to describe the light distribution through the multi-layered skin [9]. Since carcinogenic cells modify the optical properties of skin and their location is unknown prior to surgery, numerical methods can only offer an approximation of the system.

Alternatively, the concentration of chromophores can be determined from the PA images by spectral unmixing. This method allows for the decomposition of the spectral signal into a linear combination of endmembers spectra, which correspond to the different pure spectra of the chromophores present in the sample [10]. Thus, this approach requires knowledge of the dominant light-absorbing components in the sample. Due to the structural and chemical differences between tumour types and their inter-class variations, this analysis can become laborious and quite limited.

Recent research focuses on the application of deep neural network classifiers for tumour segmentation. However, these studies are mainly carried out on standard, dermoscopy, or hyperspectral two-dimensional images of skin lesions [11–16]. Common networks operationalised for classification usually include convolutional neural networks (CNN), pre-trained deeper CNNs or residual neural networks [11–14]. Although this research effort shows the consistent success of the predictions, only a few studies have investigated the application of machine learning to photoacoustic imaging on skin cancer. State-of-the-art research mainly includes studies on breast, prostate, and thyroid tumours [17–20]. The use of photoacoustic data allows for an in-depth analysis of the samples, enabling the examination of the full tumour margins before surgery.

In this project, we propose the use of three common artificial neural networks to analyse the photoacoustic spectra of ex-vivo skin samples with the aim of accurately predicting the full tumour extension prior to surgery. A multilayer perceptron (MLP) and a CNN are trained and tested on each sample for the binary classification of individual pixels. Principal component analysis and an autoencoder are used for the dimensionality reduction of the data to try to improve the classifiers' predictions. This study demonstrates the effectiveness of employing this approach by successfully estimating the carcinoma's in-depth extension. Our models are applied to ex-vivo samples. However, they can also be applied to in-vivo imaging obtained by a motion tracking system with detailed high-resolution [10].

This approach brings forth several advantages. Firstly, the solution proposed addresses the need for personalised medicine. By training the models and making predictions individually on each sample, it is possible to tailor the diagnosis to the single patient, eliminating any patient-patient bias and variation due to carcinoma and skin type [21]. Additionally, the application of this method would reduce the invasiveness of standard clinical procedures, thus improving the overall patient experience and comfort. By being technically, statistically, and conceptually reproducible [22], this project also allows for precise and unique duplicability, regardless of sample type and acquisition. Finally, by implementing a systematic and automated approach, the potential reduction in medical procedures could alleviate the burden on healthcare professionals, allowing them to allocate their attention to other critical areas.

This thesis explores the theory behind the generation of data through photoacoustic and ultrasound imaging, while also providing an overview of the three proposed neural network models. The methods section focuses on elucidating the pre-processing techniques employed to prepare the input to the networks. Furthermore, it details the operational aspects of these models in predicting the tumour's

extension. Additionally, we introduce the active contour algorithm used to delineate and finalise the predictions. Finally, we apply our framework to six different samples containing examples of malignant melanoma and basal cell carcinoma. The accuracy of the results is then assessed by comparing them to the histopathological examinations.

2. Theory

In this section, we provide the essential background knowledge required to follow the steps of this project. Initially, the working mechanism of photoacoustic and ultrasound imaging is illustrated, followed by a comparison in resolution and penetration depth with respect to other optical-based methods. Subsequently, we present the fundamental concepts of the multilayer perceptron, convolutional neural networks, and the autoencoder, emphasising their relevance and application within this study. The implementation of active contouring, which will be used to delineate the tumourous region, is finally explained.

2.1. Photoacoustic (PA) and Ultrasound (US) Imaging

Photoacoustic imaging is an absorption-based non-invasive and non-ionising technique that exploits the property of endogenous molecules present in our body to generate an optically induced ultrasonic signal when excited with short light pulses. The originating signal is thus a high-frequency acoustic wave called *ultrasound* wave. The different concentrations and chemical compositions of such biological molecules, also called chromophores, create the contrast in the image. Haemoglobin, melanin, water, and lipid are examples of anatomical and functional contrast agents present in our body [7, 23].

A wide range of wavelengths can be used to excite the sample and obtain information from different components with dominating absorption in different intervals of the spectra. As a laser pulse is delivered to the tissue, the energy absorption causes a temperature rise in light-absorbing molecules. As depicted in Figure 1, the tissue undergoes local volume expansions due to the thermally induced mechanical vibrations, which are proportional to the light-absorption coefficient of the components. This generates an acoustic pressure, which is proportional to multiple variables: the thermal expansion coefficient and isothermal compressibility of the tissue, the efficiency in the conversion from photon energy to thermal energy, and the conversion from thermal energy to mechanical, thus acoustic, energy. This pressure propagates in the sample as an acoustic wave which encodes spatial information about the tissue absorbers. The acoustic waves are collected by a transducer and the signal is processed to obtain clear images [23].

Photoacoustic imaging overcomes the high degree of light scattering that affects other ballistic and diffuse optical methods. Since ultrasound waves scatter roughly 1000 times less compared to photons, this technique allows for the maintenance of a high depth-to-resolution ratio. PA imaging can achieve a spatial resolution corresponding to 1/200 of the desired image depth, compared to 1/3 for standard diffuse optical techniques. For the same resolution, the signal depth can reach up to 7 cm underneath the tissue surface, whereas the optical diffusion limit bounds optical techniques to a 1

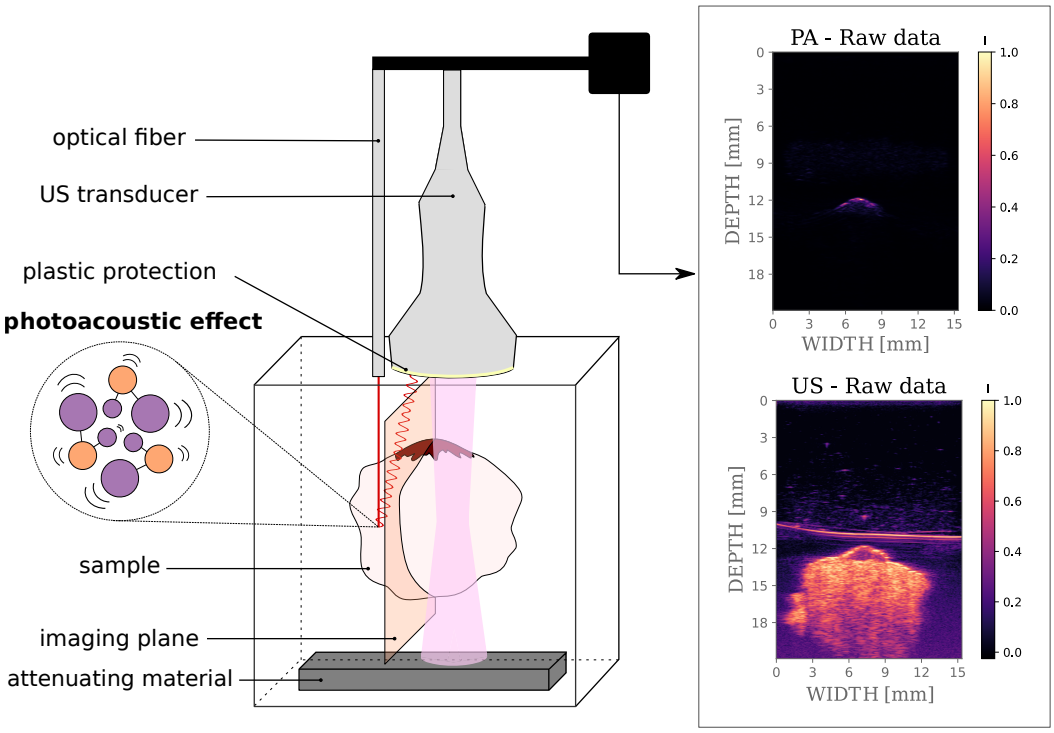

Figure 1: **Set-up for the acquisition of PA and US images:** the sample is placed in a container filled with an isotonic solution with an ultrasound attenuating material at the bottom. A layer of ultrasonic coupling gel is added between the transducer and the sample, together with a protective plastic film. The specimen is then irradiated alternatively with a laser pulse and diagnostic ultrasound. The sample's molecules undergo thermal expansions as they are excited by the laser. This originates the PA signal (right upper corner) which is collected by the transducer. The US signal (right lower corner) is instead measured by collecting the ultrasound waves that were reflected at tissue interfaces.

mm maximum penetration. Sensor quality and shape also define image quality [7]. By making use of the optical property of endogenous particles, this imaging technique has the additional advantage of being label-free. This means no external endogenous contrast or stain needs to be administered to the patient, reducing the cost and invasiveness of the procedure.

Ultrasound imaging is a similarly non-invasive and non-ionising technique that takes advantage of the propagation properties of high-frequency sound waves to reconstruct information about the inner structure of the tissue. US imaging can be executed both with a continuous or pulsed wave, the latter being the most used in clinical settings. In this case, the same transducer can be used to both produce and detect the echoes. Instead of a laser, the sample is excited with very high-frequency sound pulses. These are commonly obtained by exploiting the physical properties of piezoelectric materials: when the crystal is compressed or stretched, a charge appears at its surface. On the other hand, applying a potential causes expansions or contractions of the crystal. As previously mentioned, molecular expansion produces ultrasound waves. Thanks to its double nature, a piezoelectric crystal can be used both for the generation and detection of ultrasound waves.

The interaction of sound waves with biological matter results in different phenomena: reflection, refraction, absorption, and scattering. The nature of the interaction depends on the physical properties

of the molecules, such as density, viscosity, and elasticity. The energy is mainly reflected at the interface between different tissue types and thus carries structural information about the heterogeneity of a sample. From the outgoing signal, the distance of each reflector can be calculated. By combining information from different scan lines, two-dimensional images of the sample's cross-sections can be reconstructed and displayed. Gases can act as a barrier to US propagation due to their lower density. For this reason, ultrasonic coupling gels are generally introduced between the imaging device and the sample to remove the air interface. As the incident US beam propagates in-depth, it diverges and attenuates, affecting the image resolution. Technical adjustments can be taken to limit the effects of beam attenuation on image quality. Although ultrasound imaging offers deeper penetration due to the lower scattering nature of sound, it can only achieve low resolution compared to photoacoustic imaging due to the beam attenuation [24, 25].

When combined, these techniques offer both a structural and spectral characterisation of the sample with high resolution and deep penetration. From the resulting images, it is possible to extract the absorption spectra across the wavelengths used for sample excitation. This can be done on the individual pixels to obtain a spectral signature characteristic of the molecular composition of a specific region. Since the vibrational frequencies are very sensitive to the chemical structure of the compound, the contrast in our data is enhanced by the higher optical absorption of carcinogenic cells. This is caused by the abnormal concentration of chromophores such as melanin, haemoglobin, and collagen in the tumour tissue. Photoacoustic imaging is thus a successful technique to morphologically, functionally, and molecularly distinguish biological tissue in-depth [7].

2.2. Artificial Neural Networks (ANNs)

ANNs are computational models inspired by the nervous system of living beings. The structure of a network is composed of units called nodes and synaptic weights. The external stimuli are given by the training data, which serve as input to the neural networks. The information contained in each input node is then scaled with respect to the respective weight and propagated to the neurons in the next layer. The activation of neurons is simulated with an activation function, which corresponds to the operation the model applies to the received input determining the output. Usual activation functions can be linear, rectified linear unit (ReLU), hyperbolic tangent (tanh), and sigmoidal. When the activation function is chosen to be linear, the outgoing signal is thus a weighted sum of the inputs. A bias, or constant, is introduced at each layer, except for the output one.

Models can be classified as supervised or unsupervised according to their need for labels associated with each input point. Labels contain information about the desired output: in classification tasks, for instance, the label corresponds to the class a training data belongs to. Multilayer perceptrons and convolutional neural networks are examples of supervised models that require prior knowledge of the outcome for a set of data points to train the network. Autoencoders are instead an example of unsupervised models: since the desired output corresponds to the input itself, the usage of labels is not required. The model learns by comparison of the computed output to the desired output with respect to a preferred error function. The network weights are then updated so that the error function can be

minimised. Examples of standard errors between output and target are the mean square error (MSE) or the cross-entropy error (CEE).

The steps applied to adjust the weights and thresholds of each node make up the learning algorithm. A widely used algorithm in multilayer models is *back-propagation*, which allows computing the error and adjusting the weight of each node accordingly from output to input, hence the name. Different optimisation algorithms can be used to minimise the error function and are associated with a series of hyperparameters. Tuning of hyperparameters and optimisation method enables to reach efficiently the optimal weight values to solve the task. When the output discrepancy is within an acceptable range, the model is considered trained and is ready to perform predictions on new data.

