Deep Learning Methods for Mitosis Detection in Breast Cancer Histopathological Images: A Comprehensive Review

Nassima Dif(✉) and Zakaria Elberrichi

EEDIS Laboratory, Djillali Liabes University, Sidi Bel Abbes, Algeria
difnassima05@gmail.com, elberrichi@gmail.com

Abstract. In breast cancer histology, there are three important features for tumor grading, where the proliferation score presents a key component. The mitotic count strategy is among the used methods to predict this score. However, this task is tedious and time consuming for pathologists. To simplify their work, there is a recognized need for computer-aided diagnostic systems (CADs). Several attempts have been made to automate the mitosis detection based on both machine and deep learning (DL) methods. This study aims to provide the readers with a medical knowledge on mitosis detection and DL methods, review and compare the relevant literature on DL methods for mitosis detection on H&E histopathological images, and finally discuss the remaining challenges and some of the perspectives.

Keywords: Mitosis detection · Deep Learning · Breast Cancer · Convolutional Neural Networks · Computer Aided Diagnostic Systems

1 Introduction and Motivation

Breast cancer (BC) is considered as the most diagnosed cancer among women [12]. In histopathology, the pathologist observes stained BC biopsies with hematoxylin and eosin (H&E) under a microscope for grading. With the availability of whole slide scanners, the glass slides are digitized as whole slide images (WSIs). Their analysis is important for tumor assessment, diagnosis, and treatment.

The proliferative activity is one of the three prognostic parameters in BC, where the mitotic index is among the used methods to measure it [10]. Usually, the pathologist counts manually mitoses on the selected high power fields (HPFs) from WSIs. Though, this task is tedious, time-consuming and prone to subjectivity and inter-variability between pathologists. To reduce their workload, computer-aided diagnostic systems (CADs) are proposed. These systems are based on machine [84] and deep learning [18] methods.

The first automated experimental study on H&E tissue sections was reported by Kaman et al. in 1984 [81] for mitosis count. In 1993, Kate et al. [42] have

A. Holzinger et al. (Eds.): AI & ML for Digital Pathology, LNAI 12090, pp. 279–306, 2020.
https://doi.org/10.1007/978-3-030-50402-1_17

computerized this task on stained specimens by Feulgen owing to its capacity to highlight the DNA content. Traditionally, researchers have used machine learning (ML) to automate the mitosis detection task [84]. However, these methods suffer from some serious drawbacks related to their dependency on data representation. Since 2012, the interesting obtained error rate in the ImageNet large scale visual recognition challenge (ILSVRC) has encouraged the image vision community to exploit deep learning (DL) methods due to their capacity to learn data representation. In most recent studies, a considerable literature has been published on DL techniques for mitosis detection. Where, a significant number of published papers is based on convolutional neural networks (CNNs) [6,79].

In recent years, there has been an increasing amount of relevant review papers on the exploitation of DL methods in biomedicine [14], healthcare [59,65], medical [55], and histopathological [82] image analysis. Specifically, the analysis of breast cancer images got a significant interest in different reviews: mammography and MRI [13], breast histology [6] and nuclei detection [39,95].

The paper on deep learning in mammography and breast histology [33] details 10 relevant papers on the mitosis detection by DL methods. However, since 2018, researchers have shown an increased interest in DL strategies for mitosis detection. Despite this concern, no one as far as we know has published a related review paper. The purpose of this contribution is to provide a comprehensive review on the proposed DL methods for the mitosis detection task. The mitosis detection can be carried out on time-lapse phase-contrast microscopy images [76] and stained images by PHH3 [80] or H&E [15]. In this review, we were interested in DL mitosis detection methods on H&E stained images due to their extensive use and availability.

In this review, 28 publications have been collected from the literature review papers on DL methods for medical image analysis and Google Scholar. We used the following keywords to search for publications: 'mitosis detection', 'deep learning', 'breast cancer', 'convolutional neural networks'. First, we selected the period 2012–2019, then 2018–2019 for a maximum of recent investigations. The second strategy was to filter the concerned researches among all works that cited ICPR12 [69], AMIDA13 [86], MITOSIS-ATYPIA-14 [68], and TAUPAC16 [83] papers.

The remaining part of this chapter proceeds as follows: Sect. 2 defines the used terms in this review. Section 3 explains the background of mitosis detection and deep learning, and the related works on DL methods. The purpose of this section is to provide the DL experts with sufficient medical knowledge on breast cancer, in particular, the mitosis detection task, and to detail the DL methods to the medical image analysis community. Section 4 details the open problems and discuses some future outlook. Finally, the last section concludes this work.

1.1 Glossary

– **Breast Cancer (BC).** An uncontrolled proliferation of cells in the breast.

- **Hematoxylin and Eosin (H&E).** Staining procedure that helps to highlight the structural morphology of cells and the other features under the microscope.
- **Histopathology.** A branch of pathology, where changes in tissues are studied.
- **Mitotic Index.** An index that reveals the number of cells undergoing nuclear divisions (mitosis).
- **Whole Slide Images (WSIs).** High resolution images that represent a complete microscope slide.
- **Whole Slide Digital Scanners (WSD).** A scanner that digitizes glass slides into virtual slides, presented as whole slide images (WSIs).
- **Computer Aided Diagnostic Systems (CAD).** Systems that assist specialists in their diagnostic process and help them to make more robust decisions.
- **Machine Learning (ML).** A branch of artificial intelligence, this field helps to extract exploitable and relevant knowledge from big volumes of data.
- **Deep Learning (DL).** A branch of machine learning, based essentially on neural networks.
- **Convolutional Neural Network (CNN).** A deep learning network, inspired by the visual cortex and composed of three types of layers: convolutional layers, pooling layers, and fully connected layers.
- **Fully Convolutional Network (FCN).** A CNN variant, composed of convolutional, pooling and upsampling layers.

