Chapter 3

Automatic Interpretation of Melanocytic Images in Confocal Laser Scanning Microscopy

Marco Wiltgen and Marcus Bloice

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/63404

Abstract

The frequency of melanoma doubles every 20 years. The early detection of malignant changes augments the therapy success. Confocal laser scanning microscopy (CLSM) enables the noninvasive examination of skin tissue. To diminish the need for training and to improve diagnostic accuracy, computer-aided diagnostic systems are required. Two approaches are presented: a multiresolution analysis and an approach based on deep layer convolutional neural networks. For the diagnosis of the CLSM views, architectural structures such as micro-anatomic structures and cell nests are used as guidelines by the dermatologists. Features based on the wavelet transform enable an exploration of architectural structures at different spatial scales. The subjective diagnostic criteria are objectively reproduced. A tree-based machine-learning algorithm captures the decision structure explicitly and the decision steps are used as diagnostic rules. Deep layer neural networks require no a priori domain knowledge. They are capable of learning their own discriminatory features through the direct analysis of image data. However, deep layer neural networks require large amounts of processing power to learn. Therefore, modern neural network training is performed using graphics cards, which typically possess many hundreds of small, modestly powerful cores that calculate massively in parallel. Readers will learn how to apply multiresolution analysis and modern deep learning neural network techniques to medical image analysis problems.

Keywords: confocal laser scanning microscopy, skin lesions, multiresolution image analysis, convolutional neural networks, machine learning, computer-aided diagnosis

1. Introduction

The skin is the largest organ of the body. Its surface comprises up to two square meters. It is the organ that is in direct contact to the environment and is therefore exposed to several environmental influences such as sun radiation, temperature, infections. The skin consists of three main layers: the epidermis, the dermis and the hypodermis (subcutis), whereby each layer is subdivided into several sublayers (strata) [1]. As the outermost layer, the epidermis provides a protective barrier of the body's surface which keeps water in the body, protects against heat and ultraviolet radiation and prevents infections (caused by bacteria, fungi, parasites, etc.) [2, 3]. The horny layer (stratum corneum), which is the top layer of the epidermis, undergoes a continuous process of renovation (every 4 weeks). Keratinocytes, which represents 90% of the cell types in the epidermis, protect the body against ultraviolet radiation. Keratinocytes are derived from epidermal stem cells residing in the lower part of the epidermis (stratum basalis). During their lifetime, they migrate through the different strata of the epidermis. Via this process, they are pressed to the epidermis surface by the continuously succeeding cells. During the migration through the different strata, the keratinocytes cells undergo multiple stages of differentiation, whereby they change shape and composition and are filled with keratin. Different stages and corresponding strata are represented in **Figure 1**. Keratin, a structural protein, is the key structural material making up the outer layer of the epidermis and protects the cells from damage or stress. On their way to the outermost strata, the keratinocytes lose liquid and become hornier. Corneocytes are keratinocytes that have completed their differentiation program. They are dead cells in the stratum corneum and are shed off (by desquamation) as new ones come in. Keratinocytes protect against ultraviolet radiation by taking up melanosomes from epidermal melanocytes. The melanosomes are vesicles which contain the endogenous photo protectant molecule melanin. Melanocytes are melanin producing cells which comprise between 5 and 10% of the cells in the basal layer (stratum basalis) of the epidermis. The production of the skin pigment melanin is stimulated by ultraviolet radiation (melanogenesis). Melanocytes have several arm-like structures (dendrites) that stretch out to connect them with many keratinocytes. Once synthesized, melanin is contained in the melanosomes and moved along the dendrites to reach the keratinocytes. The melanin molecules are stored within keratinocytes (and melanocytes) in the perinuclear area, around the nucleus, where they protect the DNA against ultraviolet radiation. Thereby, a melanin molecule transforms nearly all the radiation energy in to heat. This is done by ultrafast internal conversation of the energy from the excited electronic states into vibrational modes. The ultrafast conversion shortens the lifetime of the excitation states and therefore prevents the formation of harmful free radicals.

The dermis is connected to the epidermis through a basement membrane (a thin sheet of fibres) and provides anchoring and nourishment for the epidermis. The dermis contains collagen (stability), elastic fibres (elasticity) and an extrafibrillar matrix as structural components. The papillary region (stratum papillae) in the dermis is composed of connective tissue which extends towards the epidermis. These finger-like projections are called papillae and strengthen the connection between the dermis and the epidermis. In addition to the structural components, blood vessels are present in the dermis providing nourishment for the dermal and

http://dx.doi.org/10.5772/63404

epidermal cells. Furthermore, the dermis contains hair follicles, sweat glands and lymphatic vessels. (In addition to the presented components, the dermis also contains mechanoreceptors that enable the sense of touch and thermoreceptors that provide the sense of heat). The hypodermis is beneath the dermis. Its tasks comprise energy storage, heat insulation and the connection of the skin with inner structures like muscles and bones. The hypodermis consists primarily of loose connective tissue and adipocytes (fat cells), which are grouped together in lobules (subcutaneous fat). Furthermore, the hypodermis contains larger blood vessels and nerves than those found in the dermis.

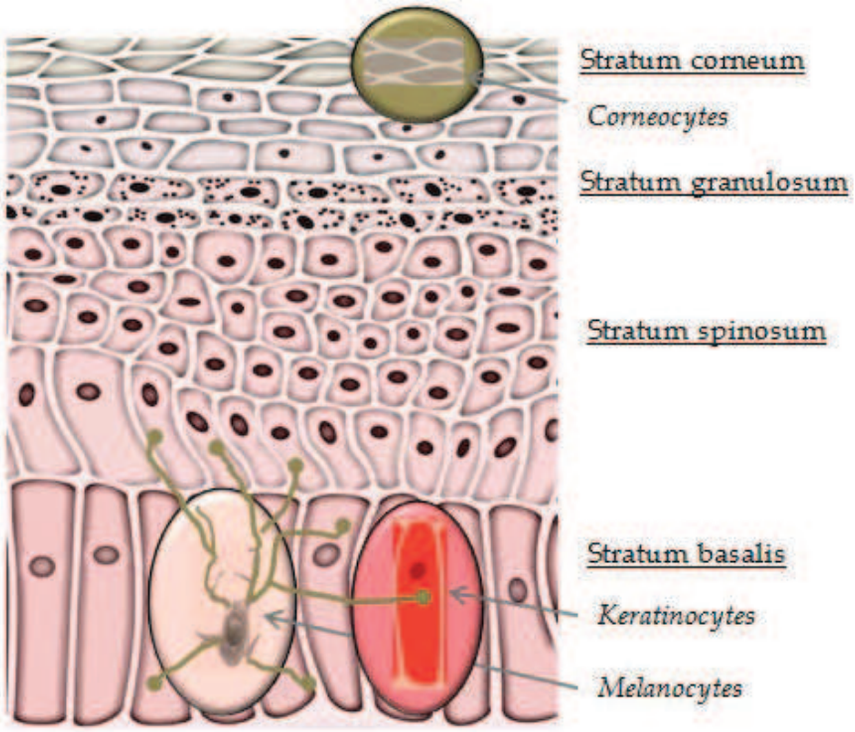

Figure 1. The layer architecture of the epidermis.

2. Malignant melanoma and benign nevi

The primary cause for the increasing number of melanomas is the extreme sun exposure during sun-bathing (especially for people with low levels of skin pigment). The malignant melanoma is a type of cancer that develops from the pigment containing melanocytes [4]. Melanomas are mainly caused by DNA damage resulting from the ultraviolet radiation [5]. It is observed that strongly pigmented people are less susceptible to (sun induced) melanomas, which demonstrates the protection function of melanin. At the early stage, melanocytes begin an out-of-control growth [5]. In a posterior stage (invasive melanoma), the melanoma may grow into the surrounding tissue and can spread out around the body through lymph or blood vessels deeper in the skin. People with melanomas at the early stage are treated by surgical removal of the skin lesion. In cases where the melanoma has spread out, patients are treated by

immunotherapy or chemotherapy. Most people are cured if spreading has not occurred. Therefore, the early and reliable recognition of melanomas at the early stage is of special importance [6]. The difference between a benign or malignant tumour is its invasive potential. If a tumour lacks the ability to invade adjacent tissues and to metastasize then it is benign, whereas a malignant tumour is invasive or metastatic. A nevus (birthmark) is a sharply circumscribed and benign chronic lesion of the skin. The melanocytic nevus results from benign proliferation of the dendritic melanocytes. Due to the pigment melanin, they are mostly brown. Nevus cells are related to the melanocytes, but they show a lack of the dendrites and are oval in shape. They are typically arranged in cell nests. The majority of acquired nevi appear during the childhood up to young adults (the first two decades of life). A melanocytic nevus present at birth is called a congenital nevus. They are rarely about one in every 100 newborns. Nevi are harmless. However, 25% of malignant melanomas arise from pre-existing nevi.