If the model lacks generalisation and misses the underlying trend of the data, the network is overtrained. When this happens, the results yield high accuracy on the training data, but the network is unable to make good predictions on unseen data. Overtraining can be avoided by model selection during the validation stage, which involves applying the network to unseen inputs and measuring its performance. Both training and validation errors thus need to be minimised when training a neural network [26, 27]. The selected model can finally be applied to a test set to make predictions.

This project focuses on the application of three feed-forward neural network models to tackle two different problems: binary classification of individual pixels and dimensionality reduction of the data.

2.2.1. *The Multilayer Perceptron (MLP)*

An MLP is a simple neural network used to treat one-dimensional data. It is composed of an input layer, one or more hidden layers, and a final output layer. The input layer presents the same number of nodes as the components of each input data point. As stated earlier, the operations performed by each layer are dictated by the choice of the activation function. When all operations are linear, the network executes simple linear regression. Non-linear activation functions make up the power of multilayer structures and keep the output bound to the function's specific boundaries [27]. An example of MLP architecture is depicted in Fig. 2.

MLPs are employed in a large variety of tasks, such as function approximation, process identification and control, system optimisation, time series forecasting, and so on. Such networks are also largely employed as superior classifiers, due to their ability to perform non-linear regression tasks [26]. In the case of a binary classifier, the network usually presents one output node with a sigmoidal activation function, which returns values in the range between 0 and 1 and can be interpreted as the probability of belonging to one of the two classes.

2.2.2. *Convolutional Neural Networks (CNN)*

Convolutional neural networks can assume different architectures according to the dimension of the input. Generally, CNNs are mainly used on two-dimensional inputs, but one-dimensional correspondents can be easily implemented. Figure 3 shows a general structure of a one-dimensional CNN. The name of the network comes from its building units which are the convolutional layers. Each convolutional layer iteratively applies a function, also known as the kernel, to each possible position of the input until the

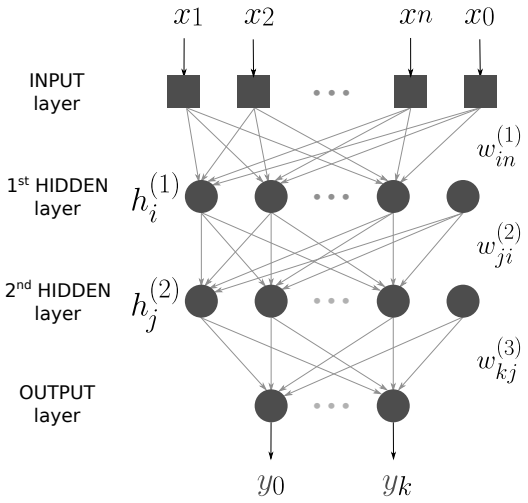

Figure 2: **Architecture of a multilayer perceptron with two hidden layers:** the input is the vector $\mathbf{x} = (x_1, .., x_n)$, with x_0 being the input bias. The number of weights w between each layer corresponds to the product of nodes in the previous layer plus one, times the number of nodes in the following layer. The nodes in the hidden layers are indicated by h. The output is the vector $\mathbf{y} = (y_0, .., y_k)$.

kernel has overlapped with all of its components. Each kernel identifies a particular spatial feature and consecutive layers focus on gradually larger feature detection. CNN layers are sparsely connected, which means that the number of weights associated with a layer is equal to the number of kernels times their size. The weights are characteristic of the kernel and their value stays the same when applied to different components of the input. The structure of the successive layer is given by the size of the dot product between the kernel and the output of the previous layer. The final output of the convolutional layers is then flattened and connected to a fully connected MLP.

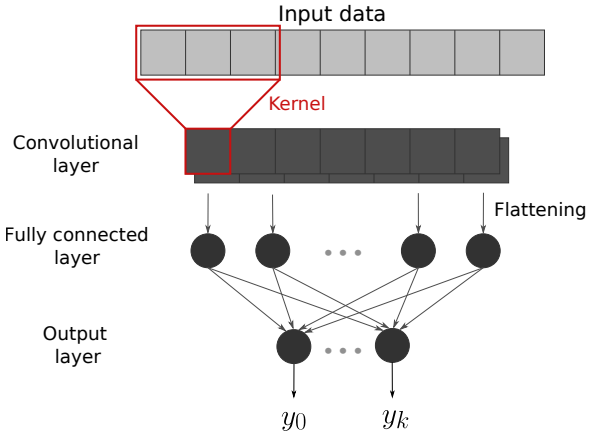

Figure 3: **Example of a one-dimensional convolutional neural network:** each convolutional layer is created by the application of a kernel to the input data. This, represented in red, is consecutively applied until it has overlapped with all its components. CNNs can be composed of multiple consecutive convolutional layers. The last convolutional layer is then flattened and passed to the dense fully connected structure to proceed with computations. The rest of the networks act as an MLP and gives output $\mathbf{y} = (y_0, .., y_k)$.

CNNs are usually employed in image comprehension, due to their ability to retain spatial information between neighbouring data. However, this architecture can be also useful when applied to one-dimensional data where some information is embedded in the neighbouring components of the data [27]. As for the MLP, CNNs which are used as binary classifiers present one final node whose value represents class identity probability.

2.2.3. *The Autoencoder*

When working with large-size datasets, the number of variables for each point can exceed the number of observations. When operating in this regime, it is hard to both visualise and analyse the information. This phenomenon represents what is called the *curse of dimensionality* in machine learning [28]. Although our data dimension, 59 as the number of wavelengths, is much lower than the number of observations, the number of pixels per sample, dimensionality reduction can still be applied to the data to aim towards the improvement of the predictions. There are numerous algorithms that can be applied to perform this task, such as single value decomposition (SVD), principal component analysis (PCA), and using an autoencoder. SVD and PCA are unsupervised algorithms that learn an orthogonal linear transformation to represent the data into a lower dimension with higher variance. PCA is a special case of SVD, where only the components with the highest variance are considered.

The autoencoder is a feed-forward fully connected model trained to reproduce the input by compressing the data into a lower dimension and later decompressing it to its original shape. A good reconstruction of the data implies a successful compression which still holds the most important information about the input. A general architecture for an autoencoder is portrayed in Fig. 4.

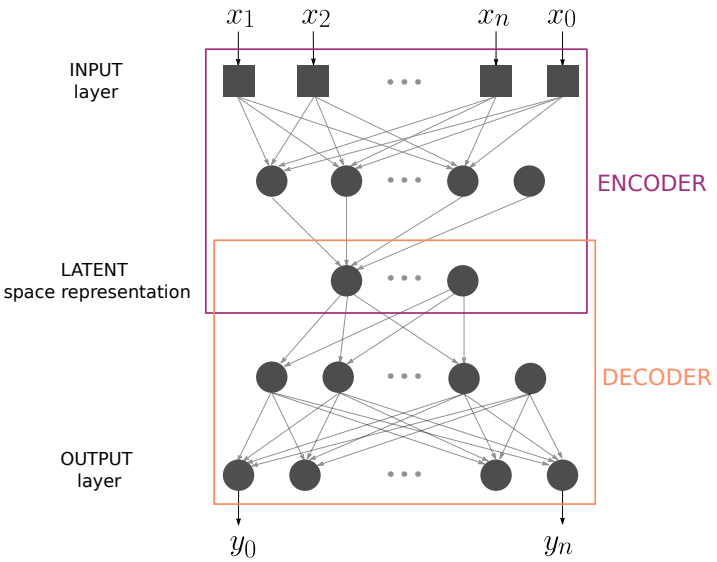

Figure 4: **Architecture of an autoencoder:** the input $\mathbf{x} = (x_1, .., x_n)$ is compressed by the encoder's consecutive fully connected layers into its latent representation. This corresponds to the bottleneck of the network. The data is then decompressed by the decoder network. The output vector $\mathbf{y} = (y_0, .., y_k)$ is a reconstruction of the input.

In correspondence with the central bottleneck, we define the latent representation. The output of this innermost layer corresponds to a reduced representation of the data. The group of compressing layers leading to dimensionality reduction is called the *encoder*, while the remaining architecture takes the name of *decoder*. A symmetric structure is usual, but not necessary. When a linear autoencoder is used to minimise the mean square error between the input and the output to train, it mathematically corresponds to computing singular value decomposition, which optimizes the same loss function. However, contrary to SVD or PCA, the autoencoder solution does not need to satisfy any orthogonality condition, allowing for alternative optima [27].

2.3. Tumour Delineation through Active Contouring

Active contouring refers to the segmentation technique based on energy minimization of a continuous deformable contour used to detect salient features in two-dimensional images. Its shape and movement are guided by internal and external forces to fall into the locally optimal solution, generally corresponding to contours and edges.

Let $\mathbf{v}$ be a vector containing the coordinates $((x(s), y(s))$ of the point s belonging to the contour. The total energy E_{tot} of the contour can be defined as:

$$E_{\text{tot}} = \int_0^1 E_{\text{pt}}\left(\mathbf{v}(s)\right) \mathrm{d}s = \int_0^1 \Big(E_{\text{int}}\left(\mathbf{v}(s)\right) + E_{\text{ext}}\left(\mathbf{v}(s)\right) + E_{\text{con}}\left(\mathbf{v}(s)\right) \Big) \mathrm{d}s. \tag{2.1}$$

The total energy is calculated as the sum of the individual points' energy E_{pt}. This is given by the sum of three characteristic types of forces. Internal forces E_{int}, for instance, stiffness or gravitational pull, are responsible for the elastic and contracting behaviour of the contour. External forces E_{ext}, which are specific to the considered image, affect the energy potential guiding the contour towards salient features. Finally, constraint forces E_{con} can be arbitrarily added to further guide the movement of the contour [29]. Equations (2.2)-(2.4) refer to an example of common forces implemented to guide the movement of the contour through the energy landscape [16]:

$$E_{\text{int}} = E_{\text{stiff}} + E_{\text{grav}} = \alpha(s)|\mathbf{v}_{ss}(s)|^2 + \beta|\mathbf{v}(s) - \bar{\mathbf{v}}(s)|^2 \tag{2.2}$$

$$E_{\text{ext}} = E_{\text{img}} \tag{2.3}$$

$$E_{\text{con}} = \gamma(\overline{\mathrm{d}} - |\mathbf{v}_s(s)|). \tag{2.4}$$

In Eq. (2.2), the energy terms respectively describe the stiffness of the contour and the gravitational pull originating from its centre of mass $\bar{\mathbf{v}}$. $\alpha(s)$, β, and γ are tunable parameters which can be dependent on the point of the contour considered. The external forces are represented by the energy potential E_{img} of the image itself: this can correspond to the pixels' intensity or more generally to a function of the feature one is aiming to highlight. Finally, Eq. (2.4) is introduced to ensure homogeneous distribution of the contour's points, with $\overline{\mathrm{d}}$ being the average distance between two consecutive points.

When working in a discrete regime, the coordinates of the ith points can be expressed as $\mathbf{v}_i = (x_i, y_i) = (x(ih), y(ih))$, with h indicating the step size between consecutive coordinates. In our case, h

is set to one and Eqs.(2.1)-(2.4) approximate to:

$$E_{\text{tot}} = \sum_{i=1}^{n} E_{\text{pt}}(i) = \sum_{i=1}^{n} \Big(E_{\text{int}}(i) + E_{\text{ext}}(i) + E_{\text{con}}(i) \Big) \tag{2.5}$$

$$\begin{aligned} E_{\text{int}}(i) &= E_{\text{stiff}}(i) + E_{\text{grav}}(i) = \\ &= \alpha(s)|\mathbf{v}_{i-1} - 2\mathbf{v}_i + \mathbf{v}_{i+1}|^2 + \beta|\bar{\mathbf{v}} - \mathbf{v}|^2 \end{aligned} \tag{2.6}$$

$$E_{\text{ext}}(i) = E_{\text{img}}(i) \tag{2.7}$$

$$E_{\text{con}}(i) = \gamma(\overline{\text{d}} - |\mathbf{v}_{i-1} - \mathbf{v}_i|). \tag{2.8}$$

Numerically, energy minimization can be solved from the discrete approximation with the Euler method [29]. Computationally, energy minimization corresponds to iteratively determining which of the nearest neighbouring sites for each point of the contour provides the lowest energy contribution. The contour is here implemented in two dimensions, but can potentially be adapted to three dimensions by considering $\mathbf{v}_i = (x_i, y_i, z_i)$.