2 State of the Art

2.1 Deep Learning

2.1.1 Convolutional Neural Networks (CNN)

Convolutional neural networks are inspired by the visual cortex. In 1962, Hubel et al. [38] proposed a hierarchical model based on complex (C) and simple (S) neuronal cells. According to their observations, Fukushima et al. [28] have developed a deep neural network for pattern recognition. This architecture was particularly useful in Lecun et al. [47] investigation, where they demonstrated the efficiency of CNN networks (LeNet) for supervised learning.

The impressive obtained error rate in the ILSVRC by the AlexNet network [46] has encouraged the computer vision community to propose more optimized architectures. Overfitting is one of the most challenging drawbacks in this field. To solve this issue, the main inspiration was to highlight the role of parameter reduction techniques. In the VGGNet architecture [75], the size of filters has been reduced to $F = 3$ to propose deeper configurations with a small number of parameters. While the Inception network [78] has assessed the significance of inception blocks. ResNet [34] has considered the use of residual blocks to prevent the vanishing gradient problem. Inception-ResNet [77] has examined the implication of residual connections in inception blocks. These findings contributed to a better exploitation of deep neural networks for medical image analysis.

2.1.2 Fully Convolutional Network (FCN)

The fully convolutional network (FCN) is a CNN variant, composed of convolutional, pooling and upsampling layers. This network was largely exploited within the semantic segmentation [52] due to its capacity to perform pixel-wise prediction. Upsampling layers play an important role in semantic segmentation, their purpose is to upsample the output of convolutional layers to obtain the same input size.

2.1.3 Region Convolutional Neural Network (R-CNN)

The region convolutional neural network (R-CNN) [31] is a CNN variant, designed for object detection. This architecture has proved its efficiency compared to the pixel-wise CNN classification method in terms of computational complexity. First, this method performs a selective search on the input image to propose the candidate regions. Then, the CNN is used as a feature extractor. Finally, the generated feature vectors are used to train the bounding box regressor based on the SVM classifier.

To speed up the R-CNN training and the prediction run time, the Fast-RCNN [30] architecture was proposed. Despite R-CNN, this architecture performs end-to-end learning, where the feature vectors are supplied to the region of interest pooling layer. Then, the obtained vector is used for classification and bounding box prediction. The faster R-CNN [67] proposes the exploitation of a separated network for the candidate regions selection instead of the selective search method to reduce the fast R-CNN computational complexity.

2.2 Generalities on Breast Cancer and Mitotic Count

The cancer is defined as an uncontrolled proliferation of cells, the most commonly diagnosed cancer among women is breast cancer (BC) with 11.6% of total cancers death [12]. For BC detection, screening tests are employed such as mammography [41,72], ultrasound [98], and MRI [20]. The ultrasound has proved its efficiency compared to the mammography test in the diagnosis of solid breast lesions [102]. These tests help for earlier detection and therefore improve the chance for surviving.

After an abnormal screening test, a breast biopsy is recommended for tumor assessment, diagnosis, and treatment. There are different types of breast biopsies: fine-needle aspiration (FNA), core needle biopsy (CNB) and excision biopsy (EB). The CNB is known as the preferred technique for histological evaluation and surgical management [91]. Since it is less expensive than the EB, and in contrast to FNA, it highlights the overall histological structure [61]. During the CNB process, a core tissue is extracted by the expert. To extract accurately the tissue from the region of interest, the ultrasound-guided core needle biopsy strategy is used [26]. Then, the tissue is sent to the pathologist for examination.

The pathologist prepares the specimens by formalin fixation and embedding in paraffin then cuts paraffin sections at 3–5 μm thickness [1]. The staining helps

to highlight the structural morphology of cells and the other features under the microscope. There are different types of staining protocols, where the hematoxylin and eosin (H&E) is the most used staining protocol, H stains the cell nuclei with blue-black and E stains the other structures with various degrees of pink [25].

The pathologist observes the stained biopsies under a brightfield microscope or a whole slide digital scanner. The scanner digitizes glass slides into virtual slides, presented as whole slide images (WSIs), under a specific magnification. Then, the pathologist selects regions of interest (ROIS) from these WSIs and analyses them based on specialized software [24]. Figure 1 highlights the difference between WSI and ROI[1].

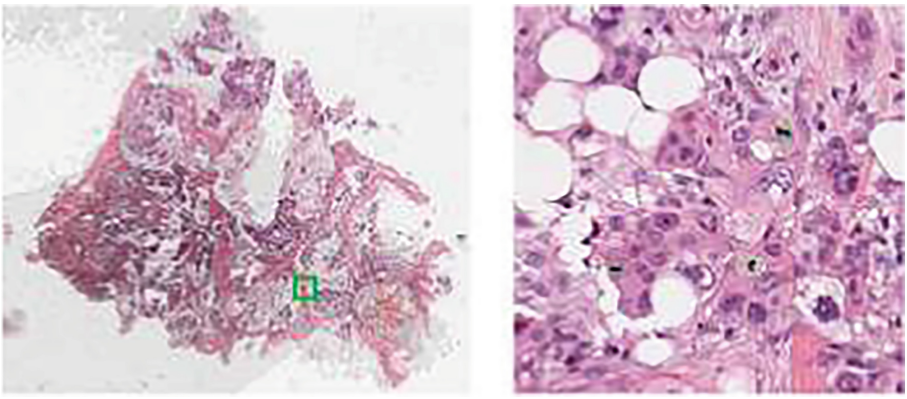

Fig. 1. The difference between a whole slide image and a region of interest from the TAUPAC16 dataset. (Color figure online)

The analysis of the stained specimens helps the pathologist to verify the presence of breast cancer. When the BC is detected, the pathologist performs a histological classification and checks the extent of cancer (in situ or invasive). The BC can be developed in epithelial (carcinoma) or stromal tissues (sarcomas), and carcinomas can be located in milk ducts or milk-producing glands, referred as ductal carcinoma (DC) or lobular carcinoma (LC) respectively [56]. The ductal carcinoma in situ is the most diagnosed cancer among women with 83% of cases [90].