3. Confocal laser scanning microscopy

In conventional microscopy, the entire field of a tissue sample is simultaneously illuminated by light and displayed. Although the brightest light intensity results from the focal point of the objective lens, other parts of the tissue are still illuminated, resulting in a large unfocused background section. This background noise diminishes the image quality. Both conventional and confocal laser scanning microscopy (CLSM) can use reflected light to image a tissue sample. The reflected light from the illuminated spot is then re-collected by the objective lens. In addition to the reflected light from the focal point, the scattered light from sample points outside the focus light (coming from places above or below the focus) is projected by the optical system of the microscope and therefore contributes to the image assembly. This causes a blurring and obscuring of the resulting image. Confocal microscopy overcomes this problem by placing a pinhole in the conjugate focal plane (hence the designation confocal) that allows only the light emitting from the desired focal spot to pass through [7]. Any light outside of the focal plane (the scattered light) is blocked. **Figure 2** shows the principle: the out of focus light (red), coming from places above the selected focal plane, is blocked by the pinhole in the conjugate focal plane. The (in focus) light from focal plane (blue) can pass through the pinhole and is detected. Therefore, a blurring is avoided and sharp and detailed images are produced (in other words: the image information from multiple depths in the sample is not superimposed). In confocal microscopy, a light beam is directed by a dichroic mirror to the objective lens where it is focused into a small focal volume at a layer within the tissue sample (**Figure 3**). A laser, with a near-infrared wavelength, is used as a coherent monochromatic light source. The same microscope objective gathers the reflected light from the illuminated spot in the sample. The dichroic mirror separates the reflected light from the incident light and deflects it to the detector. Before the light reaches the detector, the out of focus sections are blocked by the pinhole in the conjugate focal plane. The in focus light that passes through the pinhole is measured.

http://dx.doi.org/10.5772/63404

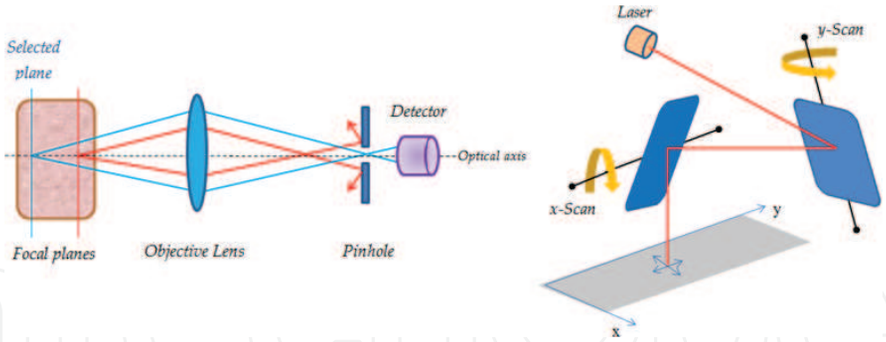

Figure 2. Principle of the confocal (left) and laser scanning (right) microscopy.

The detector, which is usually a photomultiplier tube or avalanche photodiode, amplifies and transforms the intensity of the reflected light signal into an electrical one that is recorded by a computer. In contrast to conventional microscopy, there is never a complete image of the sample at any given instant; rather only one point in the selected plane of the sample is observed. In order to create an image, light from every point in the plane (x-axis, y-axis) must be recorded. This can be done by a raster scanning mechanism which uses two motor driven high-speed oscillating mirrors, which pivot on mutually perpendicular axes. Coordination of the two mirrors, one scanning along the x-axis and the other on the y-axis, produces the rectilinear raster scan (**Figure 2**). During the scanning process, the detected signal is transferred to a computer that collects all the 'point images' of the sample and serially constructs the image pixel by pixel. The brightness of a resulting image pixel corresponds to the relative intensity of the reflected light. The contrast in the images results from variations in the refractive index of microstructures within the tissue. Information can be collected from different focal planes by raising or lowering the objective lens. Then successive planes make up a 'z-stack'. A stack

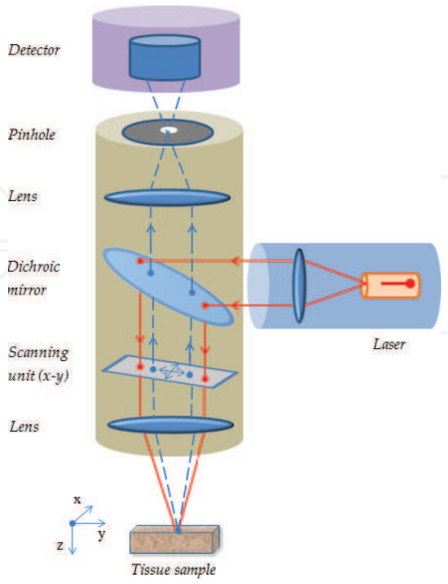

Figure 3. Principle of the confocal laser scanning microscope.

is a sequence of images captured at the same horizontal position (x- and y-axes) at different depths (z-axis). The images are taken enface (horizontally). The confocal laser scanning microscopy is performed with a Vivascope 1000 (Lucid Inc., USA) which uses a diode laser at 830 nm wavelength and a power of <35 mW at tissue level. A ×30 water-immersion objective lens with a numerical aperture of 0.9 is used with water as an immersion medium. The spatial resolution is 0.5–1.0 μm in the lateral and 3–5 μm in the axial dimension.

The images contain a field-of-view of 0.5 × 0.5 mm. Up to 16 layers per lesion can be scanned. All images, stored in BMP file format, are monochrome images with a spatial resolution of 640 × 480 pixels and a grey level resolution of 8 bits.

4. Interpretation of confocal laser scanning microscopic images

The reflectivity of the tissue depends on chemical structures. Melanin and melanosomes have a high refractive index which contributes strongly to the contrast of the resulting image [8–10]. Due to such dominating variations of the refractive index, only a certain part of the in falling light is reflected. This makes the appearance of the tissue in a CLSM image so different from conventional histological views. The power of the 830 nm laser limits the imaging depth to a maximum of 350 μm, corresponding to the papillary dermis (higher power could damage the skin). **Figure 4** shows the views of different skin layers [11]. The stratum corneum shows large polygonal anucleated corneocytes (A). Skin folds and marks appear as dark structures. The next layer is the stratus granulosum (B). The stratum spinosum (C) contains keratinocytes in a honeycomb pattern. In the stratum basalis (D), the basal cells are uniform in size and show higher reflections than spinous keratinocytes and appear very intensively. The dermatological guidelines for the interpretation of melanocytic skin lesions in CLSM views are as follows.

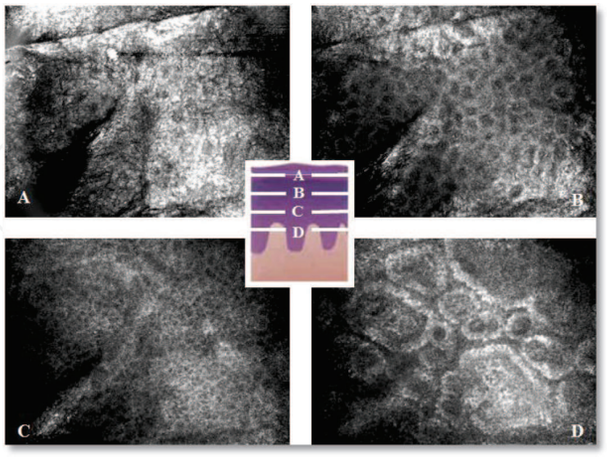

Figure 4. CLSM views of normal skin.

For the diagnosis of CLSM views of benign common nevi and malignant melanoma, architectural structures such as micro-anatomic structures; cell nests, etc., play an important role [12].

http://dx.doi.org/10.5772/63404

Monomorphic melanocytic cells, melanocytic cell nests and readily detected keratinocyte cell boarders are suggestive of benign nevi, whereas polymorphic melanocytic cells, disarray of melanocytic architecture and poorly defined keratinocyte cell borders are suggestive of melanoma (**Figure 5**). The images are taken from the centre of the tumours.

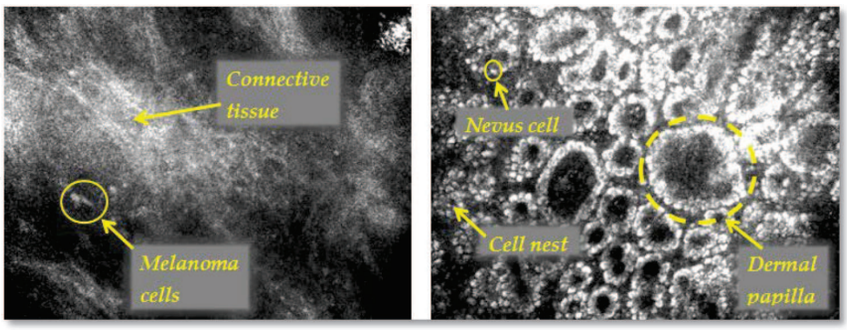

Figure 5. CLSM images of malignant melanoma (left) and common benign nevi (right).

Layers from the plane of the spinous keratinocytes (polygonal cells) to the plane of the basal cells (dermo-epidermal junction) are used for diagnosis.

5. Analysis of tissue structures at different scales

As shown in the previous section, the information at different scales (from coarse structures to details) plays a crucial role in the diagnosis of CLSM images of skin lesions. Wavelet analysis is a method to analyse visual data by taking into account scale information [13].

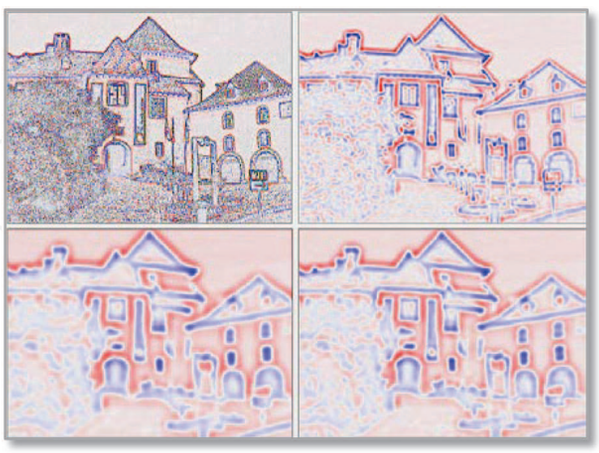

Figure 6. Scale-space sequence of a successively Laplacian of Gaussian-filtered image.

The multiple resolutions enable a scale invariant interpretation of an image. **Figure 6** illustrates the principle of scale space analysis for four levels of scale (clockwise direction). In the top left

image (scale 1), the feature detection responds to fine texture. The images at higher scales are generated by a Laplacian of Gaussian filter (LoG(x, y)), which is also known as Marr-Hildreth operator or Marr wavelet (**Figure 7**), whereby the kernel size (σ) of the Gaussian increases step by step.

$$\mathrm{LoG}(x,y) = \frac{1}{\pi\sigma^4}\left(\frac{x^2+y^2}{2\sigma^2} - 1\right)\cdot e^{-\frac{x^2+y^2}{2\sigma^2}}$$

The blue and red colours indicate positive and negative values. The images become increasingly blurred and smaller details (or regions) progressively disappear. The detected features are then associated with a larger scale scene structure. The multiresolution analysis is closely analogous to the human vision system which seems to prefer methods of analysis that run from coarse to fine and, repeating the same process, obtain new information at the end of each cycle [14] (**Figure 6** counter clockwise direction). The wavelet decomposition can be realized as a convolution of the image with a filter bank, consisting of high pass and low pass filters [15]. Whereby, for example a first-order derivative can be used as a convolution kernel for the high-pass filter and a moving average as a kernel for the low-pass filter. In our study, the filter coefficients are defined by the Daubechies 4 wavelet transform.

