# Deep Learning in Histopathology

Research Paper Business Analytics

**Bram Rodenburg**

**Supervisor:** Ted Meeds

June 2016

VU University Amsterdam
Faculty of Sciences
Business Analytics
De Boelelaan 1081a
1081 HV Amsterdam
The Netherlands

## Preface

This research paper was written as part of the Business Analytics master program at the VU University Amsterdam. The research paper (6 EC) is a required component within the program in which the student should address a topic of choice that incorporates the business-related aspect of the program as well as the computer-science and mathematics aspects. The production of the paper is planned at the end of the second semester (June) of the master.

In this paper I address how deep learning, a field of study within computer science, could assist (histo)pathologists with some of their tasks. Tasks that are often time consuming and tedious, and known for being exposed to inter-observer variability. My goal throughout this paper is twofold. On one hand, I hope to show how deep learning can assist pathologists in their work. On the other, I like to show practitioners of machine learning some problems within histopathology that are currently addressed with deep learning techniques.


Bram Rodenburg
Amsterdam, June 2016

## Management Summary

In clinical practice a task of the pathologist is to analyze human tissue for disease by microscopic examination. However, many of these tasks are exposed to inter-observer variability and are time consuming (therefore costly). Clearly, automating these tasks has two main advantages. First, the quality of diagnoses could be improved since an automated approach could give more consistent and accurate outcomes. This in turn can result in a better assessment of a patient's treatment. Second, time and costs of these tasks can be reduced since manually performed tasks by a pathologist are now automated.

The automation of tasks within pathology is now possible since in the recent years pathology labs started to move towards more digital workflows. In these approaches microscopes are substituted by scanners, resulting in the availability of digital tissue images (called whole slides). As a result, it is now possible to apply image analysis techniques - techniques that can perform some of the tasks a pathologist would normally need to do manually.

A technique for image analysis that has proven to work extremely effective in practice is called deep learning. Deep learning is a class of methods that can automatically discover representations from raw data. One such a method is the convolutional neural network (CNN). Namely, CNNs have shown to work particularly well on image detection and classification tasks. For this paper, a literature study was conducted to demonstrate how deep learning is applied to certain tasks in histopathology.

One task in histopathology is mitosis counting. Namely, the mitotic count can be used as a factor to grade the severity of breast cancer. For this task, a pathologist needs to select several regions of interest within a whole-slide and count the mitotic figures in these regions. Several deep learning techniques have been researched for the actual mitosis counting, outperforming all other non-deep learning methods. However, selecting the regions of interest has not been addressed yet in a deep learning setting.

Another task addressed with deep learning is that of gland segmentation. Namely, to determine the severity of colon cancer, a pathologists requires to obtain several statistics of the glands in the colon to make a correct diagnosis. To automatically obtain these statistics an automated approach must first need to segment the glands and determine whether the gland is malignant or benign. CNNs have shown to be able to perform these tasks also quite well.

Two other tasks addressed with deep learning are glioma grading and tissue segmentation. For the glioma grading, a deep learning approach was capable of accurately determining the grade of a glioma. For the second task, a CNN was used to segment regions in a whole slide into epithelial and stromal regions. Although this task is not directly related to grading a disease, it can be viewed as an intermediate step that could be required for such a task.

To bring automated tasks such as the previous mentioned into clinical practice at least three obstacles need to be addressed. First, more data needs to be made publicly available so that other tasks than the ones described in this paper can be addressed as well using automated methods. Second, bringing au-

tomated methods into clinical practice requires regulatory approval. Obtaining this approval can be a costly and time consuming process. Third, training deep learning models requires a sufficient amount of data. However, this data can often not leave the hospital due to regulations and privacy concerns. Therefore, methods should need to be developed that enable the training of deep learning models that can be shared between hospitals while respecting the privacy concerns.

# Contents

## Introduction

One approach to study a disease is by microscopic examination of human tissue. Often, this tissue is obtained by means of a biopsy. After a biopsy is performed in an operating room, the tissue is send to a pathology lab. In this lab a pathologist prepares the tissue and then analyzes it for a specific disease. The field concerned with analyzing human tissue for a particular decease is called histopathology.

Common practice in pathology labs is to analyze the tissue using microscopes. However, in the recent years pathology labs started to move towards digital workflows [1]. Instead of using a microscope, a scanner is used that processes the tissue and produces a digital image of the tissue, called a whole-slide image. As a result, it is now possible to apply image analysis techniques to these images - techniques that could aid pathologists in solving several existing problems.

An example of a problem that could be addressed using image analysis techniques is mitosis counting. Namely, the number of mitotic figures can be used as a factor for grading the severity of breast cancer [2]. However, obtaining the mitotic count can be time consuming [3] and is exposed to inter-observer variability (different pathologists recognizing different mitotic figures) [4]. Clearly, if an image analysis tool could perform this task, the inter-observer variability and the duration of the task could be reduced.

A technique that has proven to be very successful in discovering complex structures within high-dimensional data, such as whole-slide images, is called deep learning [5]. Deep learning methods are capable of automatically learning representations of the data that are needed for tasks such as detection and classification. There already have been several scientific competitions in which image analysis techniques were applied to histopathological problems [3][6][7]. In all of these competitions, deep learning methods were the best performing methods.

This research paper addresses how deep learning is currently applied in the field of histopathology and some opportunities on how it could be applied. Therefore, the research question addressed in this paper is as follows:

**Research Question.** *How can deep learning be used in the field of histopathology to improve the quality of diagnoses while reducing time and costs?*

To address this research question, a literature study was conducted to determine how deep learning is currently applied within the field of histopathology. This has not yet been performed. In [8] a general review of image analysis techniques was given but did not address any deep learning techniques for histopathology. In [9] the potential of deep learning in histopathology was shown but no review of the actual techniques were given.

## Organization of This Paper

The remainder of this paper is organised as follows. In the first section, a brief overview of the field of histopathology is given. In the second section, an introduction to deep learning is given. In the third section, deep learning techniques used in histopathology are reviewed. The fourth section discusses challenges that need to be addressed to successfully embed deep learning techniques into clinical practice.