3. Methods

In this section, we go through the methodology employed to obtain the results of this project. Sample preparation and data acquisition are explained to understand the choice of pre-processing techniques and labelling. Moreover, the architecture of the three artificial neural networks used is described together with the choice of hyperparameters and activation functions. Finally, we illustrate an algorithm to define the image energy landscape for active contouring.

3.1. Sample preparation and experimental set-up

The 6 tumour samples were excised by surgeons at the Department of Ophthalmology at the Skåne University Hospital. Each specimen was shaved and immersed in a Plexiglas repository containing isotonic saline solution. To achieve an adequate distance between the probe and the skin surface, as well as avoid the air-tissue interface, ultrasound gel padding was added and plastic film was wrapped around the transducer. A black ultrasound-attenuating material was placed at the bottom of the container to avoid noise due to the reflection of the exciting signal on the support. A laser source of spectrally tunable nanosecond pulses and a linear ultra-high-frequency transducer were alternatively used to image the sample, achieving an axial and lateral resolution of respectively 50 and 110 μm. The exciting laser pulse was guided in parallel to the detector through optical fibres aimed perpendicularly at the skin surface. 59 wavelengths ranging from 680 to 970 nm, with 5 nm intervals, were used to excite the sample. The diagnostic ultrasound used 40 MHz pulses. The transducer, composed of a linear array of 256 piezoelectric crystals, collected the ultrasound waves deriving from photoacoustic effect and reflection. The device was mounted on a linear stepper motor to sequentially image the sample and enable three-dimensional reconstruction from 47 slices each 0.5 mm apart. The data collected are in arbitrary units, to allow for a relative comparison of intensity between wavelength channels.

Potential motion artefacts were removed by attaching the stepper motor to an adjustable arm. The sample was later fixed in formalin and subjected to histopathological examination. The raw data were finally exported to Matlab and converted to Python arrays to proceed with the analysis.

3.2. Data pre-processing

Photoacoustic imaging results contain the intensity for 712 × 512 pixels, or 432 × 512 depending on the sample, for 47 vertical slices across the 59 imaging wavelengths. For visualization purposes, the results shown depict one vertical cross-section for each sample, for an arbitrarily chosen and specified channel. The chosen slice approximately corresponds to the slice examined by the pathologist. PA and US signals were normalized with respect to the highest intensity in the full sample for each wavelength channel. The structural information contained in the US images was exploited to segment the background from the sample. Contrary to the PA images, in which only the areas closest to the tumour have significantly higher intensity, in the ultrasound images the whole sample creates contrast (Fig. 5 **(a)**). By setting up an intensity threshold, it was possible to eliminate the background signal. The mask was then cleaned from the highly intense signal deriving from the plastic film as well as any dirt in the saline solution. Computationally, this was done by isolating for each slice the spatially largest group of points in the mask and setting any other intensity to zero. The cleaned mask was then applied to the PA data to obtain a clear signal of the sample for each slice and each wavelength channel. Figure 5 depicts the pre-processing techniques applied to sample 153 and the final data ready for labelling.

3.3. Histopathology and labelling

As stated in section 2.2., supervised models need labelled points to train. In the ex-vivo case, this information is contained in the histopathological cross-sections. These are examined by doctors assessing the extension of the tumour, which is done by identifying the anomalous cell growth with a microscope. The tumour sizes determined by histopathological examination are reported in Table 1. Complete labelling of the image by matching the two would be challenging, due to the deformation the sample undergoes when treated for histopathologic assessment. The examined cross-section might additionally not correspond to the vertical direction of the imaging system. Furthermore, multiple histopathology assessments were available for the same sample with no indication of the cross-section position with respect to the sample. However, a visual comparison with the photoacoustic images suggests that high-intensity pixels usually correspond to the area highlighted as tumorous. The idea is then to find a systematic way to group high-intensity pixels. For this, various clustering techniques were explored. As shown in Fig. 6, k-means clustering gives results very similar in size and shape to the carcinogenic region. The defined cluster was chosen as the area where to randomly pick pixels to label their spectra as tumorous. Since the second cluster usually included both sample and background, due to their similar low intensity, the healthy cluster was selected as the sample signal not belonging to the tumour cluster. However, at the interface between the two classes of points, a border of around 0.5 mm was left out to avoid mislabeling due to imprecision in comparison between original data and histopathological

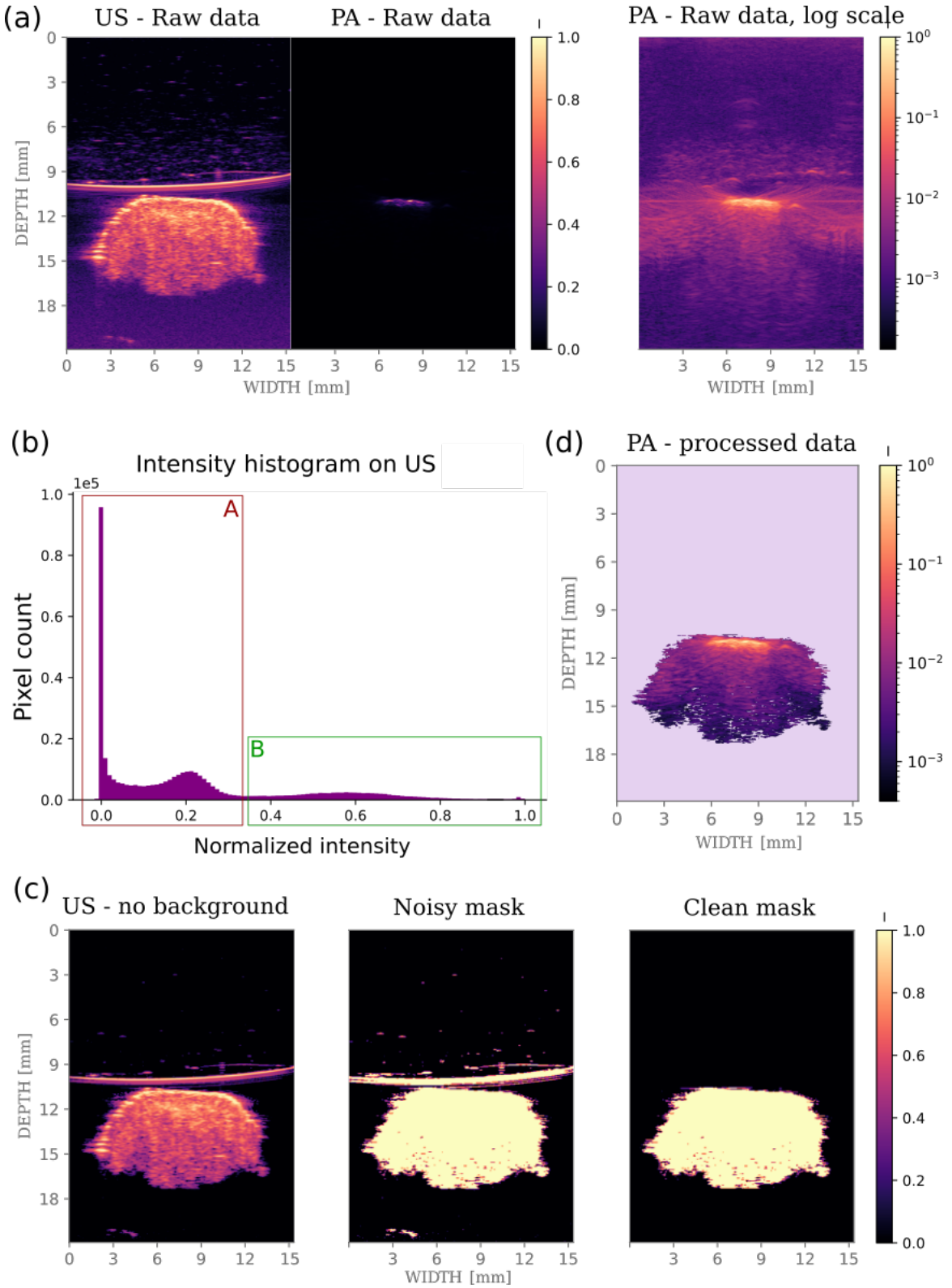

Figure 5: **Data pre-processing:** **(a)** Raw data: ultrasound and photoacoustic data with normalized intensity and a.u. for wavelength 680 nm. On the right, photoacoustic data with logarithmic colorscale for improved visualization; **(b)** Intensity histogram on ultrasound imaging, with **A** indicating the intensities belonging to background pixels and **B** the sample pixels; the threshold is set at the minimal value between these two groups; **(c)** on the left, the ultrasound image with background intensities set to zero; in the centre, the boolean mask with plastic and dirt noise; on the right the final mask corresponding to the sample's extension; **(d)** the final pre-processed PA data in logarithmic scale.

images (Fig. 6). From each sample, 2000 points were randomly chosen evenly between tumour and healthy clusters from each slice. If the number of pixels belonging to the tumour was less than 1000, the number of healthy pixels included was also reduced appropriately.

Table 1: **Histopathology examination outcome:** tumour extension according to microscopic examination. For samples 128 and 153, two slices were provided. In this case, the model's predictions were compared to an average of the two sizes. (*) Histopathological examinations of samples 389 and 394 were not available at the time of the study, but a preliminary assessment was provided.

sample number:	**128**	**131**	**153**	**389**	**394**	**266**
tumour type:	MM	MM	MM	BCC*	Dysplastic Nevus*	BCC
number of slices:	2	1	2	missing	missing	1
width [mm]:	2.43/2.96	12.6	4.23/4.93	-	-	5.20
depth [mm]:	0.76/0.51	1.00	0.60/0.60	-	-	2.84

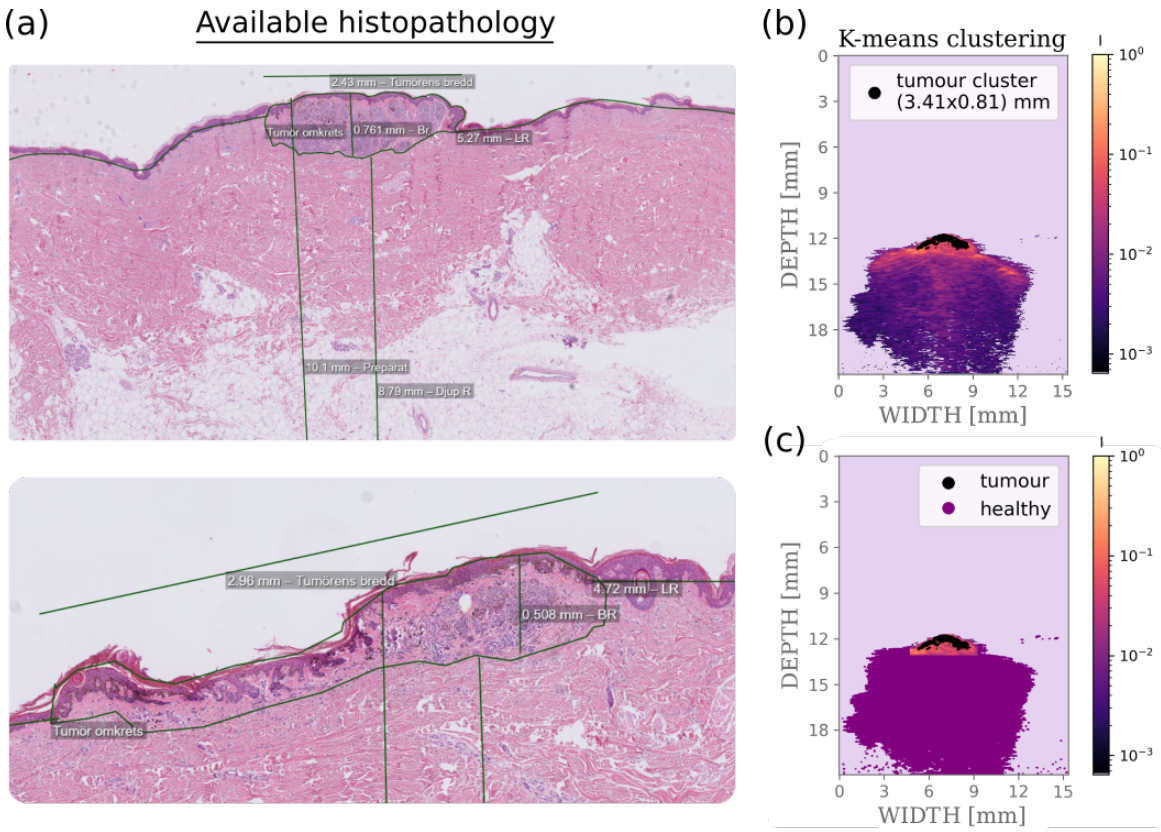

Figure 6: **Labeling strategy: (a)** Histopathology of two cross-sections from sample 128 with assessed tumour size. Anomalous cell growth can be identified in the delineated border. The tumour delineations seem to approximatively match the cluster obtained. **(b)** K-means clustering results on sample 128 for wavelength 680 nm with cluster size in mm. **(c)** In black is depicted the tumour cluster, from which 1000 random pixels are chosen as training data. In purple are the remaining healthy points, from which the other 1000 training points were selected. A safe boundary of approximately 0.5 mm is left unlabeled between the two groups due to imprecision in matching data to histopathological results.