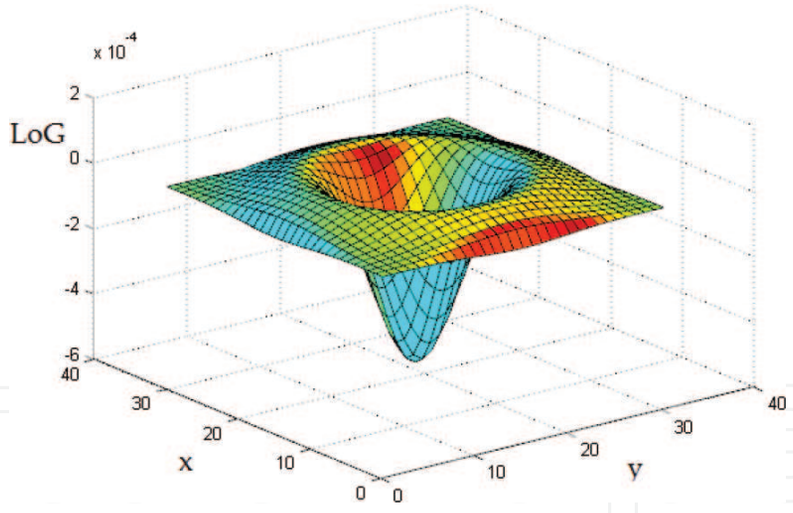

Figure 7. Shape of the Laplacian of Gaussian convolutional filter kernel.

The wavelet decomposition performs a multi resolution analysis, whereby the image is successively decomposed by the filter operations followed by sub-sampling. The (pyramidal) algorithm consists of several steps and operates as follows: at the beginning, the image rows are filtered by the high-pass filter and in parallel by the low-pass filter (**Figure 8**). From both operations result two images (which are called sub-bands), one shows details (high pass) and the other is smoothed out (low pass). The sub-sampling is done by removing every second column in both sub-bands. Subsequently, the columns of both sub-bands are high-pass and

http://dx.doi.org/10.5772/63404

independently low-pass filtered. This results in four sub-bands, which differ by the kind of filtering. Again a sub-sampling is done by removing every second row in each sub-band. This is the end of the first step. The mixed filtered (high-low pass, etc.) sub-bands are stored. Only the double low-passed sub-band is processed in the second step (**Figure 8**). The second step repeats the operations of the first step. Again this results in four sub-bands and the fourth smoothed sub-band is used as entry for the following step. At every step, the resulting sub-bands are reduced to half the resolution. The sub-bands with higher spatial resolution contain the detailed information (high pass), whereas the sub-bands with the low-resolution represent the large scale coarse information (low pass). The output of the wavelet decomposition consists of the remaining 'smooth-…-smooth' components and all the accumulated 'detail' components. In other words, via the wavelet decomposition, the image array is decomposed into several sub-bands representing information at different scales. The output of the last low-pass filtering is the mean gray level of the image.

After the dissection of the quadratic sub-bands, they are usually arranged in a quadratic configuration, whereby the three sub-bands of the first step fill 3/4 of the square, the three sub-bands of the second step fill 3/16 of the square, etc. The sub-bands representing successively decreasing scales are labelled with increasing indices (**Figure 9**). Then, the architectural structure information is accumulated along the way of the sub-bands (from coarse to fine). In image processing, it is convenient to display the smoothed image as lowest sub-band in the upper left corner of the quadratic sub-band configuration. The coefficients values in the different sub-bands reflect architectural and cell structures at different scales.

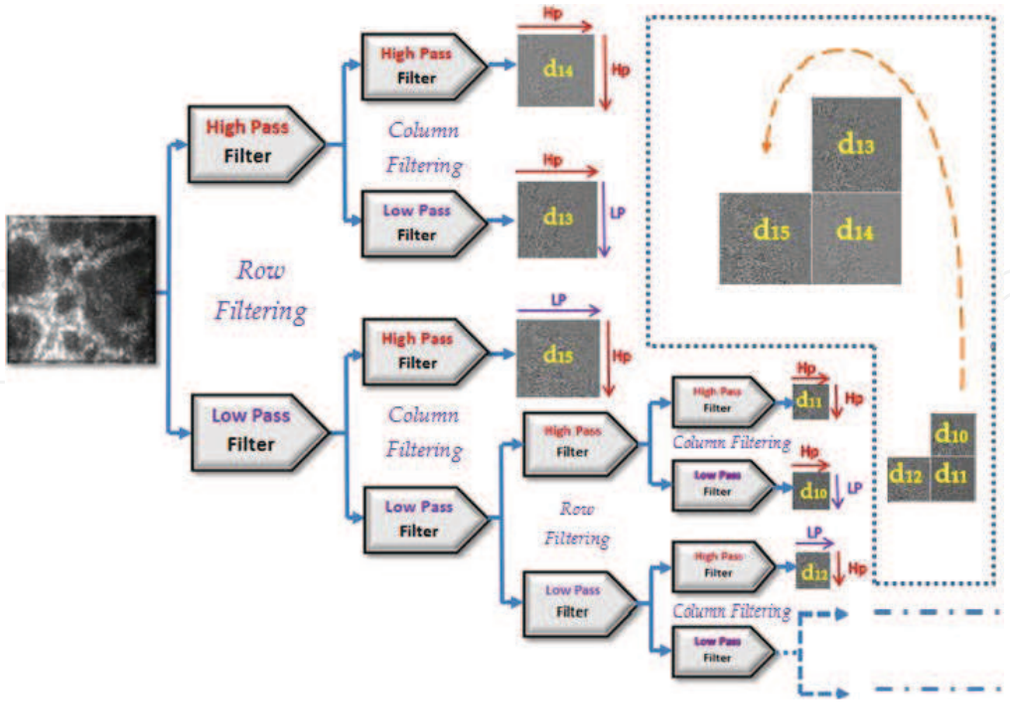

Figure 8. The multiresolution filter bank of the wavelet decomposition.

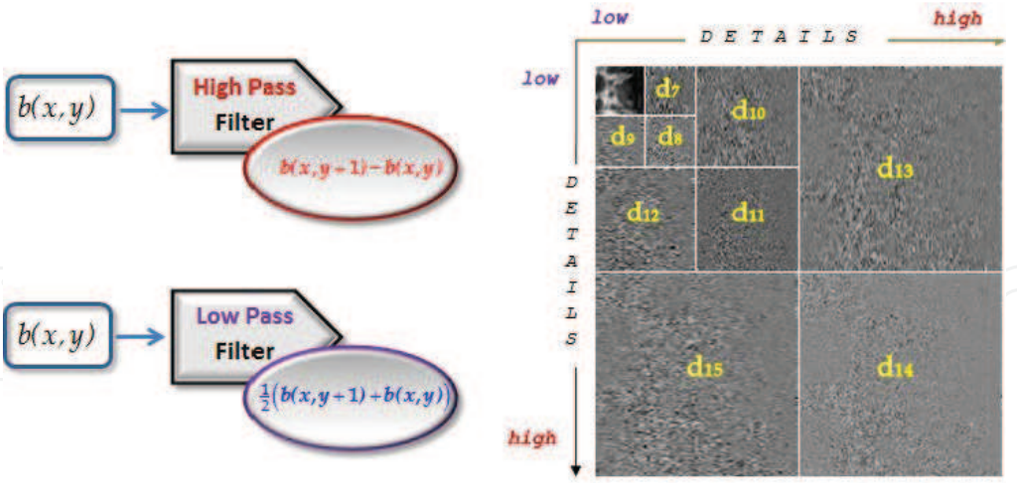

Figure 9. The sub-bands resulting from the successive high and low pass filter operations.

The tissue features are derived from statistical properties of the sub-band coefficients. For the ith sub-band of size N × N, the coefficients are given by:

$$d_i = \left\{ d_i(k,l) \middle| k,l = 1,N \right\}$$

The texture features are based on the variations of the coefficients within each sub-band and the weighted sum of all the coefficients into each sub-band. The standard deviations of the coefficients inside the single sub-bands and the energy and entropy of the different sub-bands are calculated and used as features (for details see: [16]). The standard deviation of the coefficients represents how exposed the tissue structures in the considered sub-band at the given scale are. The total energy of the coefficients in a given sub-band shows to what degree the structures at the corresponding scale contribute to the image. The distribution of the energy of the sub-bands is represented in a power spectrum, enabling an evaluation of their relative contributions.

The next task in automated image analysis is the use of machine-learning algorithms for classification purposes on hand of the feature values [17]. The algorithm learns, by use of a training set, how to assign the tissue images to given classes. Then, in future, the algorithm can apply the gained knowledge to predict the class of unknown tissue. By means of the classification procedure, the primary inhomogeneous set of CLSM samples, consisting of a mix of malignant melanoma and benign common nevi cases, is split into homogeneous subsets, which are assigned to one of the two tumour classes: common benign nevi or malignant melanoma. A homogeneous subset means that it contains only CLSM images with similar feature values, representing one specific kind of tissue. For the discrimination of the CLSM images, the CART (Classification and Regression Trees) algorithm is used [18].

The tree representation consists of different nodes and branches. There is a root node, several leaf (terminal) nodes and inner nodes (**Figure 10**). The first node in the tree is the root node. It

http://dx.doi.org/10.5772/63404

contains the feature values of the whole set of CLSM image samples. A leaf node is a homogeneous node which contains only samples belonging to the same class of tissue. The inner nodes contain more or less inhomogeneous sample sets. A branch in the decision tree involves the testing of one particular texture feature (binary tree). Then, the considered node, which is the parent node, is split into two child nodes (**Figure 10**).