# 1   Histopathology

Histopathology is the field concerned with analyzing human tissue for a certain disease. In clinical practice, the process of analyzing human tissue roughly goes as follows [1]. First, a biopsy is taken from a patient and send to the pathology lab. Next, in the lab the tissue is stained and prepared on a glass slide. The purpose of the staining is to highlight specific structures of the tissue. For example, staining tissue with hematoxylin and eosin (H&E) gives nuclei a dark purple color and other structures a pink color. After the tissue is prepared and stained, a pathologist can examine the tissue using a microscope.

## 1.1   Need for Image Analysis Techniques

In the recent years, pathology labs started to favor scanners instead of microscopes [1]. Such a scanner can process a tissue and store the scan as a digital image, called a whole-slide image. This whole-slide image can then be visualized on a screen, substituting the need to examine the tissue using a microscope. As a result of working digitally, pathology labs now start to collect whole-slide image data. The presence of this data enables pathology labs to automate some manual tasks by using image analysis techniques.

There are numerous of advantages for automating tasks performed by a pathologist. First, automating tasks can increase the quality of the diagnosis. Namely, several (histo)pathological tasks are exposed to inter-observer variability, meaning pathologists can differ in their assessment. As a result, a pathologist can (unknowingly) assign the wrong grade to the disease. Obviously, this can result in giving the patient the wrong treatment. This can have (dramatic) consequences for the patient, but it can also result in giving the patient a too costly therapy.

A second advantage is cost reduction. Enabling a software solution to perform some of the work a pathologist does can reduce the amount of pathological work. Depending on the degree of automation, a hospital could potentially decrease the number of working hours of the pathologist, assign the pathologist to a different task or even employ less pathologists.

The third advantage is more pleasant working conditions for the pathologist. Namely, some tasks within histopathology are regarded as tedious. Clearly, automating these tasks can make the work of the pathologist more pleasant. Furthermore, it also allows the pathologist to focus more on the diagnosis itself and less on the supportive tasks required to do the diagnosis.

## 1.2   Clinical Use

Throughout the remainder of this section, a brief introduction is given to some of the tasks in histopathology for which image analysis techniques have been developed. Since a complete overview is out-of-scope for this paper, only tasks are discussed for which deep learning techniques are presented in section 3 of this paper.

### 1.2.1   Mitosis Counting in Breast Cancer

A commonly used grading system for breast cancer is the Nottingham grading system [10]. The system differentiates three types of grades, indicating the severity of the cancer. To determine the grade, three morphological features are used by microscopic examination of the tissue. These features are:

1. Tubule formation

2. Nuclear grade

3. Mitotic activity

In practice, there has been up till now a lot of interest in automating the determination of the mitotic activity [3]. To determine the mitotic activity, a pathologist manually selects one or more regions in a whole-slide image. Standard practice dictates that a region is selected at the most invasive part of the tumor, at the border and with the highest cellularity [11]. Next, the pathologist counts the number of mitotic cells within this area. The size of the area is generally 2 $mm^2$, which corresponds to 8-10 high power fields (HPFs). A HPF refers to the area visible under a microscope using maximum magnification. An illustration is given in figure 1. After the number of mitosis are counted, a grade between 1-3 can be assigned to the mitotic activity (see table 1).

Determining the mitotic count, however, is regarded as a subjective procedure that is exposed to intra-observer variation. Furthermore, determining the mitotic activity is regarded as a tedious and time consuming activity that can take up to 5-10 minutes in a single area. Clearly, automating this task can be useful.
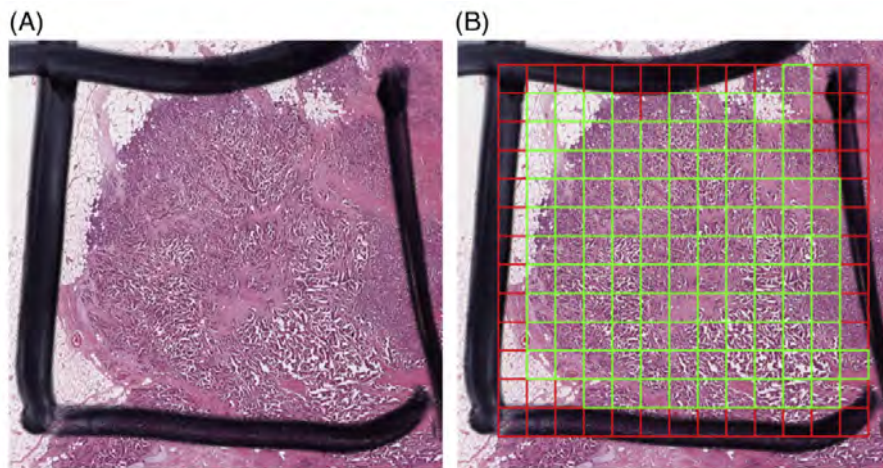


Fig. 1: (A) Selected area within a whole-slide image. (B) Grid in which each rectangle corresponds to one HPF. Source: [3]

| Mitotic Count | Score |
|:---:|:---:|
| $< 6$ | 1 |
| $6 - 10$ | 2 |
| $> 10$ | 3 |

Tab. 1: The mitotic count in a 2 $mm^2$ area is used to determine the score for the mitotic activity.

### 1.2.2 Grading Brain Gliomas

The most common malignant type of brain tumors in adults are gliomas [12]. To grade the severity of the glioma, a pathologist can use the WHO grading system. The WHO grading system distinguishes four categories I-IV. Category I (figure 2a) are the least severe gliomas, associated with long-term survival. Category IV gliomas (figure 2b) are the most severe gliomas, associated with a much lower long-term survival probability. To assign a grade to the glioma, a pathologist examines a whole-slide by looking at factors such as mitosis, nuclear atypia, microvascular proliferation, and necrosis [12].
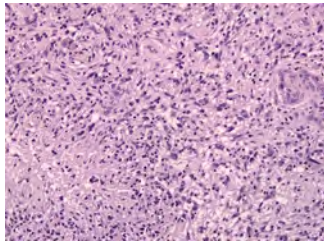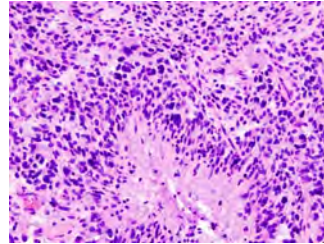


(a) Grade I glioma    (b) Grade IV glioma

Fig. 2: Two examples of H&E stained gliomas.