3.4. Machine Learning models

The idea is to use three different models with two different approaches on the photoacoustic data. A multilayer perceptron and a one-dimensional convolutional neural network were used as binary classifiers to distinguish between healthy and tumorous pixels. PCA and a non-linear AutoEncoder were used to reduce the dimensionality of the data and create a new input for classification. For the supervised classifiers, the input data consisted of the labelled spectra of a sample split with a proportion of 80-20 % in respectively training and validation sets. For the autoencoder, all the sample's pixels were split into training and validation datasets with the same proportion. After training, all models were tested on the spectra of all pixels belonging to the same sample.

3.4.1. *MLP classifier*

A simple MLP was used on both pre-processed and dimensionally reduced data. The input to the model was thus a list of respectively 59 and 2 or 10-dimensional spectra of pixels selected from the sample, according to the criteria described in section 3.3.. The architecture was one of a binary classifier with one output node. The input layer included as many nodes as the dimension of the input. The number of nodes in the single hidden layer was determined by a grid search exploring seven geometries with different combinations of learning rate, mini-batch size, and the number of epochs (Table 2 in Appendix). Model validation was done by visually assessing the performance of the network in comparison to the histopathological examination. Figure 7 shows the final architecture and hyperparameter selection of the classifier. As mentioned in section 2.2.1., the output of the MLP can be interpreted as the probability of belonging to one of the two possible classes: tumorous or healthy pixel. Since the proportion of training points is evenly balanced, the threshold for class attribution was set to 0.5. Each pixel was then mapped back and visualised on the sample to determine the success of the predictions in comparison to the histopathological results.

3.4.2. *CNN classifier*

Conventional convolutional networks allow the inclusion of spatial information in the training data so that the classification can take into account the nature of neighbouring pixels. However, due to the imprecision in matching histopathological images to our data, our images could not be used in their entirety due to the lack of labels defining the tumour region pixel by pixel. For this reason, a two-dimensional CNN which trains on the fully labelled image can not be implemented. Instead, a simple 1-dimensional CNN can be used on the pixels' spectra to introduce information about the neighbouring wavelength channels. The input to the network was both the pre-processed and dimensionally reduced data. The structure of the network was arbitrarily chosen to be as simple as possible while still performing accurately. The output layer consists of a single node to perform binary classification. Trial and error for hyperparameters tuning showed that this is less determinant compared to the network's architecture, kernel number and size. Figure 7 reports the final structure of the convolutional neural network and hyperparameter choice used to produce the results.

3.4.3. *Autoencoder for dimensionality reduction*

As mentioned in section 2.2.3., this model is unsupervised, since the desired output is a precise reconstruction of the input itself. Labelling was thus not necessary and all pixels can be included in the networks' training. The autoencoder architecture was investigated in order to minimise the reconstruction discrepancy of the data. Particularly, three different activation functions were tested and the resulting training histories for one of the samples are reported in Fig. 8. A few trial and error attempts showed that hyperparameter tuning is less important than the model's architecture. As expected, more latent nodes permit the storage of more information and decrease the reconstruction error of the network. Also, we need to keep in mind that the components found by the autoencoder are not necessarily orthonormal to each other, thus any representation without knowledge of the axis system would reflect incorrect information. Figure 7 shows the final architecture used to produce the results.

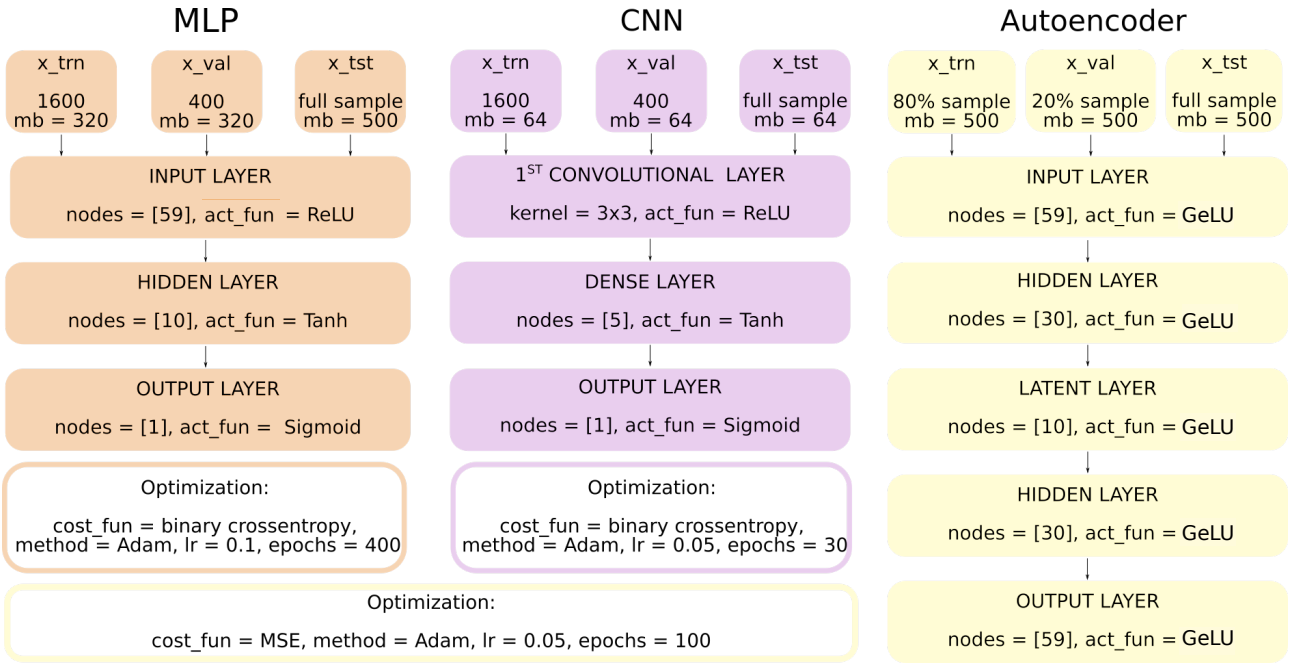

Figure 7: **Network implementations:** block diagrams of the neural networks used containing information about input data, architecture, activation functions, hyperparameters, and optimization methods. *x_trn*, *x_val* and *x_tst* are respectively the data points for training, validation and testing. *mb* stands for the minibatch size, *lr* for learning rate, *act_fun* and *cost_fun* for activation and cost function respectively. *Adam* stands for Adaptive Moment Estimation and is the optimization algorithm chosen for weight updates.

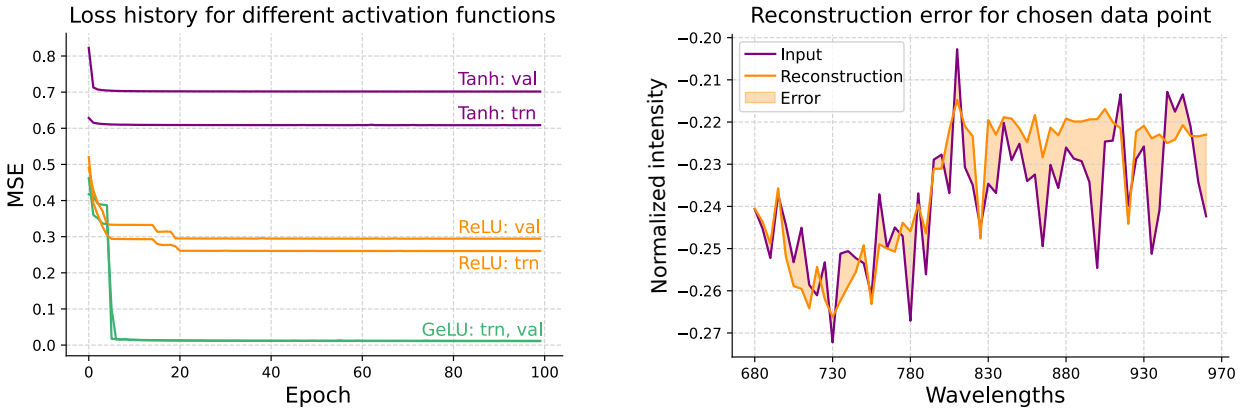

Figure 8: **Autoencoder selection:** on the left, the training history of the autoencoder when implementing three different activation functions: Tanh, ReLU, and GeLU (Gaussian ReLU). On the right, the reconstruction error of the autoencoder: the input, or target output, is represented in purple, while the output of the autoencoder is shown in orange. The orange filling depicts the reconstruction error.

3.5. Image Energy for Active Contouring

We propose the sandpiles algorithm to construct an energy landscape which serves as input to the active contour algorithm, which is finally used to refine the final predictions. The procedure consists of iteratively building figurative sand piles on top of each pixel according to their probability of belonging to the tumour class until the achievement of a steady state. This system implements neighbouring exchanges while healthy pixels act as draining sinks. This technique permits polishing of the predictions from small isolated clusters of either carcinogenic or healthy pixels while building an energy landscape for active contouring [16].

For an image of n pixels, an n-dimensional vector was initialized with the relative probabilities of being tumourous $\mathbf{t}$ and the amount of sand $\mathbf{s}$ on top of each pixel:

$$\mathbf{t} = (t_1, t_2, ..., t_n), \tag{3.9}$$

$$\mathbf{s} = (s_1, s_2, ..., s_n). \tag{3.10}$$

A connectivity matrix C was constructed to represent connections between neighbour pixels to model the sand exchange. The constant w can assume two possible values: one if two pixels are adjacent and zero otherwise:

$$C = \begin{pmatrix} 1 & w_{12} & \cdots & w_{1n} \\ w_{21} & 1 & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & 1 \end{pmatrix}. \tag{3.11}$$

An update matrix U for each iteration was then defined considering sand intake from the nearest neighbour tumourous pixels and a constant sand output proportional to b to the 4 nearest sites. In Eq.

(3.12), b was arbitrarily set to 1/8:

$$\mathbf{U} = C - 4b\mathbb{1}. \tag{3.12}$$

To include the addition of sand at each iteration according to the probability t, the update matrix U was written as an augmented matrix M, where the first column represents the sand increment independent of connections and the first empty row makes it independent of pre-existing sand as well:

$$M = \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ t & & U & \end{array}\right) = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ t_1 & 1-4b & w_{12} & \cdots & w_{1n} \\ t_2 & w_{12} & 1-4b & \cdots & w_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_n & w_{n1} & w_{n2} & \cdots & 1-4b \end{pmatrix}. \tag{3.13}$$

We finally defined the vector $\mathbf{v}$ as the vector $\mathbf{s}$ augmented of one row containing 1. The update matrix M can be then applied to the vector $\mathbf{v}$ N times until reaching a steady state:

$$\mathbf{v} = \begin{pmatrix} 1 \\ \mathbf{s} \end{pmatrix} \tag{3.14}$$

$$\mathbf{v}^{(0)} = (1, 0, \cdots, 0)^T \tag{3.15}$$

$$\mathbf{v}^{(1)} = \mathbf{M}\mathbf{v}^{(0)} \tag{3.16}$$

$$\mathbf{v}^{(2)} = \mathbf{M}\mathbf{v}^{(1)} = \mathbf{M}(\mathbf{M}\mathbf{v}^{(0)}) = \mathbf{M}^2\mathbf{v}^{(0)} \tag{3.17}$$

$$\vdots \tag{3.18}$$

$$\mathbf{v}^{(N)} = \mathbf{M}^N\mathbf{v}^{(0)}. \tag{3.19}$$

The vector representing the amount of sand at steady-state is the vector $\mathbf{v}^{(N)}$ that satisfies $\mathbf{v}^{(N)} = \mathbf{M}\mathbf{v}^{(N)} = 1 \cdot \mathbf{v}^{(N)}$. This corresponds to the eigenvectors of the matrix $\mathbf{M}$ for the eigenvalue 1. To avoid heavy computations, the steady state was approximated by multiplying $\mathbf{M}$ by itself for a large number of times, 100, and applying it to $\mathbf{v}^{(0)}$. This returned an energy landscape with higher intensities corresponding to the pixels with the highest probability of being carcinogenic.