The pathologist uses the grading and the staging systems as prognostic factors to assess the cell's appearance, size of the tumor and its proliferative behavior. Nowadays, the Nottingham grading system [27] is used for breast cancer grading.

The Nottingham histological system [27] is based on three morphological features: tubule formation, nuclear pleomorphism, and mitotic count. These features are scored from 1 to 3. The tubule formation score presents an indicator

[1] http://tupac.tue-image.nl/node/3.

of the percentage of tubular structures in the tumor area, nuclear pleomorphism indicates the degree of variability of nuclei compared to normal cell nuclei, and the mitotic count specifies the number of mitotic in the tumor and its proliferative behavior [8].

The proliferative activity presents an important prognostic parameter in the BC, it is related to the aggressiveness of cancer, where the high proliferative activity is associated with an uncontrolled cell division and therefore reveals a high risk. This activity can be measured by various methods including S-phase fraction, immuno-histochemistry of proliferation-associated to antibodies (Ki-67) and mitotic activity [10].

In oncology, the mitotic index reveals the number of cells undergoing nuclear divisions (mitosis). In the mitosis process, there are four basic phases: prophase, metaphase, anaphase, and telophase. The mitotic nucleus appears denser compared to the normal ones at the beginning of mitosis and transforms into a cell with two nuclei in telophase.

To compute the mitotic index, the pathologist identifies the representative regions of interest (ROIs) at a low magnification since the WSI may contain tens of thousands of HPFs. Each ROI corresponds to 2 mm^2 or 10 high power fields (HPFs). Then mitoses are counted manually under $\times 40$ magnification to score ROIs from 1 to 3 according to the number of mitotic per region. The mitotic count process is tedious, time-consuming (from 5 to 10 min per ROI [29]) and suffers from inter and intra-variability between pathologists [60]. This variability is related to several factors: (a) the subjective selection of the most mitotically active ROIs [11], (b) the various morphology of mitosis within its transformation process, (c) its similar appearance to other structures such as necrotic nuclei and compressed nuclei which can result a high false-positive rate, (d) the small number of mitosis compared to the normal cells nuclei. Hence, to enhance the detection task, a strict protocol must be followed [22]. As a solution to these limitations and to reduce the pathologist's workload, computer-aided diagnostic systems are proposed to automate the mitosis detection task.

2.3 Computer Aided Diagnostic Systems (CAD)

Computer-aided diagnostic (CAD) systems assist specialists in their diagnostic process and help them to make more robust decisions. CADs are based on machine (ML) [50] and deep learning (DL) methods [65].

Machine learning methods (ML) rely on data representation. Where the feature extraction process is required before the training task. These features are extracted according to the field of application and described as handcrafted features. The extraction process requires prior knowledge in the area of interest, especially in case of medical data. In computer vision, the majority of previous studies on CAD systems have emphasized the use of ML methods, due to their limited requirements in terms of computational resources and volume of data.

Deep learning (DL) is defined as a multi-level representation learning and based mainly on deep neural networks such as convolutional neural networks (CNN) and recurrent neural networks (RNN). In recent years, there has been an

increasing amount of literature on deep learning methods in different domains [32,35,45,50,100]. These rapid developments were influenced by various factors related to (a) the remarkable obtained error rate on the ImageNet dataset [46], (b) the availability of the powerful graphics processing units (GPUs), and (c) the massive available volumes of data.

One major important advantage of using DL methods is their capacity to learn data representation from raw data, which involves less human intervention compared to the traditional ML algorithms. However, despite their efficiency, DL algorithms suffer from major drawbacks: the overfitting problem on a limited volume of data, their high computational complexity, and memory requirements.

The main challenge of DL algorithms for medical image analysis is the limited number of accessible medical images. Moreover, the manual annotation of thousands of images for training requires a considerable effort by experts. On the other hand, the high resolution of histopathological images can cover these limitations by generating a large volume of patches form one digital image based on data augmentation techniques. A considerable amount of literature has been published on the application of DL methods for histopathological image analysis [82]. In the breast cancer histology, the exploitation of DL algorithms has covered several applications such as invasive breast cancer detection [19], epithelial and stromal regions segmentation [96], nuclear atypia scoring [97], and mitosis detection [18].

Several attempts have been made to propose automated methods for the mitosis detection task, in both machine [51] and deep learning [18] fields. The purpose of these studies was to resolve the different obstacles related to this automation. The mitosis has highly variable biological structures, and a similar appearance to other structures and artifacts, which can lead to a high false-positive rate. For example, in telophase, the cell contains two separated nuclei and highlights the presence of one mitosis. Furthermore, their low frequency and the limited number of cells undergoing mitotic compared to the normal nuclei cells are leading causes to the data unbalancing problem. Moreover, biopsies preparation, staining, and digitization are key issues in the generation of non-uniform histological images.

The following part reviews the proposed automatic deep learning methods for mitotic figures detection.

2.4 Datasets

The main obstacle faced by many researchers for histopathology images analysis was the availability of public big data, this issue is related to several restrictions: (a) privacy, (b) the extensive time and effort for their annotation, and (c) the variability of staining and digitization methods and scanners between laboratories.

To promote the development of robust frameworks for breast cancer histopathological image analysis, many challenges have been organized. Their purpose was to improve the performance on open access and high quality annotated datasets, where different tasks have been covered: metastasis detection

in lymph nodes (CAMELYON16, CAMELYON17), mitosis detection (ICPR12 [69], AMIDA13 [86], MITOS-ATYPIA-14 [68])), nuclear atypia scoring (MITOS-ATYPIA-14 [68]) and tumor proliferation scoring (TUPAC16 [83]).

Table 1 compares between the proposed datasets for the mitosis detection task.

Table 1. Publicly available datasets for mitosis detection.