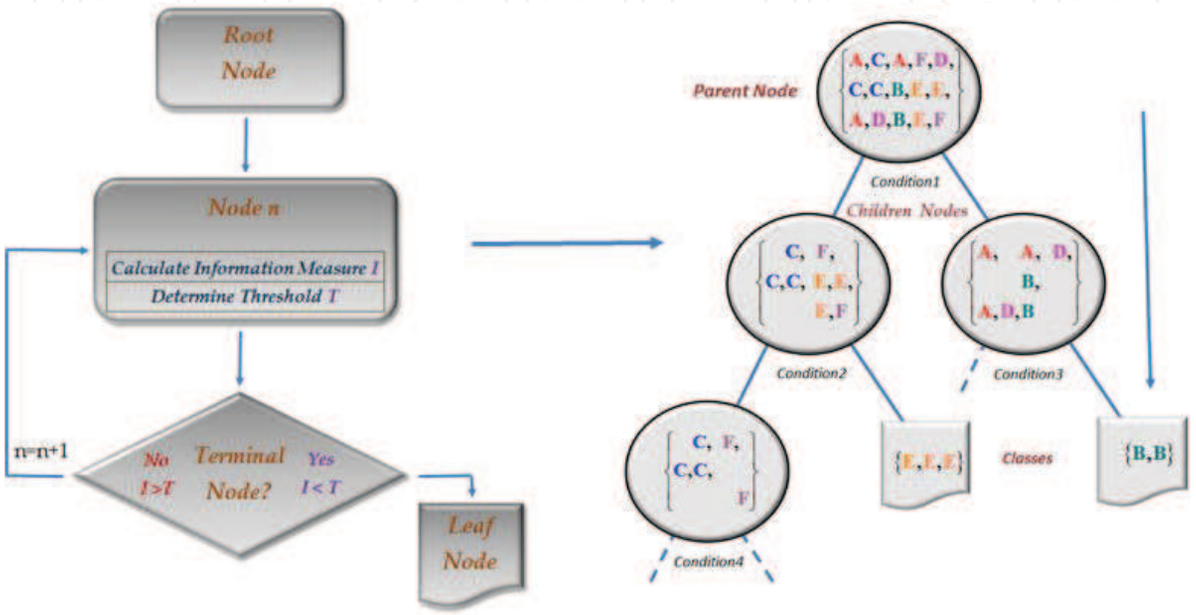

Figure 10. Generation of a decision tree.

The feature is tested by comparing its numerical value with a threshold value that divides the value range. The threshold value is selected automatically by the algorithm in such a way that the subsets of samples in the child nodes are purer than the set in the parent node. To this purpose, an information measure is used which indicates the degree of homogeneity; the value in the leaf nodes is zero and the higher the value of an inner node, the higher is its inhomogeneity. At every branch in the tree, subsets with smaller values of the information measure are generated. The decision tree is generated recursively (details are shown in: [16]). Whereby the algorithm consists in principal of three parts: the determination of the optimal splitting at every node; the decision whether the node is a leaf node or an inner node; the assignment of a leaf node to a specific class (**Figure 10**). To classify an unknown sample, it is routed down the tree according to the values of the different features. When a leaf node is reached, the sample is classified according to the class assigned to the leaf. The tree-based machine-learning algorithm captures the decision structure explicitly. That means the generated decision rules are 'Modus Ponens', with a precondition and conclusion part, and are intelligible in such a manner that they can be understood, discussed and used as diagnostic rules.

$$IF\left(...and.Condition1.and.Condition2\right)THEN\left(Class := A\right)$$

In total, 39 different features are calculated for 16 frequency bands (labelled from 0 to 15). The mean value is calculated from the first four frequency bands; therefore, 13 values result for each feature. The highest frequency bands contain only information about very fine grey level variations, such as noise, and are therefore not considered for the image analysis. The procedure for image analysis (including feature extraction and calculation) was developed with the 'Interactive Data Language' software tool IDL (IDL 7.1, ITT Visual Information Solutions). The tree classification is done by the CART analysis software from Salford Systems, San Diego, USA.

6. Biological motivation for neural networks

A neuron is an electrically excitable cell that receives, processes and transmits information as electrochemical signals. It consists of several dendrites, the soma and an axon (**Figure 11**). The soma is the cell body which contains the nucleus and all the necessary cytoplasmic cell structures. The dendrites are cytoplasmic extensions of the cell body with many branches allowing the cell to receive signals from other neurons. The axon is a special extension which carries signals away from the soma. At its terminal, the axon undergoes extensive branching, enabling communication with many target cells. The neurons maintain voltage gradients across their membranes. Ion channels, embedded in the membrane, enable the generation of intracellular-extracellular ion migrations. The resulting changes in the cross-membrane polarization generate an electrochemical pulse, known as the action potential. These changes in the cross-membrane potential are transferred as a wave of successive depolarization and repolarisation processes along the cell's axon. The axon terminal contains synapses, specialized connections to target neurons, where neurotransmitter chemicals are released. Synaptic signals may be excitatory or inhibitory. Once the pulse from the soma along the axon reaches the synapses, a neurotransmitter is released at the synaptic cleft. The neurotransmitter molecules bound at the receptors in the post-synaptic membrane (of the target neuron) and opens ion channels. Then, the electrochemical pulse is transmitted to the target neuron.

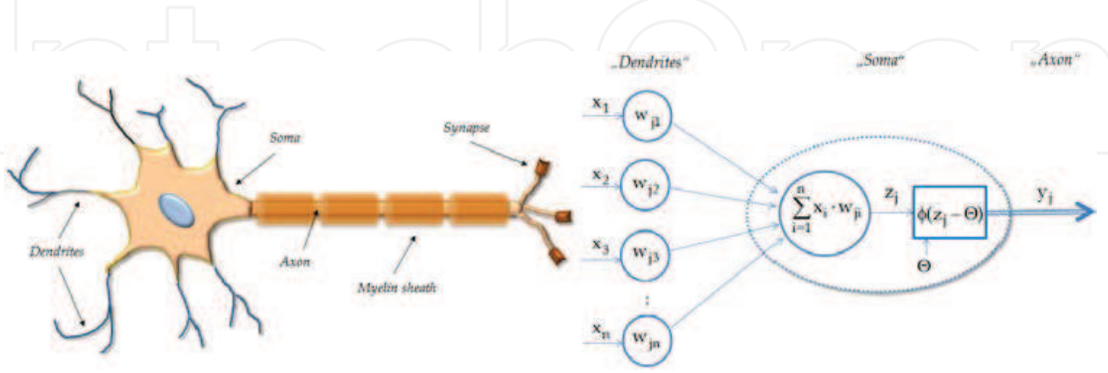

Figure 11. Microanatomy of a natural neuron (left), principle of an artificial neuron (right).

An artificial neuron is a mathematical model of a biological neuron. Artificial neurons mimic the behaviour of the biological neurons. The input of the artificial neuron is represented by a

vector:$x=(x_1, x_2, ..., x_n)$, whereby its dimension reflects the number of contributing dendrites (**Figure 11**). In the mathematical model, each 'dendrite' contributes individually through a weighted signal to the input signal. The weight factor $w_j=(w_{j1}, w_{j2}, ..., w_{jn})$ simulates the ratio of synaptic neurotransmitters, whereby positive values represent excitatory and negative values inhibitory behaviour (a weight value zero means that there is no connection between the involved neurons). The summation function represents the soma of the neuron j. The exciting and inhibiting signals are added in the function:

$$z_j = \sum_i x_i w_{ji}$$

The firing behaviour of the neuron is represented by the activation function. Its activation depends on the output of the summation function z_j and a threshold value Θ. If the summation function exceeds the threshold, the neuron is firing and transmits an output signal y_j:

$$y_j = \phi\left(z_j - \theta\right)$$

The biological motivation of the activation function is the threshold potential in natural neurons. Step and sigmoid functions are often used as transfer functions.

7. Artificial neural networks

Artificial neural networks consist of a number of artificial neurons, the computational units, which are interconnected. Each unit performs some small calculation based on inputs it receives from other units, whereby the associated weight factors can be tuned. This tuning occurs by allowing the network to analyse many examples of previously observed data. The most common type of neural network is the feed forward neural network (containing no loops), and in such networks, the computational units are organised into layers from an input layer, where data are fed into the network, to an output layer, where the result of the network's computation is outputted in the form of a classification result or regression result (**Figure 12**). Traditionally, each neuron in a layer is connected to all other neurons in the previous or subsequent layers (fully connected network). Between the output and input layers are hidden layers, and networks that consist of more than one hidden layer are known as *deep learning* algorithms. Such feed forward neural networks have been shown to be universal approximators, that is to say they can learn to approximate any continuous function to arbitrary precision, given enough hidden neurons [19]. Neural networks must be trained. The training data are previous observations that have been collected, and the task of the network is to learn a function which should map new input data to a classification label.

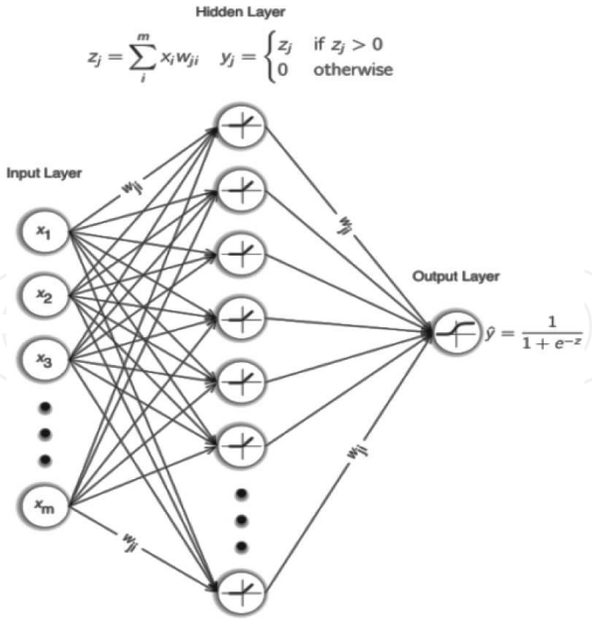

Figure 12. Structure of a feed forward artificial neural network.

In general, feed forward neural networks are supervised machine-learning algorithms. **Figure 12** shows a network with three layers: (1) an input layer, where data are fed in, (2) a hidden layer consisting of neurons that each contain an activation function that reads in data from the input neurons, performs some calculation, and outputs a value, and (3) an output layer that reads data from the hidden layer and makes a prediction based on this input. All connections between neurons have independently adjustable weights (Section 6). All layers are fully connected meaning that each neuron in the input layer is connected to every neuron in the hidden layer. The network learns by adjusting the weights between each of the connected neurons until the network makes good predictions by minimising an error function (backpropagation algorithm).