### 1.2.3 Obtaining Gland Statistics in Colon Cancer

To determine the treatment of a patient with colon cancer, morphological statistics from whole slide images are often used [7]. These statistics are mostly based on the glands within the colon. Namely, glands in the colon are visually different in benign and malignant tissue (see figure 3). To obtain these statistics, a pathologist first needs to identity the glands in the image. Clearly, if a computer could identify the glands it could also produce the statistics required for the diagnosis.
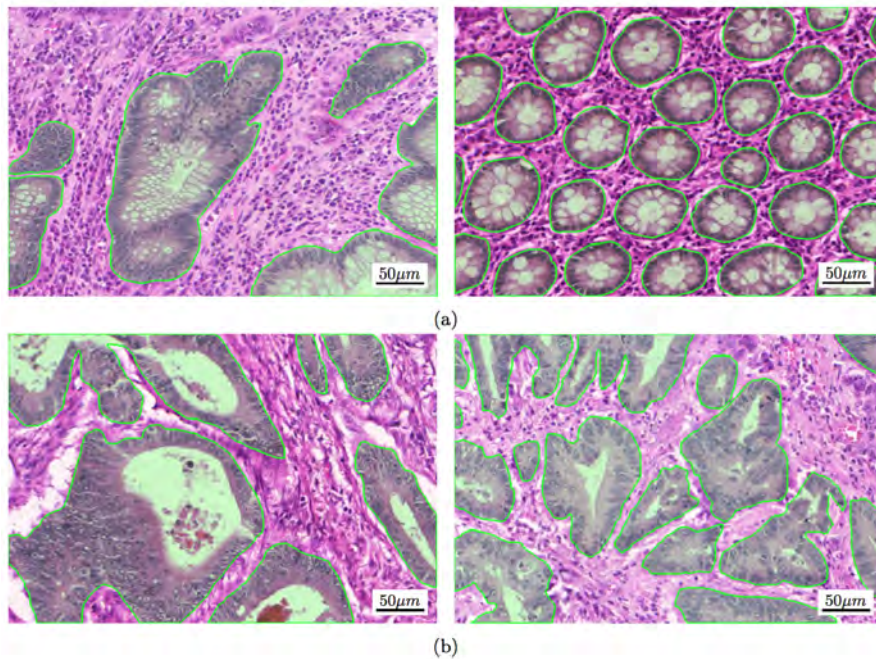
Fig. 3: Example images of gland segementation. (a) benign tissue. (b) malignant tissue. Source: [7]

## 2 Deep Learning

Deep learning is a class of techniques that can automatically discover representations from raw data for tasks such as classification and detection [5]. Deep learning models have proven to be very powerful in performing tasks on high-dimensional data, such as image recognition. A model that has proven to work particularly well in image recognition is the convolutional neural network (CNN). Throughout this section, a brief introduction is given to how CNNs work. Since the CNN is a special case of the classical feed-forward neural network, these networks are first briefly explained. After that, we move towards how CNNs are a special class of the standard feed forward neural network.

### 2.1 Feed-Forward Neural Networks

Artificial neural networks are mathematical models, inspired by the human brain, that are capable of learning complex transformations of some input to a certain output. In general, the input is a $D$-dimensional vector $\boldsymbol{x} = (x_1, \ldots, x_D)$, corresponding to for example an image. The output is in general a $K$-dimensional vector $\boldsymbol{y} = (y_1, \ldots, y_K)$, corresponding to $K$ labels we wish to assign to the input. For example, whether there is a mitotic cell in the image or not. The input-output transformation in the network is performed by

neurons - the basic building blocks of a neural network. In these networks neurons are grouped together in layers. Between these layers neurons are connected to each other, feeding the output of the neurons in one layer as input to the neurons in a consecutive layer. Furthermore, each of these connections between neurons have a weight associated with them. In this section, we assume neurons in a layer are only connected to neurons in a previous and next layer. Networks that exploit this property are called feed-forward neural networks.

In a feed-forward neural network the output $_i$ of a neuron $i$ in one layer is fed as input to the neurons in a consecutive layer. Each neuron $j$ in this consecutive layer first linearly transforms the output of the neurons in the previous layer by taking the dot product of the weights of the connections $\boldsymbol{w}_j$ with the outputs of the neurons. Next, a nonlinear activation function $\phi(\cdot)$ is applied to the output of the linear transformation (1). The output $z_j$ of this neuron is then fed to the next layer in the neural network unless it is the last layer in the network. In this case $\boldsymbol{y}$ is used as notation for the output.

$$z_j = \phi(\sum_i [w_{ji} z_i]) \tag{1}$$

Now to evaluate an input for the network, one simply starts by computing the outputs of the first layer. Next, the outputs of the first layer can then be given to second layer and so on. Finally, the last layer of the network gives the actual output. The process of transforming the input, using the network, to the output is called the *forward pass*. However, to correctly predict the output the network first needs to be trained. That is, the weights of the connections need to be determined.

A common approach to train a neural network is by stochastic gradient descent. In order to do this, a ground-truth labelled dataset $\{\boldsymbol{x}_n, \boldsymbol{t}_n\}_{n=1}^N$ is used, where $\boldsymbol{x}_n$ and $\boldsymbol{t}_n$ correspond respectively to the n'th input vector and the n'th output label. For each input vector $\boldsymbol{x}_n$ we can compare the output of the network $\boldsymbol{y}_n$ with the desired output $\boldsymbol{t}_n$. Using an certain error metric $E(\boldsymbol{y}, \boldsymbol{t})$ we can then compute the error of the network. Using this error and the gradient of the error function we can then update the weights $\boldsymbol{w}$ of the network. These updates can then be repeated through the dataset for numerous iterations. The update rule for stochastic gradient descent takes in general the following form (2), where $\tau$ corresponds to the current training iteration.

$$\boldsymbol{w}^{(\tau+1)} = \boldsymbol{w}^{(\tau)} - \eta \nabla E_n(\boldsymbol{w}^{(\tau)}) \tag{2}$$

To determine the gradient of the error function $E_n$, a technique called backpropagation is used. Backpropagation basically consists out of four steps. First, a forward pass is made through the network by applying an input vector $\boldsymbol{x}_n$ to the network and evaluating all activations of the hidden and output units. Second, all the outputs of the network are evaluated, using the error function $E_n$. Third, using the evaluated outputs, we can compute backwards to determine all the errors throughout the network. Lastly, we can use the outputs of each neuron and the corresponding errors to determine the gradient.