Active contouring was implemented in its discrete approximation, with the image energy term derived from the result of the sandpiles. Equation (2.7) becomes:

$$E_{\text{img}}(i) = \delta \cdot \mathbf{v}_i \tag{3.20}$$

where δ is an arbitrarily chosen weight associated with the image energy and $\mathbf{v}_i$ corresponds to the amount of sand at steady state for each pixel, information which is contained in the vector $\mathbf{v}^{(N)}$. A circular contour of radius 200 pixels was initialised around the highest sandpile. Iteratively, the total

energy of each possible neighbouring site was calculated for each point. The location providing lower energy was then chosen as the new point position. The algorithm was executed until the achievement of a stable final position. Precautions were taken into consideration to better guide the contour through the energy potential. The stiffness was modelled to be linearly proportional to the point distance with respect to its closest neighbours. The constraint term ensures energy dependence on the average distance between points to correct anomalous behaviour due to the heterogeneous density of the contour. Finally, large spacing between neighbours was filled with new points to avoid missing important features in the image. The results presented in this study were obtained with the following weights for each energy component of Eq. (2.5)-(2.8):

$$\alpha(s) \propto 9 \cdot d_{max}, \quad \beta = 4, \quad \gamma = 1 \quad \text{and} \quad \delta = 20000 \tag{3.21}$$

where $d_{\max}$ is the maximum distance to neighbouring points.

4. Results

In this section, we present the results of the networks' predictions after a qualitative analysis of the photoacoustic absorption spectra for the different tumour types. The final results from sandpiles and active contour algorithms are also shown.

4.1. Spectral analysis

The photoacoustic spectra of healthy and tumourous clusters are shown in Fig. 9 for all the samples. These were obtained by averaging the spectrum of each pixel included in the clusters defined in section 3.3., thus excluding a margin of 0.5 mm between the two groups. Samples 128, 131, and 153 were assessed to be malignant melanoma, while specimens 389 and 266 belong to the group of BCC. Sample 394, yet to be histopathologically examined, was identified as an abnormal nevus with chances of evolving into malignant melanoma. Melanin, blood, and collagen are the main endogenous contrast agents of photoacoustic absorption in the samples examined [10]. The spectral intensity of the MM clusters seems to always be higher in intensity due to the larger concentration of melanin. This behaviour is very different compared to the BCC spectra, which present intensities lower than the healthy tissue signal for higher wavelengths (Fig. 9). The ratio of chromophores changes between tumour classes and also vary between patients, as shown by the anomalous slope of sample 153. Focusing on the analysis of the singular sample decreases the variation between classes and patients, reducing the problem to a simple binary classification. Exciting the sample with a wide range of wavelengths gives the advantage of obtaining more information about the spectra.

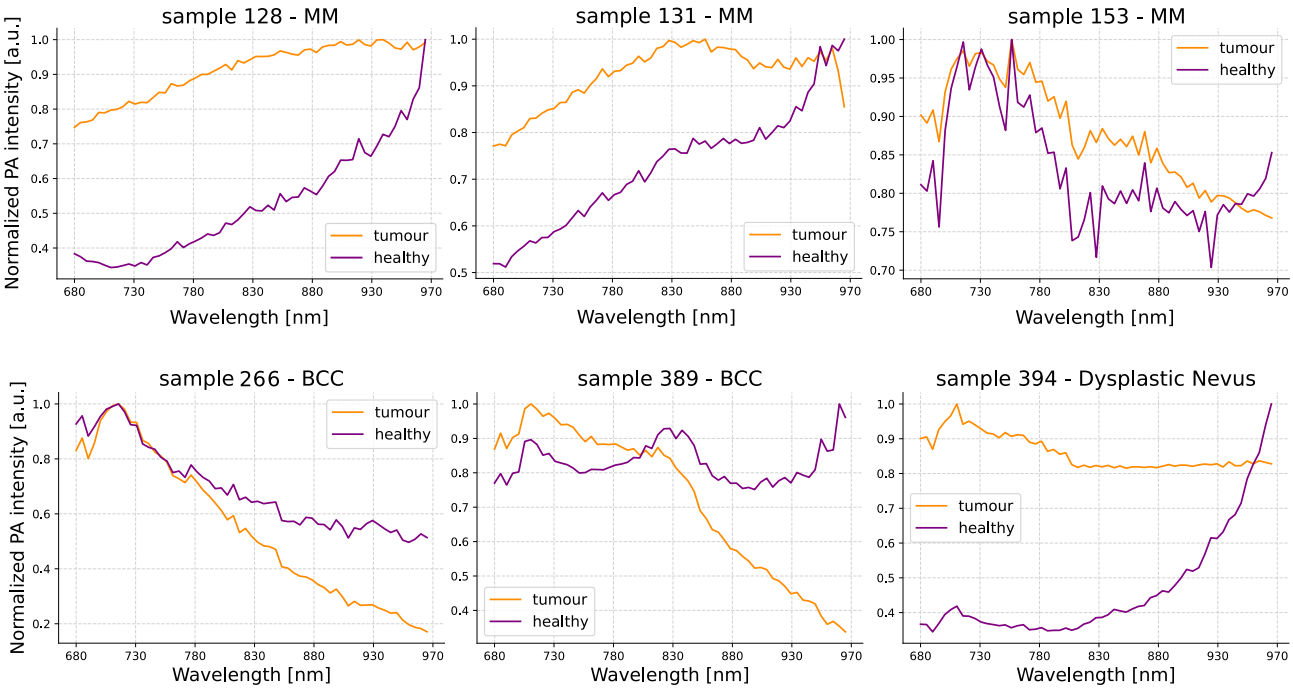

Figure 9: **Photoacoustic absorption spectra:** in orange, the average spectra of all the sample's pixels labelled as tumour, in purple the healthy spectra. The intensity is normalised with respect to the maximum value. The six tumour types are indicated in the title.

4.2. Predictions

We now show the application of our networks to three of the samples for which histopathology was available at the time of this study. This is done on the pre-processed data, on the data reduced to two principal components through PCA, and on the data compressed by the autoencoder. The network deemed as the most accurate is additionally applied to evaluate the samples lacking histopathology.

4.2.1. *Predictions on pre-processed data*

The tuned MLP and CNN were trained on 2000 random points selected from the clusters previously defined and tested on the full samples. Figure 10 shows the predictions for each pixel mapped back to the location on the original specimen. As stated earlier, the threshold on the output probability was set to 0.5 since the training points were evenly distributed between classes. As per labelling, 0 represents the healthy class while 1 corresponds to the tumour. The size reported in millimetres corresponds to the distance between extremities respectively for the horizontal and vertical directions. The predictions are only shown for the relevant vertical cross-section of the sample approximately corresponding to the slice examined by the pathologist.

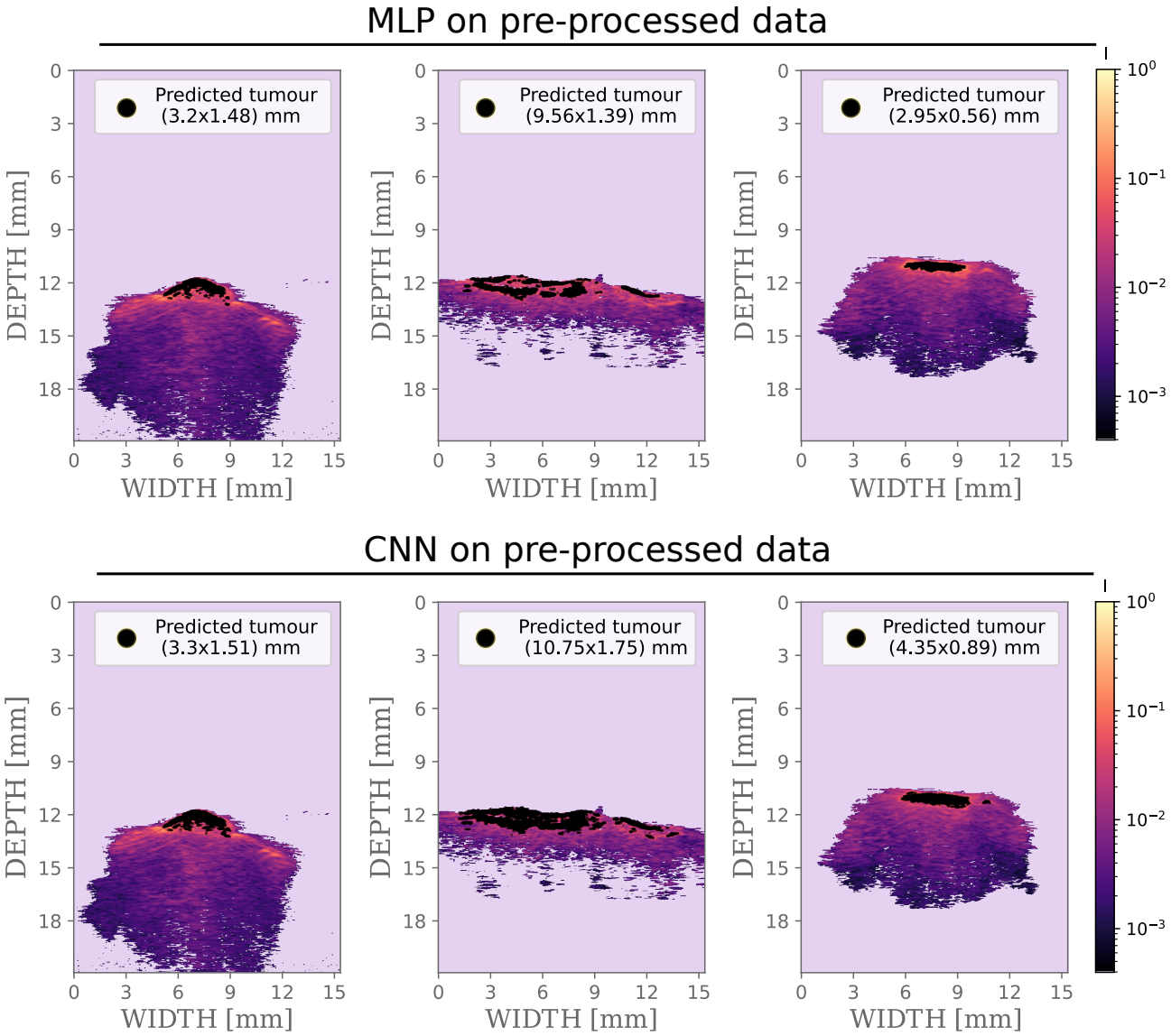

Figure 10: **MLP and CNN predictions on pre-processed data:** the spectra of samples 128, 131 and 153 were used as 59-dimensional input to the classifiers. In black are the pixels belonging to the tumour. In the legend, the size of the tumour is expressed in mm. The colour bar representing the intensity is reported with logarithmic scale. The PA image shown is the 680 nm channel.

The predictions of the classifiers are extremely close to the histopathologically measured sizes of the carcinomas (Table 1). While both networks seem to slightly overestimate the depth of the tumour, the MLP significantly underestimate its lateral extension. Moreover, the accuracy of the prediction highly varies according to the sample investigated. The training history, which is included in the Appendix (Fig. 17), reports maximal accuracy, both for the training and validation dataset. However, due to the lack of completely labelled images, the accuracy is not a faithful indicator of the networks' success. Since the purpose of our study is to establish a safe method for carcinoma delineation, a technique slightly overestimating the extension of the tumour is preferred instead of risking incomplete eradication. With the purpose of improving the estimation accuracy, dimensionality reduction is applied to the data.

4.2.2. *Predictions on dimensionally reduced data*

PCA was applied to the pre-processed data to reduce their dimensionality to both 2 and 10 principal components to produce the input to the classifiers. The choice of components was driven by visualisation purposes and comparison to the autoencoder architecture, which includes 10 latent nodes. Figure 11 depicts the distribution of the points with respect to their two principal components. The colour identifies the label that was assigned through k-means clustering. Intuitively, no unique groups can be identified. However, the distribution seems to have a similar behaviour between samples. As discussed in section 4.2.1., without precise and complete labelling of the images, the separation between the two clusters is quite broad, introducing a source of incertitude in our predictions.