Fully connected neural networks are useful where individual features of a dataset are not very informative. In image data, where an individual pixel is not likely to be very informative taken on its own, a local combination of pixels may very well be informative and represent an object of interest. However, neural networks are also far more computationally intensive than many other machine-learning algorithms, with the number of tuneable parameters quickly growing into the millions as the network increases in depth or size. Also, neural networks typically work on image data directly, without feature reduction, meaning the dimensionality of the data being analysed by neural networks is much higher than that of other algorithms, which often work on extracted features. One could therefore summarise that neural networks are most useful for high m high n problems—problems where there exist many observations (n) of high dimensional data (m). Of late, neural networks algorithms have re-emerged as a popular technique in machine learning, especially in the field of image analysis. This re-emergence has come due to a number of recent developments in neural network design as well as independent hardware developments. In real-world applications, their usage has grown beyond image analysis and has also been shown to be useful for other tasks, such as natural

http://dx.doi.org/10.5772/63404

language processing and artificial intelligence [20, 21]. Nevertheless, a number of advancements in recent years resulted in an upsurge in the usage of neural networks.

First, hardware advancements have made it feasible for larger neural networks to be trained in reasonable amounts of time. As mentioned previously, neural networks that learn on very high-dimensional data require many neurons and layers, meaning networks can consist of many millions of parameters that need to be tuned. This results in large network architectures that have, for a long time, been unfeasibly difficult to train on standard desktop workstations. However, computational enhancements have meant this is no longer the case. These computational advancements are the result of rapid developments in graphics processing unit (GPU) technology due to the ever increasing requirements of the gaming industry, resulting in great improvements in the parallel processing power of GPUs. In 3D gaming, the vast majority of processing power is spent on matrix multiplications, such as transforms and perspective calculations, in order to depict the 3D worlds of games in 2D to the user. Such calculations are, for the most part, performed using matrix and vector multiplications. Such matrix calculations can be performed in parallel, and hence gaming GPUs have evolved to be particularly suited to such parallel processing tasks. To this end, GPUs typically consist of boards with many small, less powerful cores that can perform highly parallel computations. While CPUs tend to possess 2–4 large and fast cores, GPUs possess many hundreds of smaller cores. Crucially, almost 90% of the computational effort required to train a neural network is spent on vector, matrix, and tensor operations, meaning they can benefit from all the recent technological advancements in GPU technology. Indeed, with Moore's Law no longer holding, parallelised algorithms may, in future, be the only way to analyse very large data [22]. Second, empirical data have shown that neural networks with large numbers of hidden layers outperform many algorithms at several machine-learning tasks, especially in computer vision, object recognition and object detection. Deeper and deeper neural networks, with larger and larger numbers of neurons, have achieved human-level performance at very human-like tasks, such as playing video games [23] and playing the game of Go [24]. Deeper networks, however, contain more neurons, each of which needs to perform some calculation, and have its associated weight tuned, resulting in longer training times and larger memory requirements. Again, advances in hardware and optimisation techniques have meant that ever deeper networks are now trainable within reasonable timeframes [25]. Third, more and more data are permanently stored, archived and saved than ever before. This is especially true in fields such as medicine, where large amounts of data are accumulated during routine activities. In the past, these data might have been archived or stored in offline tape drives, or even discarded. However, this is no longer necessarily true as the cost per GB of storage has declined so rapidly, meaning easier access to more data and less likelihood of data being discarded. Deep learning algorithms require large amounts of data to train and access to very large datasets, and the ability for individuals to store large amounts of data has meant they are being applied to such problems more often.

Traditional feed forward neural networks consist of layers, where each neuron is connected to every other neuron in the layers above and below it. These are known as fully connected, or affine, layers. Fully connected neural networks do not consider the spatial relation between

pixels in an image. Pixels which are close together are treated exactly like pixels which are far apart when being processed by the network. For the learning of high-level features, this is suboptimal. In terms of image analysis, one particular type of neural network algorithm has stood out as being especially adept at image classification and object recognition. This is the convolutional neural network. The idea behind convolutional neural networks is to restrict the network to take inputs only from spatially nearby neurons. In other words, the layers are not fully connected, as in the example in **Figure 12**.

8. Convolutional neural networks

In the fields of image analysis, object detection and pattern recognition, convolutional neural networks are the state of the art algorithm for practical applications. Following on from our previous work, where we applied multiresolution analysis and CART as tree-based machine-learning method (Section 5), we decided to test the applicability of convolutional neural networks at a similar classification task. Because neural networks learn their own discriminatory, high-level features, the dataset requires no pre-processing or feature extraction, with the exception of image resizing and pixel value normalisation. This is in direct contrast to our previous efforts, where a dedicated feature extraction phase was necessary. Convolutional neural networks (CNN), in effect, emulate the way in which classical pattern recognition works, where local features (edges, corners, etc.) are extracted and combined to generate higher level representations that can be used for object recognition. Convolutional neural networks are locally connected, where each neuron is connected only to those that are spatially close (local receptive fields) in the previous layer, mimicking the visual cortex of some animals. Pixels that are closer to each other are more strongly correlated than those which are further away from each other, and this is something which the convolutional neural network has been designed to be able to account for through its architecture [26].

Network architectures with fully connected layers do not take into account the spatial structure of the images. Instead of using a network architecture which is tabula rasa, convolution neural networks (CNN) try to take advantage of spatial structures in images. They use three basic ideas: local receptive fields, shared weights and pooling. It is helpful to represent the input image as a square of neurons, whose values correspond to the pixel intensities. Then, only small, localized regions of the input image are connected to a neuron in the first hidden layer. Such a region in the input image is called the local receptive field for the corresponding hidden neuron. In other words, the hidden neuron learns to analyse its particular local receptive field. If the receptive field has a size of 5 × 5 pixels, then the hidden neuron is connected by 5 × 5 weights, which are adjusted during learning. The input of the hidden neuron is given by the summation function:

$$y_j = \sum_{l=0}^{4} \sum_{m=0}^{4} w_{l,m}^{j} b_{j+l,k+m}$$

http://dx.doi.org/10.5772/63404

The value $b_{x,y}$ denotes the input activation at position (x, y). The output of the hidden neuron is given by the activation function, for example the sigmoid function. The convolutional operation can be considered as a sliding window, which travels over the image, with the window centre moving one or more pixel a time. This is defined by the stride length. If the window is moved by one pixel, the stride length is 1. For each position of the local receptive field, there is a different hidden neuron in the first hidden layer. The map from the input layer to the hidden layer (convolutional layer) is called a feature map. The weights $w_{l,m}$ defining the feature map are the shared weights. The shared weights define the convolution kernel (convolution is generally the workhorse of image processing). The pixels in the local receptive field are multiplied element-wise with the kernel. Features maps are generated using only neurons which are spatially close to each other, known as spatial connectivity. Each feature map is defined by a specific set of shared weights enabling the network to detect different kinds of features (edges, corners, etc.). The CNN therefore learns objects related to their spatial structure. For image analysis purposes, more than one feature map are required. Therefore, a complete convolutional layer consists of several different feature maps. In addition to the convolutional layers, CNNs also contain pooling layers which usually follow immediately after the convolutional layers. Pooling layers simplify the information in the output from the convolutional layer by generating a condensed feature map (this removes the positional information of the features learned, meaning the learned features are position invariant). For example, each unit in the pooling layer may summarize a region of 2 × 2 neurons in the previous convolutional layer. Pooling is done for each feature map separately. The final layer in the convolutional network is a fully connected layer. This layer connects every neuron from the last pooling layer to every one of the output neurons.

A depiction of a typical 7-layer convolutional neural network can be seen in **Figure 13**. Images are read into the network in the input layer. From this input, a number of feature maps (4) are generated, which are subsampled in a max-pooling phase. Then, both phases are repeated once more, before connecting to a conventional fully connected layer which is finally connected to the output layer. CNNs often contain multiple fully connected layers before the final output layer, and modern CNNs can contain many convolution/max-pooling pairs.

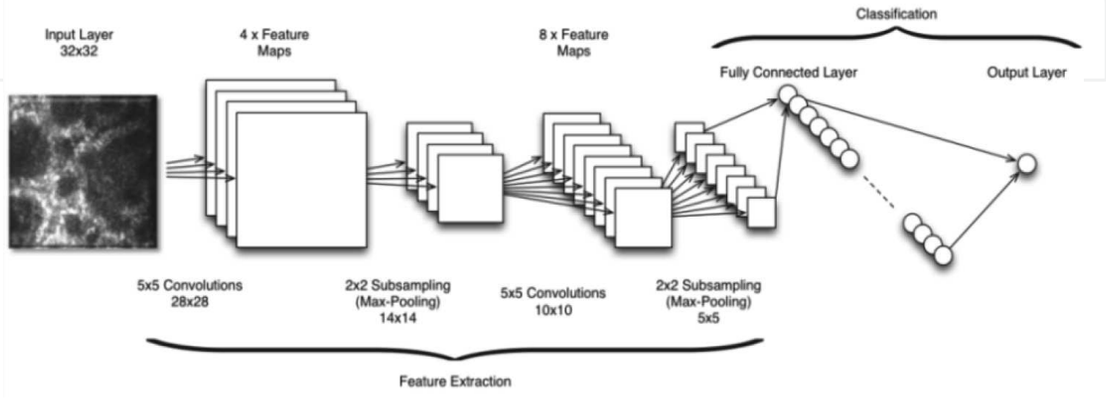

Figure 13. The structure of a typical seven-layer convolutional neural network.

Figure 14 describes the convolutional layer and max-pooling layer in more detail. The input into the convolutional neural network is a vector $\mathbf{x} \in \mathrm{R}^{1\times m}$, and the input layer has one neuron per feature. However, the layers can be thought as having their neurons arranged as depicted in **Figures 11** and **12**. In the case above, a 5 × 5 kernel is used, with a stride of 1, which results in a feature map of size 32 – 5 + 1 = 28 × 28. Typically, a convolutional layer is followed by a max-pooling layer, which acts as a type of sub-sampling, in this case halving the size of the previous feature map (**Figure 13**).