## 2.2   Convolutional Neural Networks

A CNN is a special type of feed-forward neural network that works specifically well for images. CNNs can be trained in the same manner as feed-forward neural networks. However, CNNs basically differ from 'classical' feed-forward neural networks by exploiting two properties, namely: 1) *local-connectivity* and 2) *weight sharing*. The local-connectivity property implies that neurons from one layer are only partially connected to the neurons in the previous or next layer. This in contrast to the standard feed-forward neural network in which all neurons from one layer are fully connected to the neurons in an adjacent layer. The weight sharing property implies that groups of neurons within a layer share the same parameters. The two properties together enable the CNN to learn more complex feature representations in each layer. Furthermore, it ensures training the neural network remains computationally feasible.

The type of layer within the CNN that incorporates the previous mentioned properties is called the convolutional layer (see figure 4). Groups of neurons within this layer that share the same weights are called *filters*. Each neuron is such a filter is only locally connected to the previous layer. A layer can have multiple of these filters (corresponding to the depth of a layer), each that basically will learn an aspect of the data. Furthermore, each filter also has a width and height, multiplied together corresponding to the number of neurons in the filter.
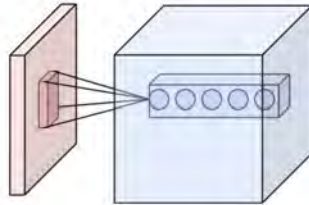


Fig. 4: An illustration of how a convolutional layer is connected to a previous layer. Here the depth of the cube corresponds to the number of filters within the layer. Source: [13]

Another type of layer that is commonly used within a CNN is the subsampling layer (see figure 5). The purpose of the subsampling layer is to reduce the dimensionality of the network. As a result, the number of parameters in the network that need to be estimated are reduced. The max-pooling layer is generally used as subsampling layer. An unit in the max-pooling layer takes the maximum of a certain area within a filter. However, it maintains the depth of the layer it is applied to.

## 3 Deep learning in Histopathology

In this section several deep learning approaches to tasks within in histopathology are discussed.

### 3.1 Mitosis Counting

As discussed in the histopathology section of this paper, the mitotic activity can be used as a factor to grade the severity of breast cancer. To determine the mitotic activity, two tasks need to be performed. First, a pathologist selects one or more areas in the whole-slide image in which to perform the mitosis counting. This task is referred to as *region of interest selection* task. After the areas are determined, the pathologist can perform the actual mitosis counting. In order to do this, the pathologist needs to correctly identify mitosis in the selected areas. This task is referred to as the *mitosis detection* task.

#### 3.1.1 Region of interest selection

Although selecting the regions of interest within a whole-slide image for the mitosis counting is required, it has not yet been addressed in a deep learning setting. Most deep learning approaches up till now solely address the mitosis detection task. However, at the time of the writing of this paper a competition was organized that incorporates both the mitosis detection as well as the region of interest selection task [11].

#### 3.1.2 Mitosis detection

In 2008, [14] were the first to apply deep learning to mitosis detection. To detect mitotic cells in an image, two different classifiers were used. First, support vector regression (SVR) was applied to the color histogram of the image to predict a mitotic color threshold. The rational behind this threshold is that mitotic nuclei exhibit a different color than ordinary nuclei. Therefore, the threshold can be used to filter obvious non-mitotic figures. Images that pass the SVR classifier are passed to the second classifier, a CNN. Although the exact architecture of the CNN is not given, it is revealed that it is loosely based on LeNet5 [15] (see figure 6). The complete mitosis detection method was trained and evaluated on a private dataset of 728 images. Each image had a resolution of 1024 by 768
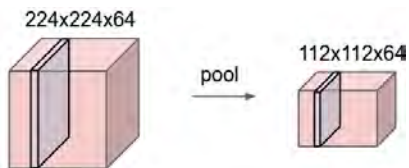


Fig. 5: An illustration of the pooling layer, in which the size of the filters is shrunk but the number of filters remained the same. Source:[13]

pixels. The dataset was manually annotated by a pathologist, identifying 434 mitotic cells in total. Final evaluation of the method on the data resulted in detecting 80% true positives and 5% false positives. However, since the data was private, it is not possible to compare the method with the performance of other methods.
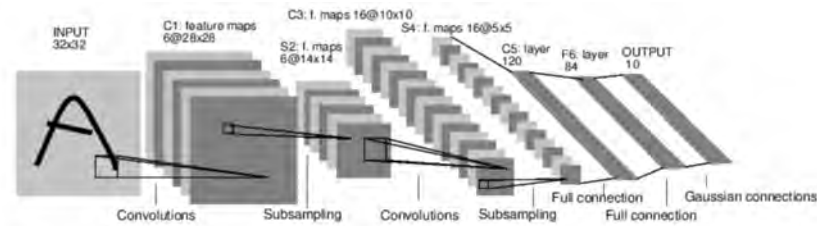


Fig. 6: The LeNet-5 architecture, as described in [15].

This problem was solved by the organization of a public mitosis detection competition held for the ICPR 2012 conference [6]. The task in this competition was to detect mitotic cells in a provided dataset. Contestants were provided public ground truth training data from two scanners and a multispectral microscope. The dataset from each scanner and the microscope contained 50 HPFs from 5 different whole slides. The slides were manually annotated by a pathologist, identifying 326 and 322 mitotic cells in the scanner and microscope datasets respectively. Approximately 30% of the data was withheld from the contestants in order to evaluate the methods proposed by contestants. Note that selecting the areas of interest in the whole-slide images was already performed, so the task was solely to detect the mitotic nuclei in the preselected areas.