The same MLP and CNN were applied to classify the compressed input. Figure 12 shows the predictions of the networks on the same samples. Since the first components are the ones responsible for the highest variance, we only show the predictions on 2 components, which are almost identical to the tumour estimation using all 10 of them.

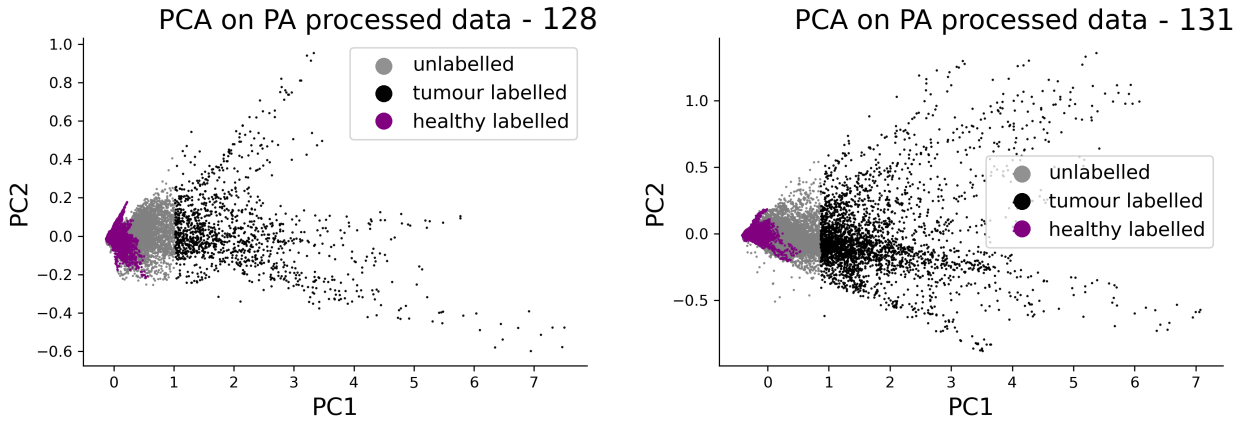

Figure 11: **PCA results:** principal component analysis is used to reduce the data to two main components. The pixels that were previously labelled are coloured to identify the tumorous and healthy regions.

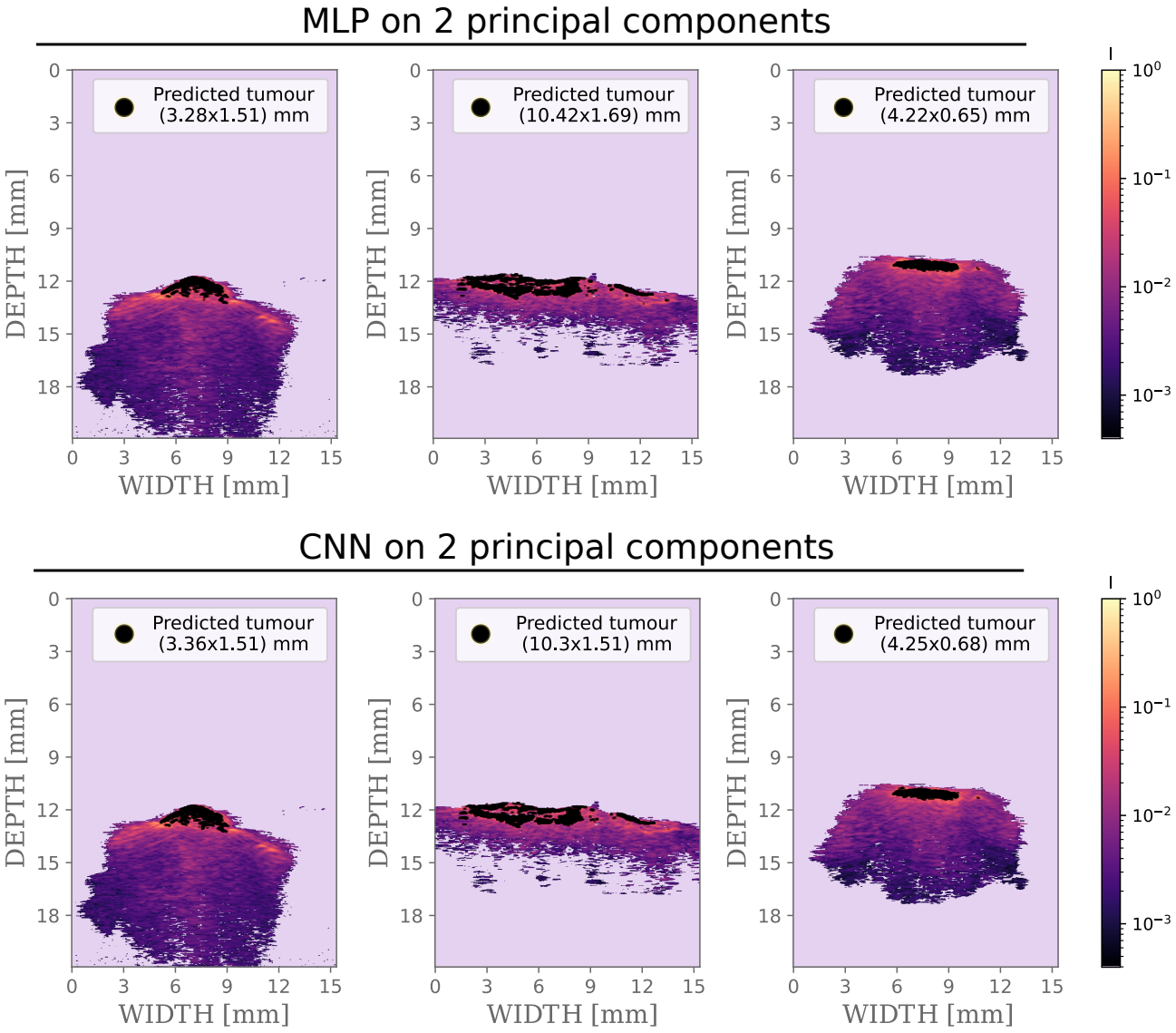

Figure 12: **MLP and CNN predictions on 2 principal components:** the spectra of samples 128, 131 and 153 were used as 2-dimensional input to the classifiers. In black are the pixels belonging to the tumour. In the legend, the size of the tumour is expressed in mm. The colour bar representing the intensity is reported with logarithmic scale. The PA image shown is the 680 nm channel.

When performing PCA, reducing to two components discards the advantage of using convolutional layers and produces results very similar to the MLP classifier. Both size estimations seem to be intermediate results between the MLP and CNN's predictions on pre-processed data.

With the intention of improving the information contained in the compressed representation of the data, these were processed by a non-linear autoencoder. The model determined in section 3.4.3. was applied to reduce the dimensionality of the data from 59 to 10. The same MLP and CNN were then used to classify the spectra of the same samples. Figure 13 presents the predictions of the classifiers on the compressed data.

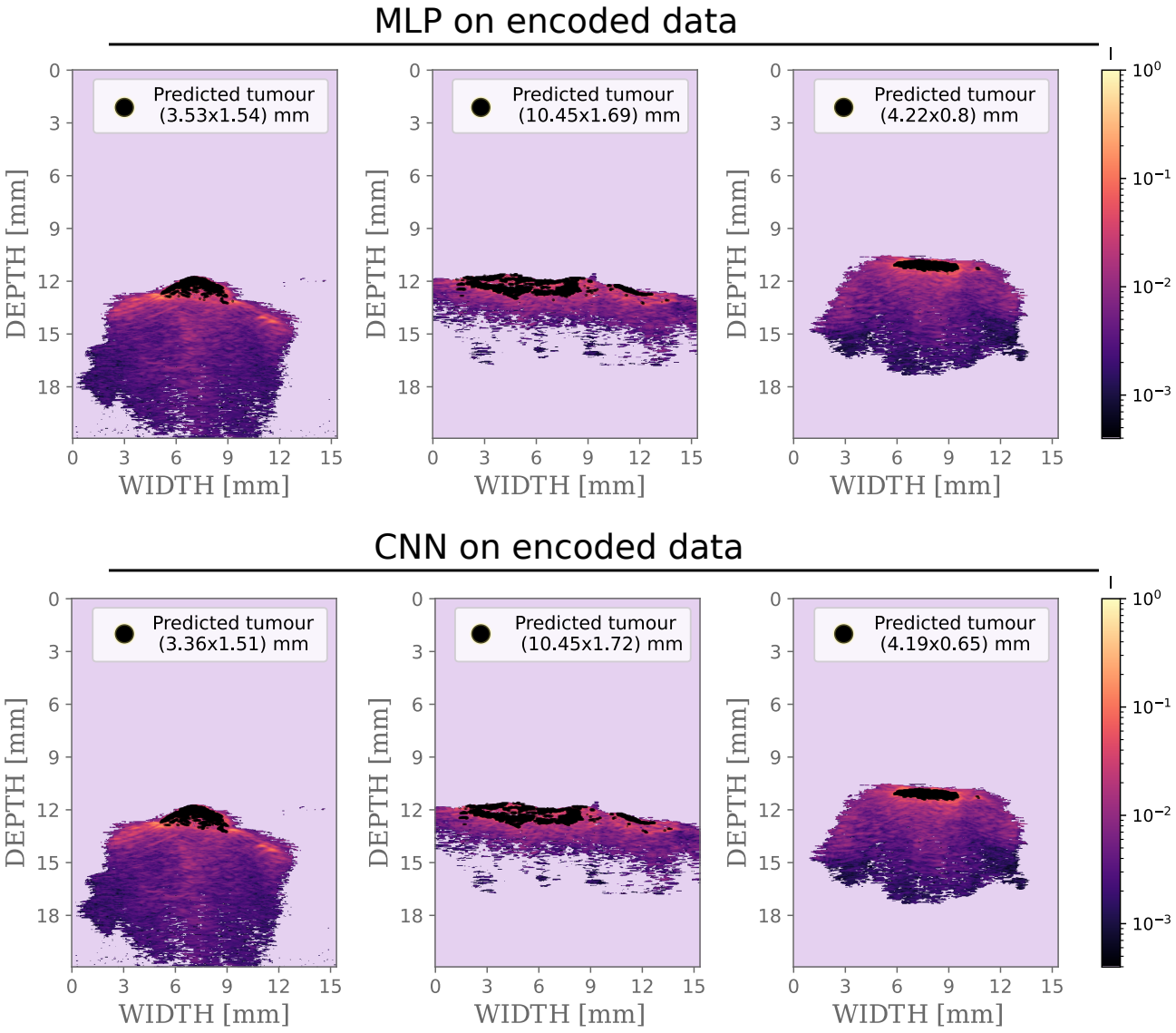

Figure 13: **MLP and CNN predictions on encoded data:** the spectra of samples 128, 131 and 153 were used as 10-dimensional input to the classifiers. In black are the pixels belonging to the tumour. In the legend, the size of the tumour is expressed in mm. The colour bar representing the intensity is reported with logarithmic scale. The PA image shown is the 680 nm channel.

The results show no significant difference compared to the predictions performed on two principal components. Nevertheless, all of the results are constantly very coherent in size and shape to the histopathological examinations. Particularly, our error needs to be reconsidered relative to the safe margin extension of healthy tissue that is currently extracted during clinical procedures. While underestimation of the tumour borders might cause only its partial eradication, a model which is scarcely overestimating its size can still constitute an improvement compared to current clinical procedures. For this reason, the CNN is chosen as the best-performing network and its predictions on the 59-dimensional data will be later processed to obtain final predictions.

4.2.3. *Predictions on undetermined samples*

The convolutional network was employed to predict the tumour extension on samples 389 and 394 which, at the time of this study, lacked histopathological assessments. Without the pathologist's opinion, the effectiveness of our estimation can not be confidently assessed. However, the predicted tumours are similar in size and intensity compared to previously analysed samples, suggesting the success of our framework on various types of skin tumours.