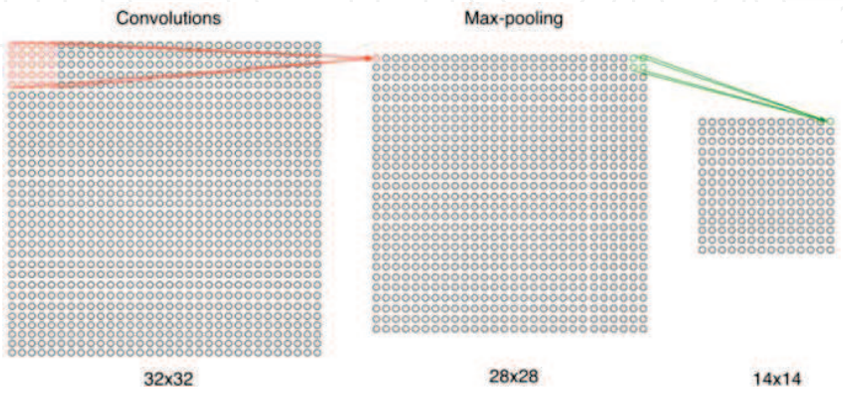

Figure 14. Principle of the convolutional layers and max-pooling layers [27].

Convolutional neural networks possess several characteristics that make them very suitable for the analysis of histological images. First, convolutional neural networks are capable of building models which are translation invariant and robust to transformations in the images, such as rotation, and they can learn features which are robust to scaling. They also generate models which are position invariant. This is especially important for microscopy imagery, where a lesion, for example, has no 'right way up', and cannot even be rotationally normalised.

9. Deep learning analysis of a CLSM image dataset

As stated previously, the goal was to train a model which would classify newly seen images as either malignant or benign. The neural network that was designed was based on the structure of the LeNet-5 convolutional neural network structure and was developed using the Keras deep learning library for Python [26]. The network consisted of a total of eight layers: the input layer, two pairs of convolutional and max-pooling pairs, two fully connected layers, and the output layer. The rectified linear unit (ReLU) was used throughout as the neuron nonlinearity. The ReLU is a computational unit which uses a ramp function [the rectifier f(x)=max(0, x)] and is currently the most popular activation function for deep neural networks. Because of the depth of network, a graphics processing unit (GPU) was used, which greatly increases the speed at which the network can train. In terms of hardware, a mid-range NVidia gaming GPU with 2 GB of dedicated video memory and 640 cores was used for training the network. The card is capable of 1306 GFLOP/s and has a memory band-

width of 86.4 GB/s. At the time of writing, the card can be purchased for under $150. The card was installed in a Linux workstation with 32 GB of RAM and a 3.5 GHz 6-core AMD processor running the Xubuntu 14.04 operating system. To illustrate the differences in computational power between a GPU and CPU, and to demonstrate the enormous impact using a GPU can have on training times, we benchmarked our code. Training the network over 20 epochs required 2 min 4 s of time, averaged over three runs, when using the GPU. When using the CPU, this time was 57 min 59 s for 20 epochs (also averaged over three runs), nearly 30 times slower. Experimenting with different parameters, or testing new network structures, can become very tedious when hours of computational power are required per run or experiment. The GPU reduces this time to minutes.

Dropout was used to control overfitting at two points in the network's structure: once after the convolutional and max-pooling pairs, and once again after the first fully connected layer. Dropout helps to control overfitting by randomly setting a certain set percentage of the neurons' weights to zero, effectively forcing the network to relearn those weights, with the intention of mitigating the learning of noise. The output of the network is finally determined by a sigmoid logistic function, squashing the results of the entire network to a value between 0 and 1. Values closer to 1 are therefore classified as being malignant, while values closer to 0 refer to a benign prediction. Such an output can also be used examine the network's confidence at a classification, with a value of 0.99 meaning a highly confident malignant prediction and a value of 0.51 representing an unconfident malignant prediction.

9.1. Input into the neural network

Images are read directly by the neural network. The only pre-processing which was performed was to resize the images from 640 × 480 to 64 × 64 pixels. Images are read by the neural network as a series of pixel values stored in a vector. Therefore, a single image is stored as a vector $\boldsymbol{x}$, so that one instance of an image $\boldsymbol{x}^{(i)} \in \mathbb{R}^{1\times m} = \left[x_1^{(i)} x_2^{(i)} x_3^{(i)} \dots x_m^{(i)}\right]$. The dataset consisted of $n = 6897$ images, each 64 × 64 pixels in size, representing a dimensionality $m = 4096$. The entire dataset is therefore stored in an $n \times m$ matrix:

$$\boldsymbol{X} \in \mathbb{R}^{n\times m} = \begin{bmatrix} x_1^{(1)} & \cdots & x_m^{(1)} \\ \vdots & \ddots & \vdots \\ x_1^{(n)} & \cdots & x_m^{(n)} \end{bmatrix}$$

To reduce the memory footprint, neural networks are typically trained using mini-batches, which are randomly selected subsets of $\boldsymbol{X}$. Targets, or labels, are stored in an n-dimensional column vector:

$$\boldsymbol{y} = \begin{bmatrix} y^{(i)} \\ \vdots \\ y^{(n)} \end{bmatrix} \big(y \in \{0, 1\} \mid 0 = \text{Benign}, 1 = \text{Malignant}\big).$$

Therefore, to input an image into a neural network, it must first be converted into a vector of pixel values. Each image vector's label is stored numerically in a separate target vector, $\boldsymbol{y}$. Once these have been prepared, a training matrix $\boldsymbol{X}_{\text{train}}$, a test matrix $\boldsymbol{X}_{\text{test}}$, and their corresponding target vectors $\boldsymbol{y}_{\text{train}}$ and $\boldsymbol{y}_{\text{test}}$ must also be generated.

9.2. Keras

Recently, a number of frameworks have been developed for deep learning, ranging from low-level, general purpose math expression compilers, such as Theano, to higher level frameworks such as Torch. For this analysis, the Keras framework was used. Keras is written in Python and is based on the Theano framework. It offers a high level control over network construction, abstracting the low-level Theano code, making it possible to design neural network structures in a layer-wise, modular fashion. Layers and functionality are added to the network piece by piece and are finally compiled into a complete network once the desired structure has been built. Users of Python can install Keras using pip, by typing pip install keras at the comment prompt. Keras has a number of requirements, including Theano (which can also be installed using pip install Theano at the command prompt). Briefly, once Keras has been correctly installed and successfully imported into the environment, a convolutional neural network is created by instantiating an object of the Sequential class, and then by adding layers to this object until the desired network is complete. For example, a convolutional layer can be added to the network using the add function: model.add(Convolution2D(…)). Configuring network properties, such as when to use dropout or specifying which activation function should be used, is also performed using the add function of the model object. The network is built in this way until the desired structure has been defined, and is then compiled using the model object's compile function. As Keras is based on Theano, the model is generated into Theano code, which itself is compiled into CUDA C++ code, and subsequently run on the GPU. Upon successful compilation the model, it can be trained on a dataset using the fit function, which takes the training data set as one of its parameters. A trained model can then be tested using the held back test data, using the trained model's evaluate function. Full Python source code for the generation of the model can be found in this book chapter's GitHub repository under https://github.com/mdbloice/CLSM-classification. This source file contains a complete implementation of the network, including the generation of all the plots and figures shown in the Section 10.

10. Results

10.1. Multiresolution analysis

Overall, 857 images of benign common nevi (408 images) and malignant melanoma (449 images) were used as study set [29]. To get more insights into the classification performance, a percentage split was performed by using 66% of the dataset for training and the remaining instances (34%) as the test set (**Table 1**). The classification results of 572 cases (276 benign

http://dx.doi.org/10.5772/63404

common nevi, 296 malignant melanomas) in the training set and 285 cases (132 benign common nevi, 153 malignant melanomas) in the test set.

CART	Training set			Test set		
	% Correct	Benign	Malignant	% Correct	Benign	Malignant
Benign	96.6	267	9	78.0	103	29
Malignant	98.0	6	290	84.1	24	129

Table 1. Classification results for features based on multiresolution analysis.

The CART classification shows a correct mean classification of 97.3% samples in the training set and a correct mean classification rate of 81.1% in the test set. In this study, the images were resized to 512 × 512 pixels. To illustrate the differences in the wavelet sub-bands of both tissues, the spectra of the wavelet coefficient standard deviations are shown for typical views of benign common nevi and malignant melanoma (**Figure 15**). The image of benign common nevi show pronounced architectural structures (so called tumour nests), whereas the image of malign melanoma show melanoma cells and connective tissue with few or no architectural structures. These visual findings are reflected by the wavelet coefficients inside the different sub-bands. The standard deviations of the wavelet coefficients in the lower and medium frequency bands (4–10) show higher values for the benign common nevi than for malignant melanoma tissue, indicating more pronounced structures at different orders of magnitude. The tissue of malignant melanoma appears more homogeneous (due to a loss of structure), and the cells are larger as in the case of benign common nevi. The standard deviations in the sub-bands with higher indices (representing finer and more pronounced structures) are lower than in the case of benign common nevi.

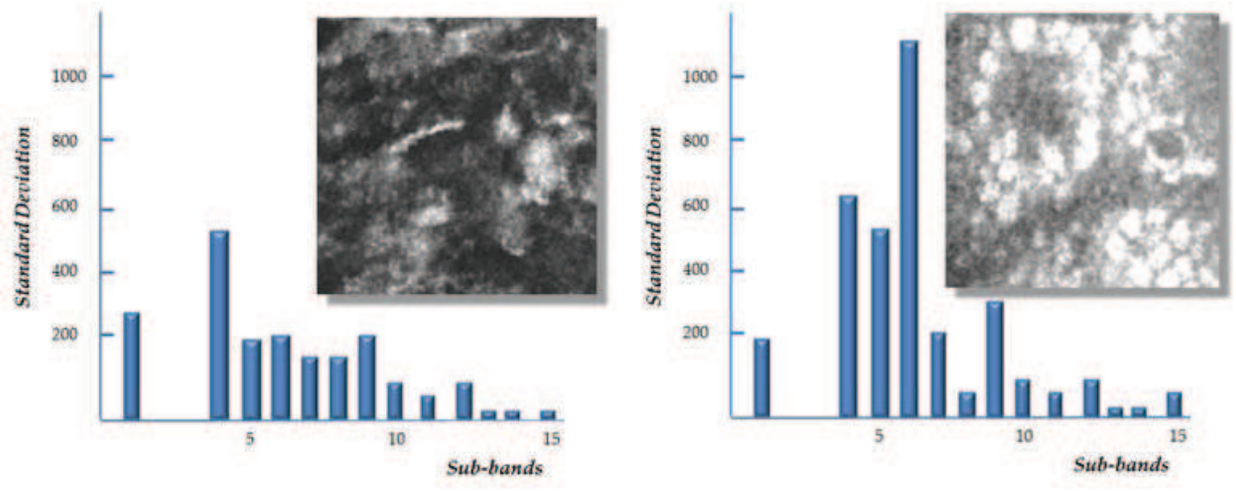

Figure 15. Sub-band spectra for benign common nevi (right) and malignant melanoma (left).