The method that won the ICPR 2012 competition was a deep learning method [16]. The proposed method was a combination of two max-pooling (MP) CNNs that operated directly on the raw RGB pixels of the image. The architecture (see figure 7) of both CNNs consisted of several pairs of convolutional and MP layers followed by two fully connected layers, of which the last one is activated by a softmax function. To train each network every pixel in the training data was labeled as either mitosis or non-mitosis. From the total dataset, 132000 smaller images were sampled, of which 50% contained a mitosis pixel in the center and the other 50% a non-mitosis pixel. Furthermore, each of these images was processed 16 times during training by rotating and mirroring the image. The weights of the network were optimized by minimizing the misclassification error over the training set. To detect mitosis in unseen images both networks were applied using a sliding window. The output of both networks was averaged to obtain the final output label. The method achieved the highest $F_1$ score and precision in the competition.

Another deep learning method for the ICPR 2012 competition was proposed by [17]. In this method, first a set of possible mitosis candidates was extracted using two color thresholds. Second, using these candidates a support vector machine was trained using two types of features: 1) handcrafted features and 2) the

| Layer | Type | Maps and neurons | Filter size | Weights | Connections |
|---|---|---|---|---|---|
| 0 | I | 3M x 101x101N | — | — | — |
| 1 | C | 16M x 100x100N | 2x2 | 208 | 2080000 |
| 2 | MP | 16M x 50x50N | 2x2 | — | — |
| 3 | C | 16M x 48x48N | 3x3 | 2320 | 5345280 |
| 4 | MP | 16M x 24x24N | 2x2 | — | — |
| 5 | C | 16M x 22x22N | 3x3 | 2320 | 1122880 |
| 6 | MP | 16M x 11x11N | 2x2 | — | — |
| 7 | C | 16M x 10x10N | 2x2 | 1040 | 104000 |
| 8 | MP | 16M x 5x5N | 2x2 | — | — |
| 9 | C | 16M x 4x4N | 2x2 | 1040 | 16640 |
| 10 | MP | 16M x 2x2N | 2x2 | — | — |
| 11 | FC | 100N | 1x1 | 6500 | 6500 |
| 12 | FC | 2N | 1x1 | 202 | 202 |

| Layer | Type | Maps and neurons | Filter size | Weights | Connections |
|---|---|---|---|---|---|
| 0 | I | 3M x 101x101N | — | — | — |
| 1 | C | 16M x 98x98N | 4x4 | 784 | 7529536 |
| 2 | MP | 16M x 49x49N | 2x2 | — | — |
| 3 | C | 16M x 46x46N | 4x4 | 4112 | 8700992 |
| 4 | MP | 16M x 23x23N | 2x2 | — | — |
| 5 | C | 16M x 20x20N | 4x4 | 4112 | 1644800 |
| 6 | MP | 16M x 10x10N | 2x2 | — | — |
| 7 | C | 16M x 8x8N | 3x3 | 2320 | 148480 |
| 8 | MP | 16M x 4x4N | 2x2 | — | — |
| 9 | FC | 100N | 1x1 | 25700 | 25700 |
| 10 | FC | 2N | 1x1 | 202 | 202 |

Fig. 7: The architectures of the two CNNs used in [16].

output of a CNN. The CNN itself was trained using all positive mitosis instances and approximately 1000 randomly chosen negative non-mitosis instances. Each instance was processed multiple times by mirroring and rotating. The CNN itself was modelled after the LeNet 5 architecture, using two convolutional layers. The method achieved an $f$-measure of 0.659 and 0.589 on respectively the color scanner and multispectral scanner images.

The dataset used for the ICPR 2012 competition was however relatively small. Furthermore, the dataset also did not take variability in the tissue appearance and staining into account. To address this problem, in 2013 the AMIDA13 challenge was organized [3]. The dataset from the challenge consisted of 23 whole slides images. From these images, HPFs (represented as images of 2000 by 2000 pixels) were extracted that at least contained one mitotic cell. Furthermore, variability in tissue appearance and staining was also taking into account. Namely, the slides were selected from a longer period of time to incorporate the differences in staining. Furthermore, pathologists from different institutions were asked to annotate the data in order to reduce the inter-observer variance.

The same team that won the 2012 competition also won the AMIDA13 competition. The team used a similar approach to their previous work [16], but instead employed a Multi-Column MP CNN [3]. Three CNNs were trained on 20 million samples extracted from the dataset. Of these 20 million samples, 10 percent were randomly sampled images with a mitosis pixel in the center. The remaining images were images in which the center had a non-mitosis pixel. Of these remaining images, 50 percent were non-mitosis images that looked relatively similar to the mitosis images. The training of each network took approximately 3 days for GPU optimized implementation. The output probabilities of the CNNs were averaged and used to obtain the final mitotic figures.

Another deep learning method was proposed by [18], in which an approach was presented that combines a lightweight CNN with handcrafted features. The approach consists out of two stages. In the first stage a CNN on the raw pixels and a random forest classifier using handcrafted features are trained independently to classify whether an instance contains mitosis or not. If both classifiers
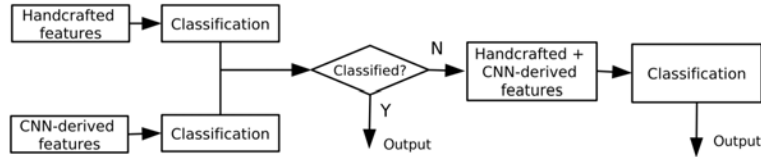
Fig. 8: Classification workflow as described [18].

agree upon the label, the label is assigned to the instance. For the instances on which the classifiers disagree, a second stage exists. In this stage, a third classifier is trained that combines the outputs of the previous two classifiers and makes the final decision on the label (see figure 8 for an illustration of the process).