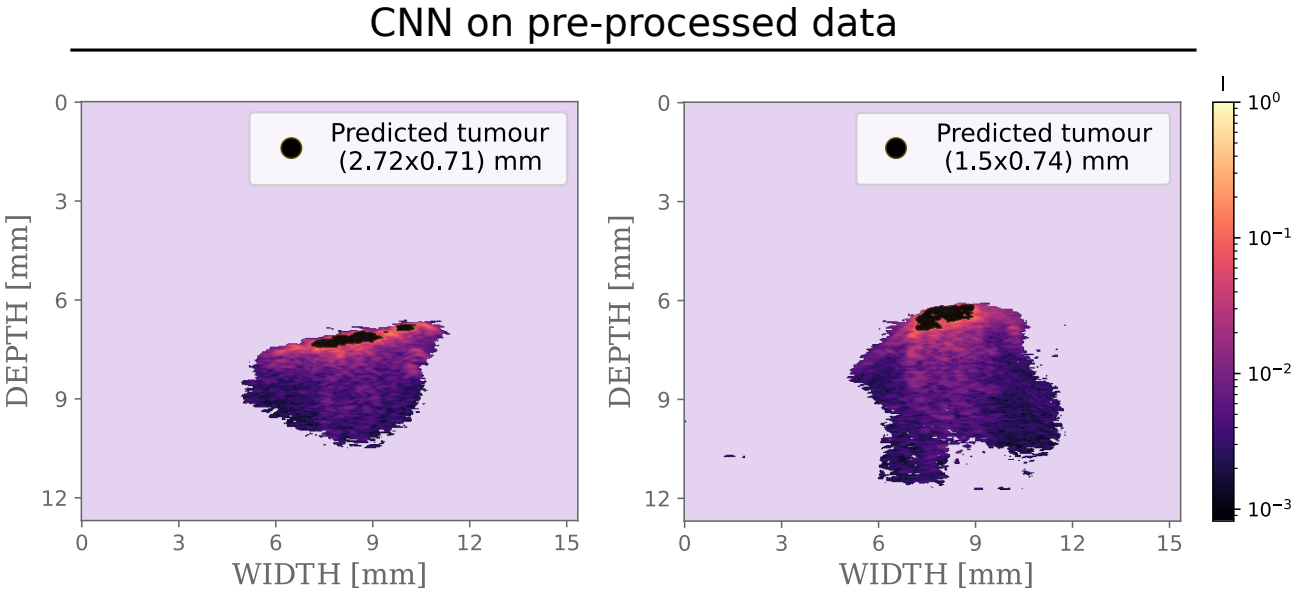

Figure 14: **CNN predictions on undetermined tumour samples:** with the same approach, the pixels' spectra of samples 389 and 394 were used as input to the CNN classifier. In black are the pixels belonging to the tumour. In the legend, the size of the tumour is expressed in mm. The colour bar representing the intensity is reported with logarithmic scale. The PA image shown is the 680 nm channel.

4.3. Tumour segmentation

In this section, we present the results of the segmentation algorithms employed on the best prediction for some of the samples. Notably, this includes the results of the sandpiles and active contouring of the tumour region.

4.3.1. *Image Energy Landscape*

The predictions of the CNN on the pre-processed data are chosen as the basis for the post-processing algorithm. These were used to construct an energy landscape for active contouring through the previously illustrated sandpiles algorithm. Figure 15 shows the final height of the sandpiles at the steady state for samples 153 and 128. The amount of sand on each pixel is normalised with respect to its maximal value in the sample. For visualisation purposes, the sample is given an arbitrary intensity to be distinguished from the background. However, during the implementation, zero intensity was given to the healthy part of the sample and to the background.

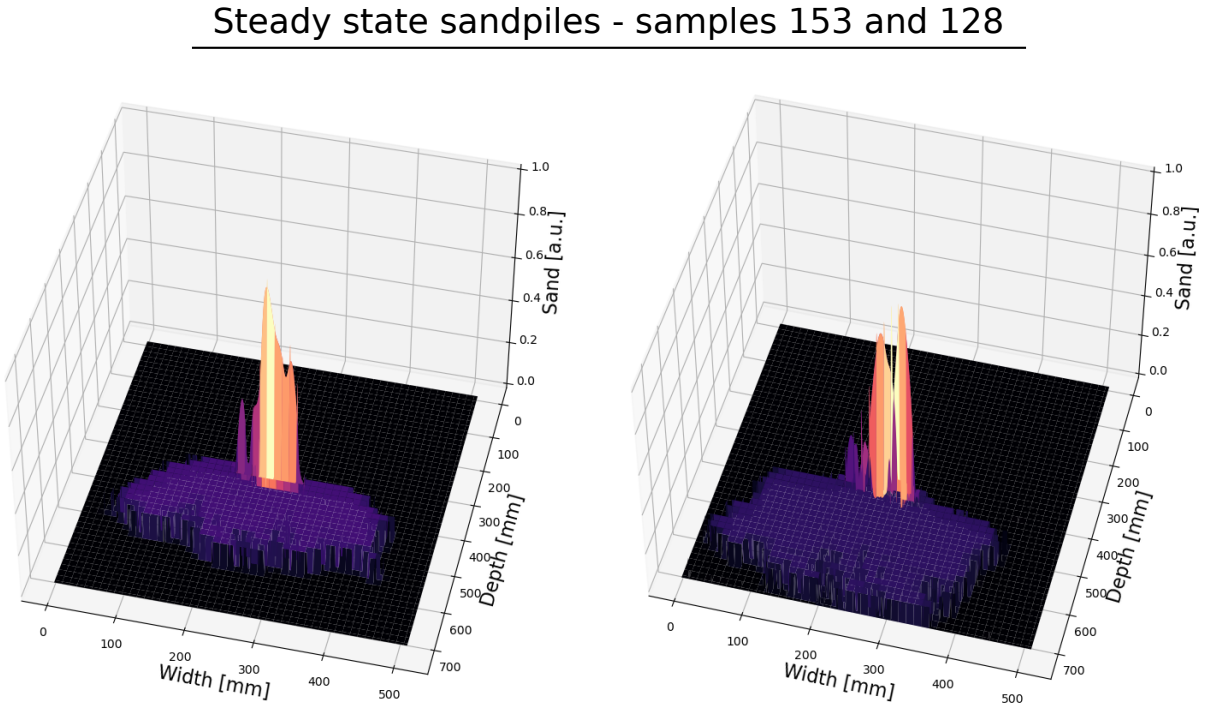

Figure 15: **Sandpiles results:** on the left, the steady state sandpiles for sample 153. On the right, the final sandpiles for sample 128. The colour is proportional to the intensity, which is also indicated on the z axis.

4.3.2. *Active Contouring*

The contour was initialised as a closed circular boundary centred around the pixels with the highest probability of being carcinogenic. The algorithm was run on the different samples. Figure 16 shows the initial and final position of the contour after 250 iterations. The extremities of the contour were measured for a quantitative comparison with the size of the assessed tumours and the classifier predictions.
As it can be noticed in Fig. 16, some of the lower sandpiles allow the contour to go past them, while this stops before higher intensity areas. Additionally, small groups of healthy pixels contoured by carcinoma were included in the tumour area, as one would intuitively assume. This segmentation technique can remarkably improve border delineation by correcting the predictions of the networks and by creating a clean contour of excision.

5. Discussion

The samples analysed in this thesis belong to two of the main classes of skin carcinoma: malignant melanoma and basal cell carcinoma. As expected, these show different photoacoustic spectra due to their distinctive chemical composition. However, for the same type of carcinoma, very dissimilar behaviours can be observed between patients. Although the tumour cluster frequently shows considerably higher intensity, the trend of the spectra varies for different samples. Processing each sample individually allows the networks to specialise in the unique spectra of the patient's skin and tumour type, thus

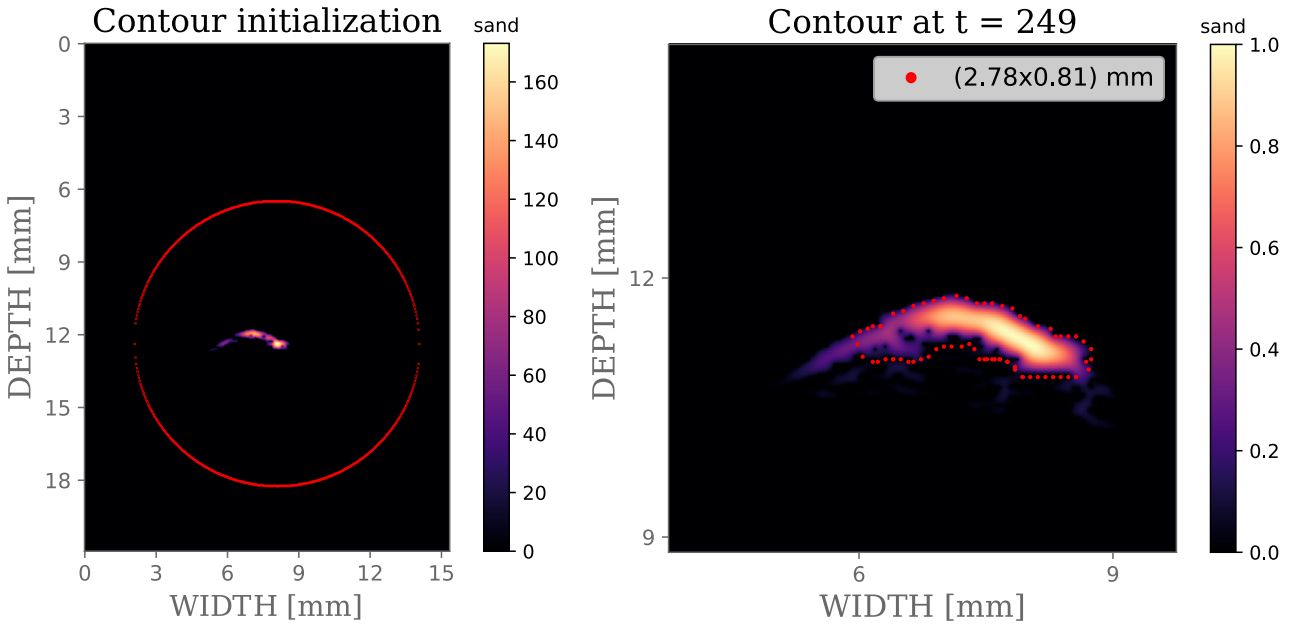

Figure 16: **Active contouring:** in red, the position of the contour at the initial and final time on the image energy landscape for sample 128. The final size of the contour is indicated on the upper right corner in mm.

simplifying the classification. This approach also compensates for the lack of samples, since each pixel constitutes a data point. On the other hand, multiple samples would improve model selection and their robustness by enabling us to test our approach on a larger variety of tumour types and patient variation.

K-means clustering offers a simple way to isolate high-intensity pixels typically belonging to carcinogenic tissue. Nevertheless, a supervised algorithm increases the amount of information given to the network through the introduction of labelling or neighbouring spectral intensities. Model selection additionally permits us to adjust our results and further improve class prediction. Ultimately, a classifier returns both the class identity and the probability of belonging to that class for each pixel. The latter is critical for the construction of the energy landscape, which is used during the final delineation of the tumour by the active contour algorithm.

For the samples shown, the network was trained on one slice. However, training on all of the slices was implemented to include more information about different sections of the specimen. The predictions can be interpolated to allow for a three-dimensional view of the tumour volume. When analysing the full extension of the imaging results, highly intense noise might be present at the extremities of the samples. This is most likely due to the high absorbance of blood residual from excision. The problem can be solved by neglecting predictions on areas of the samples belonging to the border. In-vivo measurements will not present this issue.

The predictions seem to overall confirm the presence and shape of the tumour in accordance with the histopathological examination. In particular, the CNN offers a safer model due to its minor overestimation of the region of interest. In the case of our data, dimensionality reduction can subtly improve the prediction of the MLP classifiers by increasing the number of pixels included in the tumour

area. To confidently assess the best-performing model, more tumour variations would be needed. Nevertheless, our approach attempts to establish a pipeline to treat medical data that might present a high number of dimensions and for which data compression might improve data storage and network performance.

Although a spatial relation between neighbouring pixels could not be included by the implementation of a two-dimensional CNN, the sandpiles algorithm allows the inclusion of such information through the sand exchange system. With healthy pixels acting as sinks, small clusters of wrongly classified tumour points disappear when surrounded by healthy tissue. Small clusters of healthy points surrounded by carcinoma are instead taken care of by the implementation of active contouring.

To assess the performance of the network, histopathological sections were always used to identify the tumour shape and size. As explained earlier, the reference to histopathology is only approximate, due to the deformation the sample undergoes during the procedure. Regardless, the aim of this research is to establish a procedure that allows us to make assessments on new samples which have not yet been excised and examined. This was already started by testing our framework on undetermined samples.