Dataset			ICPR12	AMIDA13	MITOS-ATYPIA-14	TUPAC16	TUPAC16 auxiliary
Scanners			Aperio (A) Hamamatsu (H) Microscope (M)	Aperio (A)	Aperio (A) Hamamatsu (H)	Aperio (A)	Aperio (A) Leica SCN400 scanner (L)
WSI	Total		5	23	–	821	73
	Dimension		–	–	–	50000 × 50000	–
HPF	Total		50	596	136	–	656
	Dimension	A	2084 × 2084	2000 × 2000	1539 × 1376	–	2000 × 2000
		H	2252 × 2250	–	1539 × 1376		L: 5657 × 5657
		M	2767 × 2767		–		–
Mitoses	Train		226	550	–	–	1552
	Test		100	533		–	
Pathologists			1	2	3	–	3
Winner			IDSIA [18]	IDSIA [86]	CUHK team	LUNIT [62]	LUNIT [62]

The (a) ICPR 2012 is a small size dataset composed of 5 WSIs, which have been collected from one laboratory and annotated by one pathologist. This dataset has not considered the problem of inter variability between pathologists and laboratories, which limits the power of the trained models in terms of generalization. To improve the proposed systems, more challenging datasets have been published: AMIDA13 and MITOS-ATYPIA-14. The (b) AMIDA13 has been collected at different time points and contains a considerable number of annotated HPFs (596) by 2 pathologists. The MITOS-ATYPIA-14 is a larger dataset composed of 1136 HPFs and annotated by 3 pathologists. However, these datasets have not automated the full grading task, since, the ROIs were selected manually by pathologists. Moreover, the pathologist computes manually the proliferation score according to the detected mitosis by the automatic system. For a fully automatic workflow, the TUPAC16 dataset addresses the possibility to predict automatically the tumor proliferation score from the WSI, where two auxiliary datasets have been provided: TUPAC16 auxiliary for mitosis detection and regions of interest for the automatic selection of ROIs. TUPAC16 auxiliary is an extension of the AMIDA13 dataset with 50 supplementary WSIs.

2.5 Deep Learning Methods for Mitosis Detection

Figure 2 displays the distribution of the 28 selected papers per year. It highlights a considerable amount of researches in 2018 including January 2019. The first contribution was published in 2008, most studies in this period have emphasized

the use of machine learning approaches because of: their large exploitation in computer vision, the lack of powerful resources and publicly available mitosis detection datasets. Since 2012, there has been a growing interest in deep learning methods due to the availability of datasets, the optimization of DL architectures, open-source libraries and pre-trained models.

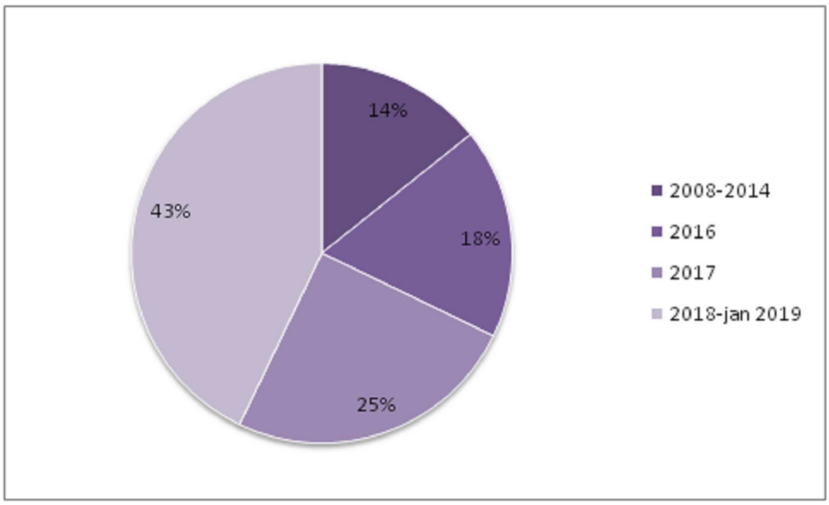

Fig. 2. The distribution of mitosis detection papers per year.

The mitosis detection task with DL methods has several drawbacks related to:

- The limited number of medical data.
- The limited number of mitotic figures because of their low frequency.
- The high false positive rate.
- The high variance between the digitized histopathological images under different conditions.
- Overfitting problems.
- The required computational resources and memory storage.

To solve these limitations, several attempts have been made in the state of the art, where different strategies have been exploited, such as:

- Regularization strategies to reduce overfitting problems.
- Transfer learning, fine tuning, and the exploitation of CNN as a feature extractor to reduce the training run time complexity and overfitting problems.
- FCN and deep detection methods to enhance the precision and to reduce the computational complexity.
- Regression networks to reduce the inference time.

– Multi-scale learning to exploit the contextual information and to enhance the detection task.
– Two stages learning methods to solve the high false positive rate problem.

2.5.1 Regularization Methods

Despite the availability of mitosis detection datasets, the number of simple remains limited for DL applications. Moreover, these datasets are unbalanced, since, the number of mitotic figures is restricted compared to the other structures. Thus, to address these complexities, prior studies have noted the importance of regularization methods, such as: the exploitation of small models [18,21,88], the data augmentation techniques (random patch extraction, translation, rotation, mirroring and flipping), transfer learning [15], fine-tuning [16,94], the use of CNNs as feature extractors [5], ensemble learning [15] and learning from crowds [4]. The purpose of these investigations was to improve the generalization capacity of the generated models. Another key component was the exploitation of stain normalization techniques [44,54,66] to reduce the inter variability between labs. The purpose of these strategies is to covert the processed slides under various conditions to a normalized space [54]. This step is important for the exploitation of the generated models within other labs.

Table 2 resumes the used stain normalization techniques as a preprocessing before the application of DL methods. As far as we know, the SVD-geodesic based stain normalization technique [54] is the most commonly employed in the automatic mitosis detection field [4,21,62,64,85,99]. For more information, Saafin et al. [70] reviewed the relevant literature on stain normalization methods for digital pathology image analysis. Despite the importance of this preprocessing, many significant papers have ignored this step [15,18]. However, the stain normalization is not required when training and testing a model on generated images under common conditions, but it helps to exploit this model within heterogeneous labs.