After the competition ended, [19] proposed a deep cascaded neural network approach that consists out of two phases. In the first phase a CNN is used for fast retrieval of mitosis candidates in the image. The model processes non-overlapping patches of 94 by 94 pixels extracted from the image and assigns an output score to the patch in the image. The CNN is composed out of three pairs of convolutional and max-pooling layers, followed by two fully connected layers. In the second step three CNNs are used to detect mitosis in all positive patches determined by the first CNN. To detect mitotic cells in these patches, the CNNs moves through the patch by using a sliding window. The CNNs in the second step were based on CaffeNet [20]. However, the CNNs differ in the number of neurons in the output layers.

## 3.2 Grading Gliomas

In [12] deep learning is used for the automated grading of gliomas. For this task, a pipeline of two CNNs was used. The first CNN in the pipeline classifies whether the grade of the tumor is $IV$ or $II - III$. If the first CCN assigns grade $II - III$ to the tumor, a second CNN is used to provide whether the grade is actually $II$ or $III$. Grade $I$ tumors were not considered since these are usually cured by surgical resection.

The architecture of the first CNN is modelled after the LeNet-architecture [15] (see figure 9). The CNN consists of 8 layers, including convolution, pooling, ReLU and fully connected layers. The last layer in a softmax layer. The second CNN is deeper than the first CNN, incorporating 19 layers. To classify tissue in a whole-slide image, the image was first split into smaller tiles of 1024 by 1024 pixels. If a tile contained less than 90% tissue it was rejected. The accepted tiles were segmented into smaller images of 256 by 256 pixels, called e-microbiopsies (see figure 10). The e-microbiopsies were eventually fed to the classification pipeline. The proposed method was trained and evaluated on publicly available whole-slide images from The Cancer Genome Atlas (TCGA). The first and second CNN achieved an accuracy of respectively 96% and 71% on this dataset.
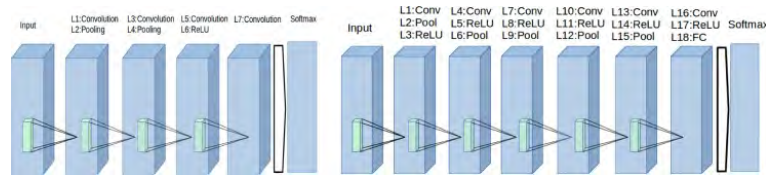
Fig. 9: Overview of the architecture of the two CNNs used for glioma grading [12].. (Left) Shows the architecture of the first CNN in the pipeline. (Right) Shows the architecture of the other CNN.
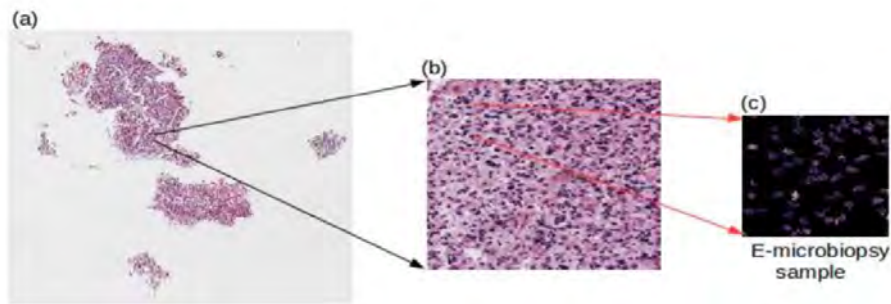


Fig. 10: (A) An whole-slide image is split into tiles of size 1024 by 1024 pixels. (B) A tile is split into smaller images of 256 by 256 pixels called e-microbiopsy samples. (C) The e-microbiopsy samples are eventually fed to the classification pipeline. Source: [12]

## 3.3   Segmenting Epithelial and Stromal Regions

In [21] the task of segmenting and classifying epithelial and stromal regions in both H&E and IHC stained tissue images is presented. The address this task a two step process is described (see figure 11). First, the tissue is segmented using machine learning clustering techniques. H&E stained slides were segmented using the Normalized Cut (Ncut) algorithm or the Simple Linear Iterative Clustering (SLIC) algorithm, whereas the IHC stained images were segmented using a fixed-size windows. Second, the segmented areas were classified as either epithelial or stromal by employing a CNN.

The CNN consisted out of two concecutive pairs of convolutional and pooling layers, followed by two fully connected layers. The final classification was performed by using a support vector machine (SVM) or soft-max. The architecture of the CNN is as follows. First, two consecutive pairs of layers consisting of a convolutional and pooling layer were used. Second, two fully connected layers were employed. Lastly, the classification was performed using a support vector machine (SVM) or using soft-max.

The complete approach was trained and evaluated on two different datasets. The first dataset was acquired from the Netherlands Cancer Institute (NKI)
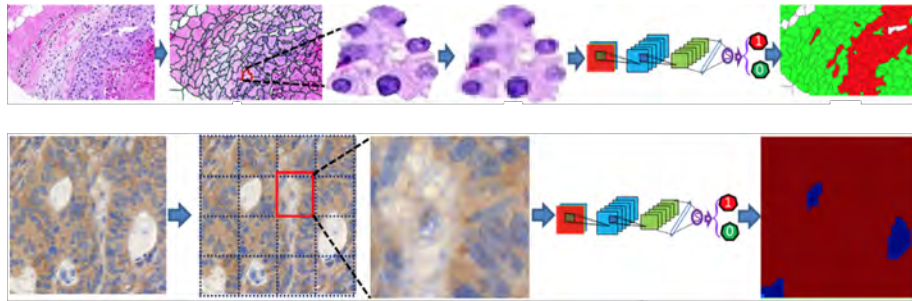
Fig. 11: (Top row) Segmentation and classification of a H&E stained whole-slide image. (Bottom row) Segmentation and classification of a IHC stained whole-slide image. Source: [21]

and the Vancouver General Hospital (VGH). The dataset included 157 breast cancer H&E stained image regions in which the epithelial and stromal regions were manually annotated. The second dataset was acquired from the Helsinki University Central Hospital. The dataset consisted of 27 IHC stained colorectal cancer slides. On both datasets the approach outperformed non-deep learning methods.