The analysis of the classification tree shows that seven classification nodes indicate benign common nevi and six nodes malignant melanoma. The visual examination of the selected nodes demonstrates characteristic monomorphic melanocytic cells and melanocytic cell nests for benign common nevi [28, 29]. Contrary polymorphic melanocytic cells, a disarray of melanocytic architecture and poorly defined or absent keratinocytic cell borders are characteristic for malignant melanomas.

10.2. Convolutional deep learning neural network

For this study, a dataset consisting of 6897 CLSM images of skin lesions was obtained from our university hospital. The dataset consisted of images of skin lesions in layers of various depths. Before training, the images were randomised and placed into a training set and test set, with the training set consisting of 5000 images and the test set consisting of 1897 images (**Table 2**). It is important to note that, in the case of this project, each image was treated individually, and not treated as belonging to one particular patient or even lesion. The test set, therefore, contained different layers or lesions from potentially the same patient as the training set, as a single patient may have had several scans or may have been examined on multiple occasions.

	Full Dataset	Training Set	Test Set
Total	6,897	5,000	1,897
Benign	3,607	2,655	952
Malignant	3,290	2,345	945

Table 2. The distribution of the classes in the whole dataset and in the training and test set.

Class imbalance occurs when a training set has far more samples of one particular class than another. For example, a small class imbalance existed in the dataset analysed in this chapter, with the samples of benign nevi slightly outnumbering the samples of malignant melanoma (there existed 317 more samples of the former compared to the latter). There are a number of techniques which can be employed to address class imbalance, such as data augmentation (generating synthetic data from your original dataset) or simply by discarding samples to better balance the dataset. In the case of our dataset, class imbalance was not at the degree as to make it problematic. When the training set and test sets were split, however, we ensured that the test set was largely balanced. Class imbalance can also affect how results, such as accuracy and precision/recall, should be perceived when analyzing a trained model on a highly imbalanced test set.

The network, after training for 20 epochs, achieved 93% accuracy on the unseen test set. The model's accuracy on the test set during training, as well as the model's error rate on the training set through each of the 20 epochs is shown in **Figure 16**.

http://dx.doi.org/10.5772/63404

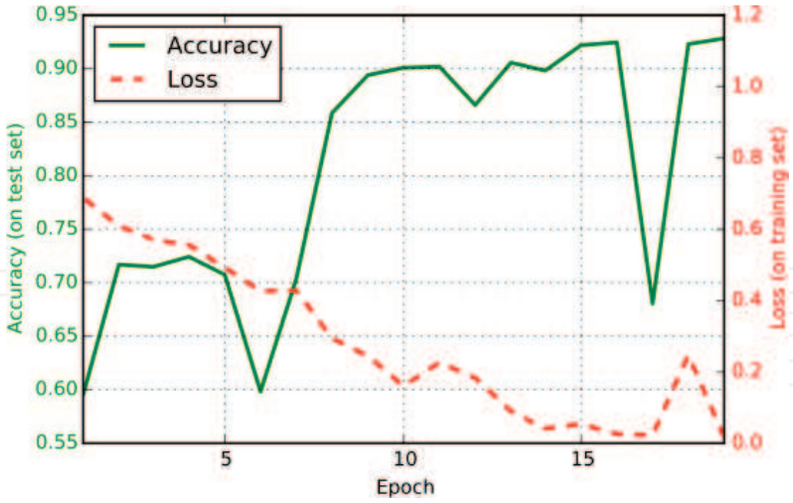

Figure 16. The model's accuracy on the test set and its logistic loss against the training set.

Loss on the training set eventually reduces to almost 0 (meaning it is at this point overfitting heavily), while the accuracy of the model on the unseen test set fluctuates but is tending towards an accuracy of approximately 90%. The accuracy of the final model after epoch 20, when training was terminated, was 93%. A confusion matrix, shown in **Figure 17**, describes the model's accuracy on the test set, in terms of absolute numbers of predicted and actual labels for both the benign and malignant classes.

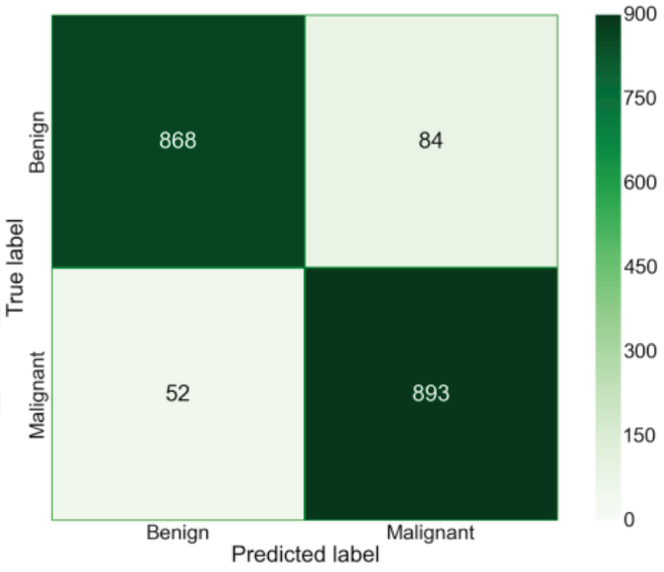

Figure 17. Confusion matrix.

Here, all true/false positives and true/false negatives can be seen. From these values, the precision, recall (sensitivity), and F_1 score (a weighted average of the precision and recall, given by $F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$) were calculated, as shown in **Table 3**.

	Precision	Recall (Sensitivity)	F_1 score	Support
Benign	0.94	0.91	0.93	952
Malignant	0.91	0.94	0.93	945
Avg/total	**0.93**	**0.93**	**0.93**	**1897**

Table 3. The generated model's precision, recall and F_1 score measured against the test set.

Table 4 describes the results of the model in absolute terms, with results for the model's predicted labels for both classes versus the actual labels for each class. As well as this, the total number of actual and predicted labels is shown.

	Actual		
Predicted	**Benign**	**Malignant**	**Total**
Benign	**868**	84	952
Malignant	52	**893**	945
All	920	977	1897

Table 4. The generated model's predicted labels versus the actual labels, measured on the test set.

10.3. Transfer learning

Transfer learning is a term that can be applied to several aspects of machine learning. In the case of neural network-based machine-learning approaches, transfer learning often refers to the act of using a pre-trained network as the starting point for a learning procedure, rather than starting with a network which has been initialized with random weights. This is often performed as a time-saving measure, but can also be done when the new data to be classified is scarce. Also, it can be performed only when the data used for pre-training is similar to the new data which should be classified. Furthermore, it constrains the practitioner into using a network which has the same architecture of the pre-trained model. Therefore, it is not useable in all situations, and it does not make sense to use, say, a network pre-trained on the ImageNet dataset (a commonly used benchmarking dataset, containing millions of samples of 1000 classes of images) in the context of CLSM lesion classification.

However, there exist several types of laser scanner-based approaches to skin lesion analysis, where the use of transfer learning may be beneficial. Other methods in the field include two photon excitation fluorescence microscopy, second harmonic imaging microscopy, fluorescence-lifetime imaging microscopy and coherent anti-stokes Raman microscopy. Whether or not transfer learning could indeed be implemented in this context would depend entirely on how well the features learned during pre-training match the features that exist in the new data (in other words, whether the learned features transfer well from one domain to the other). For example, several new methods produce colour images, which would mean the features learned in the analysis described here would likely not transfer well to this new domain (of course, colour images could be converted to greyscale). However, it is conceivable that other technol-

ogies, that also produce greyscale images, could make use of a pre-trained network, and thus benefit from pre-trained weight initialisation and therefore transfer learning.

The machine-learning community often makes available pre-trained networks for others to use, such as in the Model Zoo (https://github.com/BVLC/caffe/wiki/Model-Zoo). Some of the networks available on the Model Zoo took many weeks to train on powerful hardware, and is considered a very useful resource by many who do not have the time or the computational resources available to them for such an involved learning task. Of course, a pre-trained network could be made available for the CLSM or skin lesion analysis community, if the network was trained on a sufficiently large dataset and if indeed the learned features would transfer well to other domains.

11. Discussion

Confocal laser scanning microscopy is a technique for obtaining high-resolution optical images with depth selectivity. It enables the noninvasive examination of skin cancer in real-time. This makes CLSM very suitable for screening and early recognition of skin tumours, which augment the success of the therapy. The training of pathologists to acquire and refine their visual diagnostic skills is very time-consuming. To implement diagnostic capabilities on a computer, it is of considerable interest to understand how the diagnostic process unfolds and which texture features are critical for a successful diagnosis. For medical diagnosis, it is important to duplicate the automated diagnostic process.

The multiresolution approach with wavelets features mimics the diagnostic guidelines of the dermatopathologist, as they use multiscale features for the examination of CLSM views. The decision rules generated by machine-learning algorithms, such as CART, represent explicit knowledge that can be used to analyse and refine the diagnostic process. The generated rules can be implemented in viewer software which enables a visual evaluation of the diagnostic performance by the dermatologist. This can be used as a training aid for ongoing dermatologists in education. As shown in the Section 10, the algorithm performance allows a correct classification of 78.0% of the benign common nevi cases and 84.1% of the malignant melanoma in the test set. In contrast, sensitivity and specificity of 85.5 and 80.1% are reached by the human observer (overall performance 82.8%).