Table 2. The used stain normalization techniques in the proposed mitosis detection methods.

Reference	Stain normalization method
(Albarqouni et al.) [4] (Das et al.) [21] (Veta et al.) [85] (Paeng et al.) [62] (Zerhouni et al.) [99] (Rao et al.) [64]	SVD-geodesic based stain normalization technique [54]
(Wu et al.) [94] (Kausar et al.) [43] (Akram et al.) [2]	Color transfer between images [66]
(Beevi et al.) [7]	A nonlinear mapping approach using image-specific color deconvolution [44]
(Shah et al.) [74]	Stain specific standardization method [9]

2.5.2 Pixel-Wise and Patch-Wise Classification Strategies

Many recent papers have used DL methods for mitosis counting on the H&E stained slides. To the best of our knowledge, Malon et al. [57] have made the first attempt to automate this task based on convolutional neural networks. They used a set of 728 images at ×400 magnification, then the SVM classifier was trained on the obtained results to automate the grading process. Tables 3 and 4 present the proposed deep learning methods for mitosis detection.

The proposed DL methods for automatic mitosis detection are categorized into pixel [4,18,85,99] and patch wise classification [40,94] strategies. The pixel-wise classification method is considered as a semantic segmentation, where each pixel is labeled separately as mitosis or non-mitosis. Cireşan et al. [18] proposed a max-pooling CNN as a pixel-wise classifier for mitosis detection, where they averaged the output of three classifiers to improve the generalization capacity. The best-obtained results in both ICPR and AMIDA challenges provide strong evidence on the efficiency of this method. However, one major drawback of this approach is its high inference time: 8 min per HPF. Moreover, the pathologist selects many ROIs (HPFs) from the same WSI for analysis. This makes this method time consuming and not feasible for clinical use.

To reduce the time complexity, the patch wise classification strategies have been widely considered. First, patches or mitosis candidates are generated and subsequently, trained as mitosis or not mitosis. In the patch generalization process, the images are converted to blue ratio to highlight the candidate nuclei, due to their high blue intensity in the stained digital slides. Then a segmentation method is performed based on different mechanisms: globally fixed and local dynamic thresholding [88], k-means clustering algorithm [5], aggressive and weaker color threshold and grid search [58], krill held algorithm (KHA) [7], otsu's thresholding method [21,62], globally binary thresholding [87].

The related literature to the classification of mitosis candidates has highlighted several use cases of DL methods: training a network from scratch [21,40], transfer learning [15] or fine tuning [43,94], the use of CNNs as feature extractors [5,7] and the combination between handcrafted and CNN features [58,71,88].

2.5.3 Training from Scratch and Fine-Tuning

Janowczyk et al. [40] trained the cifar-10 AlexNet network to classify the mitosis candidates, based on extracted patches at x20 magnification. However, this low magnification can be a major source of uncertainty for CNN. In another study, Das et al. [21] have evaluated the effectiveness of a shallower CNN on the decomposed sub-patches by the Haar wavelet decomposition method. These methods proved their efficiency for mitosis classification. On the other hand, training a network from scratch is time-consuming and can lead to an overfitting problem due to the limited amount of data. To overcome these limitations, transfer learning and fine-tuning methods have been explored in several studies [43,94]. In these methods, pre-trained models are adapted to the new classification task,

Table 3. Deep learning methods for mitosis detection (1).

Method	Segmentation	Classification	Training	Dataset
[57]	Color histogram	CNN SVM	From scratch	A set of 728 images at 400X magnification
IDSIA [18]	Max pooling CNN		From scratch	ICPR12
[40]	Blue-ratio	CNN: cifar-10 AlexNet network	From scratch	–
[21]	Blue-ratio + Otsu's thresholding	CNN	From scratch	ICPR12 MITOS-ATYPIA-14
FF-CNN [94]	Blue-ratio	FF-CNN	Fine tuning: AlexNet model	MITOS-ATYPIA-14
[5]	K-means clustering	CNN for feature extraction SVM for classification	From scratch	MITOS-ATYPIA-14
[7]	Krill Held Algorithm (KHA)	CNN for feature extraction Softmax for classification	Fine tuning: caffe VGGNet model	MITOS-ATYPIA-14 Regional Cancer Centre (RCC)
HC + CNN [88]	Blue-ratio images + laplacian of Gausian + globally fixed and local dynamic thresholdings	Logistic regression model on CNN features Random forest classifier on handcrafted features	From scratch	ICPR12
[58]	Aggressive and weaker color threshold and grid search	CNN (LeNet) for feature extraction SVM for classification	From scratch	ICPR12
[71]	Blue-ratio images + morphological erosion and dilation operations	CNN	From scratch	ICPR12 MITOS-ATYPIA-14 AMIDA13
CasNN [15]	FCN	CNN	Fine tuning (CNN)	ICPR12 MITOS-ATYPIA-14
DeepMitosis [48]	Segmentation: FCN Detection: faster R-CNN Verification: CNN (ResNet50)		–Fine tuning (FCN) from VGGNet16– Transfer learning (R-CNN) from VGG_CNN_M_1024 –From scratch (CNN)	ICPR12 MITOS-ATYPIA-14
MITOS-RCNN [64]	MITOS-RCNN based on faster-RCNN		Fine tuning VGG-16 layers	ICPR12 MITOS-ATYPIA-14 AMIDA13
[49]	Lightweight R-CNN		From scratch	ICPR12 MITOS-ATYPIA-14
[16]	DRN		Fine Tuning from [17]	ICPR12
[93]	DRN + Hough voting		From scratch	AMIDA13
AggNet [4]	–	Multi-scale CNN	From scratch	AMIDA13
MFF-CNN [43]	Blue-ratio	MFF-CNN	Fine tuning from a caffeNet model	MITOS-ATYPIA-14
[87]	Blue ratio + global binary thresholding	CNN	From scratch	ICPR12 TAUPAC16

(*continued*)

Table 3. (*continued*)

Method	Segmentation	Classification	Training	Dataset
MSSN [53]	–	CNN	From scratch	ICPR12 MITOS-ATYPIA-14
[2]	–	CNN	From scratch	MITOS-ATYPIA-14 TAUPAC16
Wide resNet [99]	CNN (wide ResNet)		From scratch	ICPR12 MITOS-ATYPIA-14 TAUPAC16
L-view [62]	Otsu's method + binary dilatation	–CNN (L-view based on residual blocks) –SVM for tumor scoring	From scratch	TAUPAC16
[92]	Blue-ration + thresholding methods	– DRN + Hough transform –Decision tree for tumor scoring	From scratch	TAUPAC16

Table 4. Deep learning methods for mitosis detection (2).