The procedure was tested on an additional non-pigmented sample, 266. In this case, the tumour does not present an increased concentration of melanin, which is the main contrast agent for photoacoustic imaging. Without high-intensity pixels, the clustering method is not able to return a systematic separation between healthy and tumour tissue. This suggests the need for a different approach in the case of non-pigmented carcinoma. A different strategy could consist of an analysis of the US images, which highlights instead the structural difference between healthy and anomalous tissue.

From the analysis of our samples, it is obvious that the quantity of healthy skin removed during clinical examination is significantly larger than the area of the carcinoma. When the tumour is situated where a thicker layer of tissue can be removed, the affected region is usually excised with safe margins of around 4 mm for low-risk tumours, and up to 6 mm for higher-risk tumours [10]. However, since partial eradication would increase the number of surgeries needed, larger portions might be removed according to the doctor's assessment. Our approach enables us to explore the full extension of margin without the need for excision, contrary to histopathology, in which the analysed slices only represent around 1-2% of the full margin area [30]. Our implementation could potentially eliminate the need for surgeries in case of false positives. It would also considerably decrease the amount of healthy tissue excised along with the invasiveness and patient discomfort associated. Considering this, our incertitude on the tumour borders would be negligible as extremely smaller safe margins could be established.

6. Conclusion and Outlook

The results obtained suggest the success of our approach for the treatment of photoacoustic data in the presence of endogenous contrast agents. Specifically, segmentation of skin cancer is possible with relatively high accuracy without any need for preliminary surgery. These artificial neural network tools can be used to both analyse and compress the data to a lower dimension. Although dimensionality reduction did not consistently improve the predictions of the classifiers, it can still constitute a possible

strategy to treat data with numerous dimensions. Regardless of the small discrepancy, automation of such a routine procedure would highly impact skin cancer diagnosis and treatment.

Although the final contour was shown only for single cross-sections, the procedure can be extended to three dimensions. This would include a three-dimensional redefinition of the sandpiles algorithm as well as the introduction of a higher-dimension contour surface. For implementation of the latter, one could choose among multiple options, for instance: interpolation between the final contours on each cross-section, or the definition of a three-dimensional contour with the addition of appropriate forces controlling the movement of its surface.

More complex structures of MLPs or CNNs could be explored. A finer hyperparameter tuning of the networks, along with an improvement of the active contour algorithm can further decrease the error on the final prediction. Together with more samples, this would allow us to establish robust classifiers trained to accurately distinguish and segment tumour pixels.

Our procedure can definitely highlight the correct shape and approximate size of the tumour with a few tens of millimetres of residual compared to the histopathological examinations. As previously discussed, such error is negligible when compared to the current clinical safe margins of excision. This approach would innovate current skin cancer treatment by eliminating the need for histopathological diagnosis. It will offer automatic assistance during tumour identification and segmentation.

References

[1] WHO: World Health Organization. **Radiation: Ultraviolet (UV) radiation and skin cancer**.

[2] NIH: National Institutes of Health. **Skin cancer**.

[3] Randy Gordon. **Skin cancer: an overview of epidemiology and risk factors**. *Semin. Oncol. Nurs.*, 29:160–169, 2013.

[4] Cristiane Benvenuto-Andrade, Stephen W. Dusza, Anna Liza C. Agero, Alon Scope, Milind Rajadhyaksha, Allan C. Halpern, Ashfaq A. Marghoob. **Differences Between Polarized Light Dermoscopy and Immersion Contact Dermoscopy for the Evaluation of Skin Lesions**. *Archives of Dermatology*, 143:329–338, 2007.

[5] Ronald P. Rapini. **Pitfalls of Mohs micrographic surgery**. *J. Am. Acad. Dermatol.*, 22:681–686, 1990.

[6] Jenny Hult, Ulf Dahlstrand, Aboma Merdasa, Karin Wickerström, Rehan Chakari, Bertil Persson, Magnus Cinthio, Tobias Erlöv, John Albinsson, Bodil Gesslein, Rafi Sheikh, Malin Malmsjö. **Unique spectral signature of human cutaneous squamous cell carcinoma by photoacoustic imaging**. *J. Biophotonics*, 13, 2020.

[7] Gharieb, Reda. ***Photoacoustic Imaging for Cancer Diagnosis: A Breast Tumor Example***. 2020.

[8] Halil Arslan , Bahar Pehlivanöz. **Determination of fluence rate distribution in a multilayered skin tissue model by using Monte Carlo simulations**. *Turkish Journal of Physics*, 43:286–92, 2019.

[9] Omnia Hamdy, Ibrahim Abdelhalim. **Diagnosing different types of skin carcinoma based on their optical properties: A Monte-Carlo implementation**. *IOP Conf. Ser. Mat. Science and Eng.*, 2021.

[10] Magne T. Stridh, Jenny Hult, Aboma Merdasa, John Albinsson, Agnes Pekar-Lukacs, Bodil Gesslein, Ulf Dahlstrand, Karl Engelsberg, Johanna Berggren, Magnus Cinthio, Rafi Sheikh, and Malin Malmsjö. **Photoacoustic imaging of periorbital skin cancer ex vivo: unique spectral signatures of malignant melanoma, basal, and squamous cell carcinoma**. *Biomed. Opt. Express*, 13:410–425, 2022.

[11] Jitendra V. Tembhurne, Nachiketa Hebbar, Hemprasad Y. Patil, Tausif Diwan. **Skin cancer detection using ensemble of machine learning and deep learning techniques.** *Multimed Tools Appl.*, 2023.

[12] Lequan Yu, Hao Chen, Qi Dou, Jing Qin, Pheng-Ann Heng. **Automated Melanoma Recognition in Dermoscopy Images via Very Deep Residual Networks**. *IEEE Transactions on Medical Imaging*, 36:994–1004, 2017.

[13] Jianpeng Zhang, Yutong Xie, Yong Xia, Chunhua Shen. **Attention Residual Learning for Skin Lesion Classification**. *IEEE Transactions on Medical Imaging*, 38:2092–2103, 2019.

[14] Khalid M. Hosny, Mohamed A. Kassem, Mohamed M. Foaud. **Skin Cancer Classification using Deep Learning and Transfer Learning**. *9th Cairo International Biomedical Engineering Conference (CIBEC),* 90-93, 2018.

[15] Tehseen Mazhar, Inayatul Haq, A. Ditta, S. Mohsan, Faisal Rehman, I. Zafar, J. Gansau, L. P. W. Goh. **The Role of Machine Learning and Deep Learning Approaches for the Detection of Skin Cancer.** *Healthcare (Basel)*, 2023.

[16] Emil Andersson, Jenny Hult, Carl Troein, Magne Stridh, Benjamin Sjögren, Agnes Pekar-Lukacs, Patrik Edén, Bertil Persson, Malin Malmsjö, Victor Olariu, Aboma Merdasa. **Towards precision skin tumor diagnostics with hyperspectral imaging and spectroscopically-guided machine learning**. *Unpublished manuscript.*

[17] Gayathry S. Warrier, T. M. Amirthalakshmi, K. Nimala, T. Thaj Mary Delsy, P. Stella Rose Malar, G. Ramkumar, and Raja Raju. **Automated Recognition of Cancer Tissues through Deep Learning Framework from the Photoacoustic Specimen**. *Contrast Media and Molecular Imaging*, 2022.

[18] Jiayao Zhang, Bin Chen, Meng Zhou, Hengrong Lan, Fei Gao. **Photoacoustic Image Classification and Segmentation of Breast Cancer: A Feasibility Study**. *IEEE Access*, 7:5457–5466, 2019.

[19] Jiayan Li, Yingna Chen, Wanli Ye, Mengjiao Zhang, Jingtao Zhu, Wenxiang Zhi, Qian Cheng. **Molecular breast cancer subtype identification using photoacoustic spectral analysis and machine learning at the biomacromolecular level**. *Photoacoustics*, 30:100483, 2023.

[20] Yingna Chen, Chengdang Xu, Zhaoyu Zhang, Anqi Zhu, Xixi Xu, Jing Pan, Ying Liu, Denglong Wu, Shengsong Huang, Qian Cheng. **Prostate cancer identification via photoacoustic spectroscopy and machine learning**. *Photoacoustics*, 23:100280, 2021.

[21] Hugh M. Gloster Jr. , Kenneth Neal. **Skin cancer in skin of color**. *J. Am. Acad. Dematol.*, 55:714–60, 2006.

[22] Matthew B. A. McDermott, Shirly Wang, Nikki Marinsek, Rajesh Ranganath, Luca Foschini, Marzyeh Ghassemi. **Reproducibility in machine learning for health research: Still a ways to go**. *Sci. Transl. Med.*, 13, 2021.

[23] Lihong V. Wang, Song Hu. **Photoacoustic Tomography: In Vivo Imaging from Organelles to Organs.** *Science*, 335, 2012.

[24] Nimrod M. Tole. **Basic physics of ultrasonographic imaging**. *WHO Library Cataloguing-in-Publication Data*, 2005.

[25] Jacqueline Dinnes, Jeffrey Bamber, Naomi Chuchu, Susan E. Bayliss, Yemisi Takwoingi, Clare Davenport, Kathie Godfrey, Colette O'Sullivan, Rubeta N. Matin, Jonathan J. Deeks, Hywel C Williams. **High frequency ultrasound for diagnosing skin cancer in adults**. *Cochrane Database Syst. Rev.*, 2018.

[26] Ivan Nunes da Silva, Danilo Hernane Spatti, Rogerio Andrade Flauzino, Luisa Helena Bartocci Liboni, Silas Franco dos Reis Alves. ***Artificial Neural Networks: A practical course***. Brazil, 2018.

[27] Charu C. Aggarwal. ***Neural Networks and Deep Learning: A Textbook***. 1st edition, 2018.

[28] Ian Goodfellow, Yoshua Bengio, Aaron Courville. ***Deep Learning***. 2016. `http://www.deeplearningbook.org`.

[29] Michael Kass, Andrew Witkin, Demetri Terzopoulos. **Snakes: Active contour models**. *International Journal of Computer Vision*, 1:321–331, 1988.

[30] Stanislav N. Tolkachjov, David G. Brodland, Brett M. Coldiron, Michael J. Fazio, George J. Hruza, Randall K. Roenigk, Howard W. Rogers, John A. Zitelli, Daniel S. Winchester, Christopher B. Harmon. **Understanding Mohs Micrographic Surgery: A Review and Practical Guide for the Nondermatologist**. *Mayo. Clin. Proc.*, 92:1261–1271, 2017.

Appendix

Table 2: **MLP gridsearch:** ranges of hyperparameters tested during model selection. The loss was not reported since always null, but a visual assessment was made of the predicted tumour areas in comparison to histopathological examinations.

number of nodes	learning rate	batch size	epochs
[1, 2, 3, 4, 6, 10, 15]	[0.001, 0.005, 0.01, 0.05, 0.1]	[160, 320, 800]	[400, 800, 1200]

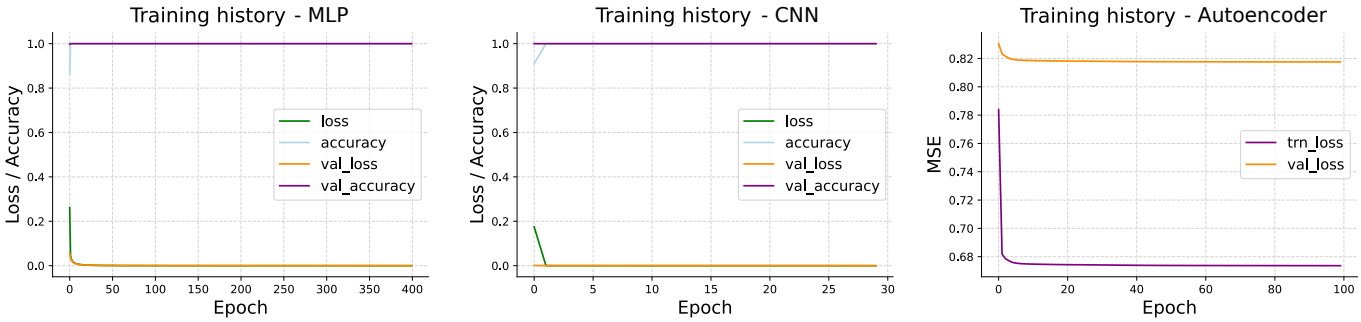

Figure 17: **Networks' training history:** left and centre, training and validation error for the two classifiers, which minimises the binary cross-entropy error. Right, training and validation history for the autoencoder, which minimises the mean square error. All training histories correspond to the usage of the networks on the pre-processed sample 128.