## 3.4    Gland Segmentation

To improve image analysis techniques for grading colorectal adenocarcinoma, [7] organised a gland segmentation contest. Gland segmentation is an important image analysis technique since it can be used to obtain morphological statistics that can be used for the actual grading of the colorectal adenocarcinoma. In the challenge, contestants were asked to develop algorithms that are capable of segmenting and classifying glands in H&E stained images. The contestants were provided a dataset consisting of 165 images containing from stage T3 or T4 colorectal adenocarcinoma sections. From these images, 52 were selected and evaluated by a pathologist. The task of a contestant is to segment the glands in these images and classify the gland as either benign or malignant. At the end of the contest, all contestants proposed methods that incorporated a CNN for the segmentation task.

In [22] a three-step strategy is used for the gland segmentation and classification task. First, images are preprocessed using H&E stained color devolution. Of these deconvolved images only the red channel was retained. Second, two CNNs are employed for the segmentation and classification task. The first CNN, called ObjectNet, (see fig. 12a) is used to assign one out of four classes to a pixel. These classes are: 1) background benign, 2) gland benign, 3) background malignant and 4) gland malignant. However, since ObjectNet was not capable of separating physically close glands, a second CNN called SeperatorNet was used. This net was trained for a binary classification task to predict separate objects.

(a) *Object-Net* architecture
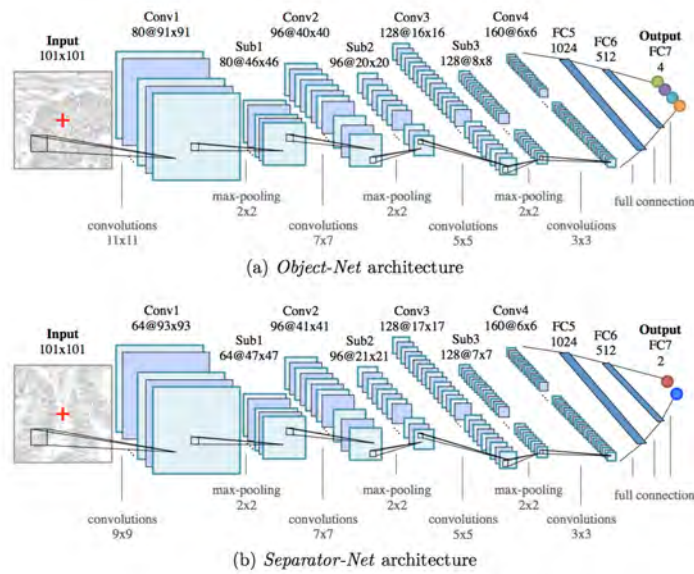


(b) *Separator-Net* architecture

Fig. 12: CNN architectures of the colon gland segmentation approach in [22]. (a) Shows the architecture of Object-Net while (b) shows the architecture of Seperator Net. Both are modelled after the LeNet-5 CNN.

The output of each CNN was individually fed to separate softmax functions to produce two probability distributions. In the third step, the output probability distributions were combined to assign the final class labels to a pixel. To train the CNNs, images from the dataset were rotated. Furthermore, to improve the execution speed of the proposed method, the original 775 by 522 pixels images were resized to 387 by 261 pixels images. Lastly, the complete method was applied to images using a sliding window approach.

## 4   Challenges and Opportunities

Although research on deep learning in histopathology made real progress, there still exists a gap between the proposed methods in research and actually using those methods in clinical practice. Throughout this section, three challenges are presented that need to be addressed to bring deep learning into clinical practice.

### 4.1   Availability of Data

As the mitosis detection competitions [6] [3] and the gland segmentation competition [7] showed, making data available in the form of a challenge can greatly help advance image analysis techniques in histopathology. Namely, these challenges can greatly incentivize others to develop new methods. Furthermore, these challenges also make the evaluation of different types of methods more transparent and easier to compare.

However, there are clearly many other tasks within histopathology that could benefit from image analysis techniques but that not yet have been addressed yet. Making data publicly available could create an incentive for others to develop and test methods for this. Clearly, this can help to create better and more advanced methods.

### 4.2   Regulatory Approval

A major obstacle for a business or institution can be to get regulatory approval for using deep learning in clinical practice [23]. Regulatory approval is given in the US by the Food and Drug Administration whereas in the EU this is done by the European Medicines Agency (EMA).

To get approval in the US clinical software that employs deep learning needs to be cleared under the agency's 510(k) process. However, this process is known for being both time and cost consuming. According to [24] the average cost to get a 510(k) product from concept to market is 31$ million. Moreover, it takes a company on average 51 months from first communication with the FDA to approval. This in contrast to CE approval in the EU that takes on average 11 months according to the same report. Although the costs of solely a software product is possibly lower than the average, getting approval is a huge barrier for getting deep learning software into clinical practice.

### 4.3   Privacy

Another challenge for both business and hospitals that needs to be addressed is that of privacy. Namely, hospitals and other health care providers are often required to keep medical data of a patient confidential. As a result, medical data can often not leave the institution. This imposes a challenge for deep learning techniques since these techniques often are highly dependent on vasts amount of data to be available for training. Clearly, both businesses as well as health care providers need to determine how to address this problem.

A possible solution to this problem is employing decentralized learning [25]. That is, instead of having a central data repository on which the learning algorithm is trained, hospitals keep their data local. Instead, an algorithm is locally trained on the data of the hospital and the parameters of the algorithm are then pushed to a central host. From this central location the parameters can then again be redistributed among other hospitals. As a result, health care providers can still maximally make use of deep learning techniques while keeping their data confidential.

## Conclusion

In this paper was reviewed how deep learning is used for several tasks within the field of histopathology. Namely, it was discussed why image analysis techniques such as deep learning can be beneficial for (histo)pathologists. Furthermore, it was shown how deep learning is applied to 1) mitosis counting, 2) glioma grading, 3) gland segmentation and 4) epithelial and stromal region segmentation. Lastly, three challenges for bringing deep learning into clinical practice were discussed.

For future work, a more extensive review of deep learning techniques within histopathology is possible. Namely, in this work only a handful of tasks to which deep learning is applied were described. However, these tasks are not the only tasks addressed with deep learning in histopathology.