Method	Segmentation	Classification	Training	Dataset
[85]	Max pooling CNN		From scratch	AMIDA13 Dataset from two pathology labs in the Netherlands [3]
[63]	Max pooling CNN with one dropped fully connected layer		From scratch	AMIDA13
[80]	Brown and blue cannels	CNN	From scratch	TAUPAC16 Dataset from three different hospitals in the Netherlands
[74]	Otsu's method + binary dilatation	MitosNet (CNN variant)	From scratch	Dataset from three international pathology centers

where the weights are transferred to another target network, then a subset of layers is retrained according to the new classification problem.

Wu et al. [94] fine-tuned their deep fully fused convolutional neural network (FF-CNN) based on the AlexNet model. The FF-CNN fuses multi-level features by linking the output of Conv3 and Conv4 to the fully connected layer. Their application has outperformed the winner of the ICPR2014 challenge, which proves the capacity of fine-tuning in the mitosis detection task. One additional

advantage of using fine-tuning strategies is their limited computational requirements in terms of GPU's capacity, where a standard CPU is enough to complete this task.

2.5.4 Feature Extraction with CNN

In other investigations, CNNs have been used as feature extractors. Albayrak et al. [5] employed the CNN network for feature extraction, LDA and PCA methods for feature reduction and the SVM algorithm for classification. For a more optimized run-time complexity in the feature extraction phase, other studies [7] suggest the use of fine-tuned models as feature extractors, where the last four convolutional layers of the Caffe VGGNet model have been retrained.

Both machine and deep learning strategies proved their efficiency in the mitosis detection task. ML methods are based on handcrafted or DL based features. Thus, to take advantage of these two techniques, several studies have addressed the hybridization between both handcrafted and DL features. Wang et al. [88] proposed a cascade approach (HC + CNN), where they trained separately classifiers with CNN-based features and handcrafted features, followed by a third classifier in case of confusion between the decision of the two classifiers. The final class was computed based on an averaging between the generated models. In further studies, Malon et al. [58] combined nuclear features (texture, color, and shape) and CNN features (LeNet 5 [47]), and Saha et al. [71] incorporated 24 handcrafted feature in the first fully connected layer of the CNN. These investigations highlight the efficiency of the hybridization compared to the handcrafted or CNN features when employed separately.

2.5.5 FCN and Deep Detection Methods

To reduce the considerable inference time for the mitosis detection process, other researches have suggested the exploitation of the fully convolutional network (FCN) as a coarse retrieval. Chen et al. [15] proposed a hybrid method based on the FCN to retrieve mitosis candidates and a fine-tuned CaffeNet model for classification. This method has reduced the inference time from 8 min [18] to 0.5 s per HPF. Other investigations suggest converting the obtained DL models into FCNs to speed up the detection process [2,62,94].

Nevertheless, Li et al. [48] have critiqued the use of the FCN to infer the location of mitosis, since, it ignores the regional information. To enhance this process, they showed for the first time the role of deep detection methods for mitosis detection. Their hybrid framework (Deepmitosis) is composed of deep detection (DeepDet), verification (DeepVer) and segmentation (DeepSeg) networks. The main component is the DeepDet network, which localizes mitosis based on the faster R-CNN [67]. Another earlier study by Rao et al. [64] proposed a novel variant of the faster R-CNN (MITOS-RCNN) for small object detection. In another research, Li et al. [49] developed a lightweight

region-based CNN inspired by the RCNN [31], their main purpose was to propose a fast system on CPU computers.

2.5.6 Regression Networks

Another strategy to adjust the inference time for the clinical use was to formulate the mitosis detection task as a regression problem [16,92,93]. Chen et al. [16] proposed a method based on the deep regression network (DRN) with fully convolutional kernels. This network is composed of convolutional (CLs) and deconvolutional layers (DLs). The CLs perform the down-sampling phase for feature extraction, whereas DLs are used to restore the original input size. To prevent overfitting, they fine-tuned the off-the-self deepLab model [23]. Wollmann et al. [93] combined between the deep residual network and the hough voting method. The architecture of this network is composed of three parts: downsampling, factor disentangling part and a pixel-wise classification. The pixel-wise classification provides two branches, which are combined based on the hough voting layer. This method reduces the computational time compared to the other ensemble learning methods due to its single training process.

2.5.7 Multi-scale Learning

The previous studies have suggested training DL methods on a single scale image. On the other hand, the contextual information is important, since the pathologist can observe digital slides from different scales. For an accurate detection task, other studies have been interested in multi-scale learning. Albarqouni et al. [4] proposed an augmented architecture (AggNet), based on a multi-scale CNN and an aggregation layer. To improve the generalization, this network was retrained based on the crowd's annotation labels. This study has been the first attempt to thoroughly examine the CNN networks for generating ground truth labeling from non-expert crowd annotations, in the biomedical context. In another paper, Kausar et al. [43] developed a multi-scale FCNN model (MFF-CNN) based on two different scales FF-CNNs [94] and a fusion layer.