Although the CART algorithm discriminates the training set automatically (unsupervised), the feature extraction algorithm is predefined. Algorithms based on artificial neural networks do not perform or require hand-defined analyses of the image features with predefined (filtering) methods. Instead, they use neural computation inspired by the visual system of mammals. Neural networks process an image by use of a hierarchical processing architecture which mimics the way the visual cortex processes visual stimuli from the primary cortex (V1) to different layers (V2–V8) which are selective for different components of the visual stimuli such as orientation, colour, size, depth and motion. Neural networks are well suited for detecting similarities in images. However, the distributed representation of the acquired knowledge complicates the extraction of the diagnostic information. They deliver nothing

about the inference mechanism leading to a classification in a form that is easy readable for the human observer. Nevertheless, we can demonstrate a real example as to why artificial neural networks will play an ever more important role in automated medical diagnostic systems. A recent work reported that pigeons (columba livia) proved to have a remarkable ability at discriminating benign from malignant human breast histopathology images and at detecting cancer relevant micro calcifications in mammogram images after differential training with food reinforcement [30]. The discrimination was done by the pigeons via two distinctively coloured response buttons. For a correct discrimination, food was immediately provided by a dispenser. The pigeons proved not only to be capable of image memorization but were able to extend the learned skills to novel tissue images. It results that their diagnostic skills are like that of trained humans. It should be noted that the capabilities were acquired without the benefit of verbal instructions as in the case with human education. The low-level vision capabilities of pigeons appear to be equivalent to those in humans; feedforward and hierarchical processing seem to dominate. It can be assumed that pigeons do not explicitly analyse the images with predefined criteria and explicit instructions as humans do. The reinforcement training of the pigeons resembles the training of artificial neural networks. Given the high diagnostic accuracy of the pigeons they may serve as a model for the development and amelioration of artificial networks (or vice versa). We still do not know in detail how pigeons differentiate such complex visual stimuli but colour, size, shape, texture, and configurational cues seem to participate. Their visual discrimination performance may guide the basic research in artificial neural networks in order to develop computer-assisted image diagnostic systems. Experienced dermatopathologists reported that a beginner (a person in education) examines the CLSM views strictly according to the dermatological guidelines (Section 4), as the computers do by multiresolution analysis. Based on the large amount of previously viewed specimens, an experienced person reports the CLSM views more by its visual appearance (personal communication). This is similar to the image analysis performed by a trained neural network. The receptive field of a sensory neuron is a particular region in the visual system in which a stimulus will trigger the firing of that neuron. In vision research, it is known that a cat's visual cortex only develops its receptive fields if it receives visual stimuli in the first months of life [31]. The receptive fields in the primary visual cortex can be thought as 'feature detectors' or 'flexible categorizers'. This means that they learn the structure of the input patterns and become sensitive to combinations that are frequently repeated [14]. This also demonstrates the importance of convolutional neural networks in image processing and analysis.

In this work, and given the relatively small dataset size, the performance of the trained neural network model is encouraging. However, the results must be considered as a proof of concept, and not a model that could be used in a clinical setting, despite the good accuracy of the trained model. For example, the images were collected from a single department, at one hospital in a single region in Austria. To judge the potential real-world accuracy of a trained model would require a far larger dataset, collected from several regions worldwide, and carefully curated to ensure no unintentional bias is introduced (by only collecting data from patients of a certain age range, for example). By training a model on a far larger dataset such a model could be used in real-world clinical settings as a diagnosis aid.

The work here shows that deep layer neural networks have the capacity to learn the high-level discriminatory features required to classify malignant and benign skin lesions. This can be achieved without any dedicated feature engineering phase, data pre-processing or a priori domain knowledge. In the case of the CLSM image classification task presented here, all that was required was a labelled dataset of previous observations. However, what is also true is that neural networks require far more training data than traditional machine vision methods that work on extracted features. This is due to the very high dimensionality of the data, which in our case was $\mathbb{R}^{4096}$, in contrast to the analysis of the extracted features where the dimensionality was $\mathbb{R}^{39}$. To compensate for a far higher dimensionality, a much larger dataset is, therefore, a necessity. In other words, deep learning neural networks are most suitable for situations where you encounter data with 'high m, high n' properties—high dimensional data, like images, of which many samples exist—such datasets are common in the medical domain, meaning deep learning should be of especial interest to researchers in the area of healthcare informatics.

As parallelized hardware advances, Moore's law begins to plateau, and the amounts of data being stored increases, algorithms that take advantage of this perfect storm will become more and more relevant. We have shown in this chapter that classical approaches to image classification can indeed be emulated by deep neural networks fed with large amounts of observed data. In fields such as medicine, where data are in such abundance, highly parallelized algorithms may be the only approach that can deal with such large data sources in a meaningful way. Fortunately, this is no longer the domain of specialized research institutes with access to cluster computing: such algorithms are trainable without large investments in hardware and can be performed on a standard desktop workstation equipped with a modestly priced GPU.

Author details

Marco Wiltgen* and Marcus Bloice

*Address all correspondence to: marco.wiltgen@medunigraz.at

Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Graz, Austria

References

[1] Schaller A, Sattler E, Burghof W, Röken M: Color Atlas of Dermatology. 1te ed. Thieme Verlag. Stuttgart, New York. 2012; 978–3131323415

[2] Sterry W, Paus R: Thieme Clinical Companions Dermatology. Thieme Verlag. Stuttgart, New York. 2006 3–13–1359110

[3] Bolognia J.L, Jorrizo J.L, Schaffer J.V: Dermatology: Expert Consult Premium Edition. Saunders. 3 edition UK. 2012; 978–0723435716

[4] Markovic S.N, Erickson L.A, Flotte T.J, Kottschade L.A. Metastatic malignant melanoma. G Ital Dermatol Venereol. 2009;144(1):1–26.

[5] Oliveria S, Saraiya M, Geller A, Heneghan M, Jorgensen C. Sun exposure and risk of melanoma. Arch Dis Child. 2006;1(2):131–138.

[6] Friedman R, Rigel D, Kopf A. Early detection of malignant melanoma: the role of physician examination and self-examination of the skin. CA Cancer J Clin. 1985;35(3): 130–151.

[7] Pawley J.B. Handbook of Biological Confocal Microscopy. 3rd ed.. Springer, Berlin; 2006. 0-387-25921-X.

[8] Paoli J, Smedh M, Ericson M.B. Multiphoton laser scanning microscopy—a novel diagnostic method for superficial skin cancers. Semin Cutan Med Surg. 2009;28(3):190–195.

[9] Patel D.V, McGhee C.N. Contemporary in vivo confocal microscopy of the living human cornea using white light and laser scanning techniques: a major review. Clin Exp Ophthalmol. 2007;35(1):71–88.

[10] Rajadhyaksha M. Confocal microscopy of skin cancers: translational advances toward clinical utility. Conf Proc IEEE Eng Med Biol Soc. 2009;1:3231–3233.

[11] Hofmann-Wellenhof R, Pellacani G, Malvehy J, Soyer H.P. (eds). Reflectance Confocal Microscopy for Skin Diseases. Springer, Berlin Heidelberg; 2012; 978-3-642-21996-2.

[12] Pellacani G, Cesinaro A.M, Seidenari S. In vivo assessment of melanocytic nests in nevi and melanomas by reflectance confocal microscopy. Mod Pathol. 2005;18:469–474.

[13] Prasad L, Iyengar S.S. Wavelet analysis with applications to image processing. CRC Press, Boca Raton; 1997.

[14] Marr D. Vision. W.H. Freeman, New York, 1982.

[15] Strang G, Nguyen T. Wavelets and Filterbanks. Wellesley-Cambridge Press. MA USA; 1996.

[16] Wiltgen M. Confocal laser scanning microscopy in dermatology: Manual and automated diagnosis of skin tumours. In: Chau-Chang Wang editor. Laser Scanning, Theory and Applications, Intech Publisher. Croatia. 2011;133–170. ISBN 978–953–307–205–0)

[17] Murphy, K.P. Machine learning: a probabilistic perspective. MIT press, USA. 2012.

[18] Breiman L, Friedman J, Olshen R.A, Stone C.F. Classification and Regression Trees. Chapman & Hall, New York, London; 1993.

[19] Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. Neural Netw. 1989;2(5):359–366.

[20] Kalchbrenner N, Grefenstette E, Blunsom P. A convolutional neural network for modelling sentences. *arXiv preprint arXiv:1404.2188*, 2015.

[21] Collobert R, Weston J, Bottou L, Karlen M, Kavukcuoglu K, Kuksa P. Natural language processing (almost) from scratch. J Mach Learn Res. 2011;12:2493–537.

[22] Waldrop M.M. The chips are down for Moore's law. Nature. 2016;530:144–147.

[23] Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, Graves A, Riedmiller M, Fidjeland AK, Ostrovski G, Petersen S. Human-level control through deep reinforcement learning. Nature. 2015;518(7540):529–33.

[24] Silver D, Huang A, Maddison CJ, Guez A, Sifre L, van den Driessche G, Schrittwieser J, Antonoglou I, Panneershelvam V, Lanctot M, Dieleman S. Mastering the game of Go with deep neural networks and tree search. Nature. 2016;529(7587):484–9.

[25] Williams DR, Hinton GE. Learning representations by back-propagating errors. Nature. 1986;323:533–6.

[26] LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proc IEEE. 1998;86(11):2278–324.

[27] Michael A. Nielsen, Neural Networks and Deep Learning, Determination Press, USA. 2015. URL: http://neuralnetworksanddeeplearning.com/

[28] Lorber A, Wiltgen M, Hofmann-Wellenhof R, Koller S, Weger W, Ahlgrimm-Siess V, Smolle J, Gerger A. Correlation of image analysis features and visual morphology in melanocytic skin tumours using in vivo confocal laser scanning microscopy. Skin Res Technol. 2009;15:237–241.

[29] Wiltgen M, Gerger A, Wagner C, Smolle J. Automatic identification of diagnostic significant regions in confocal laser scanning microscopy of melanocytic skin tumours. Methods Inf Med. 2008;47:14–25.

[30] Leveson R.M, Krupinski E.A, Navarro V.M, Wasserman A. Pigeons (Columba livia) as trainable observers of pathology and radiology breast cancer images. Plos One 10(11). 2015; 1–21.

[31] Hubel DH, Wiesel TN. Period of susceptibility to physiological effects of unilateral eye closure in kittens. J Physiol. 1970;206(2):419–436.

IntechOpen