## References

[1]  N. Stathonikos, M. Veta, A. Huisman, and P. van Diest, "Going fully digital: perspective of a dutch academic pathology lab," *Journal of Pathology Informatics*, vol. 4, no. 1, p. 15, 2013.

[2]  E. A. Rakha, M. E. El-Sayed, A. H. Lee, C. W. Elston, M. J. Grainge, Z. Hodi, R. W. Blamey, and I. O. Ellis, "Prognostic significance of nottingham histologic grade in invasive breast carcinoma," *Journal of Clinical Oncology*, 2008.

[3]  M. Veta, P. J. van Diest, S. M. Willems, H. Wang, A. Madabhushi, A. Cruz-Roa, F. Gonzalez, A. B. Larsen, J. S. Vestergaard, A. B. Dahl, D. C. Cireşan, J. Schmidhuber, A. Giusti, L. M. Gambardella, F. B. Tek, T. Walter, C.-W. Wang, S. Kondo, B. J. Matuszewski, F. Precioso, V. Snell, J. Kittler, T. E. de Campos, A. M. Khan, N. M. Rajpoot, E. Arkoumani, M. M. Lacle, M. A. Viergever, and J. P. Pluim, "Assessment of algorithms for mitosis detection in breast cancer histopathology images," *Medical Image Analysis*, vol. 20, no. 1, pp. 237 –248, 2015.

[4]  H. F. Frierson, R. A. Wolber, K. W. Berean, D. W. Franquemont, M. J. Gaffey, J. C. Boyd, and D. C. Wilbur, "Interobserver reproducibility of the nottingham modification of the bloom and richardson histologic grading scheme for infiltrating ductal carcinoma," *American Journal of Clinical Pathology*, vol. 103, no. 2, pp. 195–198, 1995.

[5]  Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.

[6]  L. Roux, D. Racoceanu, N. Loménie, M. Kulikova, H. Irshad, J. Klossa, F. Capron, C. Genestie, G. Naour, and M. Gurcan, "Mitosis detection in breast cancer histological images an icpr 2012 contest," *Journal of Pathology Informatics*, vol. 4, no. 1, p. 8, 2013.

[7]  K. Sirinukunwattana, J. P. W. Pluim, H. Chen, X. Qi, P. Heng, Y. B. Guo, L. Y. Wang, B. J. Matuszewski, E. Bruni, U. Sanchez, A. Böhm, O. Ronneberger, B. B. Cheikh, D. Racoceanu, P. Kainz, M. Pfeiffer, M. Urschler, D. R. J. Snead, and N. M. Rajpoot, "Gland segmentation in colon histology images: The glas challenge contest," *CoRR*, 2016. [Online]. Available: http://arxiv.org/abs/1603.00275.

[8]  M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener, "Histopathological image analysis: A review," *IEEE Reviews in Biomedical Engineering*, vol. 2, pp. 147–171, 2009.

[9]  G. Litjens, C. Sánchez, N. Timofeeva, M. Hermsen, I. Nagtegaal, I. Kovacs, C. Hulsbergen-van de Kaa, P. Bult, B. van Ginneken, and J. van der Laak, "Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis," *Scientific Reports*, vol. 6,

[10]  C. Elston and I. Ellis, "Pathological pronostic factors in breast cancer. i. the value of histological grade in breast cancer: Experience from a large study with long-term follow-up.," *Histpathology*, vol. 5, pp. 403–410, 1991.

[11]  *Tumor proliferation assesment challenge 2016*, Accessed: 2016-06-29, 2016. [Online]. Available: `http://tupac.tue-image.nl`.

[12]  M. Ertosun and D. Rubin, "Automated grading of gliomas using deep learning in digital pathology images: A modular approach with ensemble of convolutional neural networks," *AMIA Annual Symposium Proceedings*, 2015.

[13]  A. Karpathy, L. Fei-Fei, and J. Johnson, *Convolutional neural networks for visual recognition*, Accessed: 2016-06-29. [Online]. Available: `http://cs231n.github.io/convolutional-networks/`.

[14]  C. Malon, M. Miller, H. Burger, E. Cosatto, and H. Graf, "Identifying histological elements with convolutional neural networks," in *Proceedings of the 5th International Conference on Soft Computing As Transdisciplinary Science and Technology*, ACM, 2008, pp. 450–456.

[15]  Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, 1998.

[16]  D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," 2013.

[17]  C. Malon and E. Cosatto, "Classification of mitotic figures with convolutional neural networks and seeded blob features," *Journal of Pathology Informatics*, 2013.

[18]  H. Wang, A. Cruz-Roa, A. Basavanhally, G. H., N. Shih, M. Feldman, J. Tomaszewski, F. Gonzalez, and A. Madabhushi, "Mitosis detection in breast cancer pathology imgaes by combining handcrafted and convolutional neural network features," *Journal of Medical Images*, 2014.

[19]  H. Chen, Q. Dou, X. Wang, J. Qin, and P.-A. Heng, "Mitosis detection in breast cancer histology images via deep cascaded networks," 2016.

[20]  Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, and S. Guadarrama, "Caffe: Convolutional architecture for fast feature embedding," 2014.

[21]  J. Xu, X. Luo, G. Wang, H. Gilmore, and M. Madabhushi, "A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images," *Neurocomputing*, 2016.

[22]  P. Kainz, M. Pfeiffer, and M. Urschler, "Semantic segmentation of colon glands with deep convolutional neural networks and total variation segmentation," 2015.

[23]  G. Freiherr, *Will the fda be too much for intelligent machines?* Accessed: 2016-06-25, 2015. [Online]. Available: `http://www.itnonline.com/content/will-fda-be-too-much-intelligent-machines`.

[24]   J. Makower, A. Meer, and L Denend, "Fda impact on u.s. - medical tech-
       nology innovation," 2010.

[25]   R. Shokri and V. Shmatikov, "Privacy-presevering deep learning," in *Pro-
       ceedings of the 22Nd ACM SIGSAC Conference on Computer and Com-
       munications Security*, ACM, 2015, pp. 1310–1321.