2.5.8 Two Stages Learning Methods

The mitoses are characterized by their low frequency, which can bias the nature of the generated dataset for classification. For example, Cireşan et al. [18] have generated a training set that includes only 6.6% of mitosis pixels. Hence, this may cause a serious class unbalancing and a high false-negative (FN) rate issues. Various approaches have been proposed to solve these limitations [2,40,48,53,87] by exploring the advantages of the two stages learning methods.

Wahab et al. [87] proposed a method based on a two-phase CNNs. In the first stage, the CNN classifies mitosis into easy, normal and hard non-mitosis, the mitosis candidates and the hard non-mitosis are augmented by both

rotations and flipping, while the easy non-mitoses are under-sampled by the blue ratio histogram-based clustering. Then, the generated dataset is retrained by the second phase CNN. In another research, Ma et al. [53] proposed a two-stage deep method. First, a multi-scale and similarity learning convnets (MSSN) was used to treat the FN problem. Subsequently, a similarity prediction model was trained to reduce the high false-positive rate. Akram et al. [2] proposed a deep learning-based self-supervised algorithm. First, CNN was trained on the two sets: BG-rand and FG-Lab which contains the background samples and the centred patches on mitosis. Then, the false-positive detected samples noted as BG-hard were exploited with FG-WSI for retraining the CNN model. This work has analyzed the effect of semi-supervised learning through the use of the extracted mitosis patches from the unlabelled dataset (FG-WSI). Li et al. [48] developed a DeepVer to verify the false positives that have been provided by the DeepDet network. Even though the efficiency of these hybrid systems, the obtained results reveal that the DeepVer model did not improve the performance on the ICPR12 dataset. Another strategy to reduce the FP rate was the exploration of a weighted fitness function [99]. Zerhouni et al. [99] exploited the wide residual network in a pixel-wise classification strategy. To strengthen their training set, they fused between three heterogeneous datasets: ICPR12, MITOS-ATYPIA-14 and the auxiliary mitosis detection dataset in the TAUPAC16 challenge.

2.5.9 Detection from WSI

The previously cited studies are restricted to the detection of mitosis from HPFs. Nevertheless, the pathologist must select manually HPFs from the WSI. Thus, to automate the full detection task, the manual selection of ROIs should be automated. In construct to ICPR12, AMIDA13 and MITOS-ATYPIA-14 challenges, the TUPAC16 has explored the prediction of the proliferation score directly from WSIs. The availability of this dataset has encouraged many researchers to propose frameworks for tumor proliferation score prediction [62,92]. These methods are composed of three main steps: HPFs extraction, mitosis detection, and tumor proliferation score prediction.

Paeng et al. [62] used Otsu's method and the binary dilatation to extract tissue blobs. The extracted patches represent a square of 10 consecutive HPFs. Then, the L-view network was trained on the associated regions to a high cell density. Finally, the tumor proliferation score was predicted based on the number of detected mitosis, 21 handcrafted features, and the SVM classifier. The best results in the TAUPAC16 challenge provide strong evidence about the efficiency of this method. Wollman et al. [92] exploited the threshold-based attention mechanism for the ROI extraction. Then, a DNN network with the Hough transform method was employed for mitosis detection. Finally, the decision tree classifier was trained on the obtained results for the mitosis count.

2.6 Results

Table 5 compares the obtained results by the proposed methods for the mitosis detection task in terms of recall (R), precision (P) and f-measure or accuracy (F1/Acc).

Figure 3 highlights the obtained results on the ICPR12 dataset. These results indicate the efficiency of the faster RCNN for detection [48]. Furthermore, CNNs [21] tend to perform better than their hybridization with handcrafted features

Table 5. The obtained results by the deep learning methods for mitosis detection.

Dataset	Method	Precision	Recall	F-measure/ Accuracy
ICPR12	[21]	0.845	0.837	0.841
	DeepMitosis [48]	0.854	0.812	0.832
	[87]	0.83	0.76	0.79
	[16]	0.779	0.802	0.79
	[49]	0.78	0.79	0.784
	IDSIA [18]	0.88	0.70	0.782
	MSSN [53]	0.776	0.787	0.781
	HC + CNN [88]	0.84	0.65	0.7345
	[58]	0.747	0.590	0.659
	CasNN [15]	0.460	0.507	0.482
AMIDA13	[18]	0.610	0.612	0.611
	[93]	0.547	0.686	0.609
	AggNet [4]	0.441	0.424	0.433
MITOS-ATYPIA-14	[21]	0.996	0.987	0.981
	[5]	–	–	Acc 0.968
	[7]	0.874	0.901	0.886
	CasNN [15]	0.804	0.772	0.788
	MSSN [53]	0.379	0.617	0.470
	DeepMitosis [48]	0.431	0.443	0.437
	MFF-CNN [43]	0.405	0.453	0.428
	[49]	0.40	0.45	0.427
	FF-CNN [94]	–	–	0.393
TUPAC16 auxiliary	[87]	0.57	0.53	0.55
	[62]	–	–	0.652
	[80]	–	–	0.480
ICPR12 + MITOS-ATYPIA-14 + AMIDA13	[71]	0.92	0.88	0.90
	MITOS-RCNN [64]	–	–	0.955
ICPR12 + MITOS-ATYPIA-14 + TAUPAC16 auxiliary	[99]	–	–	0.648
MITOS-ATYPIA-14 + TAUPAC16 auxiliary	[2]	0.613	0.671	0.640

[88], or their exploitation as features extractors [58]. However, their use in a pixel-wise strategy is too expensive for inference. Thus, the key aspect figures into gathering the appropriate selection among various parameters: architecture, strategy, and the network's hyper-parameters. Despite the fast inference time of CasNN [15], this method is less accurate compared to the other approaches, which can be justified by the limits of the FCN for mitosis location inference.

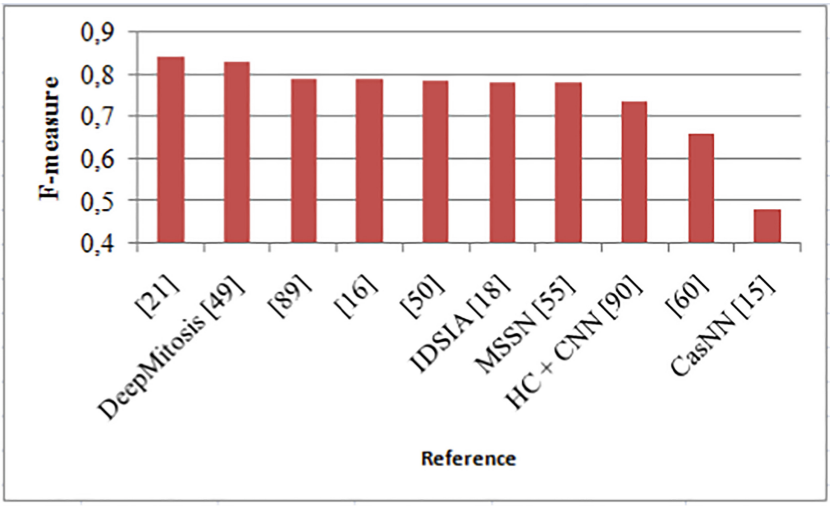

Fig. 3. The obtained results on the ICPR2012 dataset.

Few studies have examined the DL methods on the AMIDA13 dataset. The best results were obtained by the max pooling CNN in terms of f-measure value [18], whereas the proposed method by Wollmann et al. [93] yields the best recall rate. The AggNet [4] reported significantly a lower level of f-measure compared to the previous results. This can be explained by the noisy annotations by non-experts in the crowd.

Despite the ICPR12 dataset, the results on the MITOS-ATYPIA-14 dataset (Fig. 4) highlight the effectiveness of the exploited DL methods as a feature extractors [5,7] and the cascade method CasNN [15] compared to the other approaches [43,48,49,53,94]. The reported results by Albayrak et al. [5] reveal the effectiveness of this approach, where the results have been improved from 0.786 to 0.969 by the feature selection strategy. However, the number of selected features (10) to distinguish the complex morphology of mitosis may be reviewed.

The obtained results on the ICPR12, AMIDA13 and TUPAC16 auxiliary datasets provide additional evidence on the problem of inter variability between pathologists and laboratories. Hence, the annotation of the AMIDA13 by various pathologists, and collecting the TUPAC16 auxiliary dataset from diverse laboratories can justify their low accuracy compared to the ICPR12.

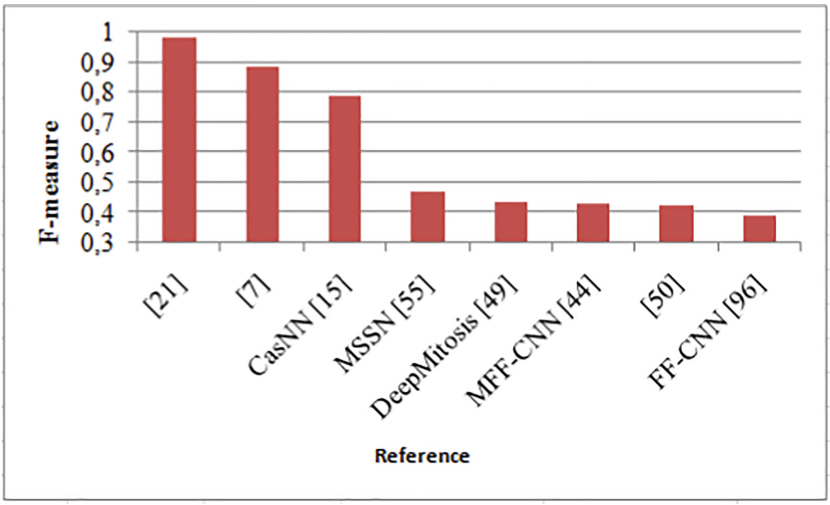

Fig. 4. The obtained results on the MITOS-ATYPIA-14 dataset.

To prevent overfitting, other studies [2,64,99] combined between different datasets for training and testing. Considerable results were obtained by Saha et al. [71] and Rao et al. [64] by combining the ICPR12, MITOS-ATYPIA-14, and AMIDA13 datasets. Saha et al. [71] improved the performance of their deep learning framework with 14% by including 24 significant handcrafted features from a total of 55. However, the importance of handcrafted features is not validated in other studies, this could be attributed to the nature of the selected features and the deep learning architectures.

Table 6 compares between obtained results by the proposed methods for the automated tumor proliferation scoring (TAUPAC16 dataset). The best results have been achieved by Paeng et al. [62] in terms of quadratic weighted cohen's kappa score.

2.7 Computational Time and Materiel

Table 7 resumes the capacity of the exploited GPUs and the processing time of the DL methods for mitosis detection. Powerful GPUs were employed [16,48] and parallelized [64,94] to accelerate the training and the inference time.

The results emphasize the high computational time of the pixel-wise method [18] compared to the other approaches [15]. For a fair comparison, we regrouped these methods by GPU type. Some investigations depend on CPUs [87,88] owing to their restricted requirements related to the shallower CNNs and the small datasets (ICPR12). However, the achieved inference time by Wang et al. [88] (1.5 min per HPF) is not feasible for clinical use, which points out the importance of GPUs.

Table 6. The obtained results on the TAUPAC16 dataset.

Dataset	Method	Quadratic weighted Cohen's kappa score (K)	Spearmans correlation coefficient
TAUPAC16	[62]	0.567 [0.464, 0.671]	0.617 [0.581 0.651]
	[80]	0.471 [0.340, 0.603]	0.519 [0.477, 0.559]
	[92]	0.42	–

The proposed DRN network by Chen et al. [16] is 6 times slower compared to its use by Wollmann et al. [93] on a less powerful GPU as it was influenced by other parameters such as the patch size. The most optimized computational time was obtained on parallel GPUs (<0.5 s) due to their distributed treatment [43,64,94].

Another important parameter is the complexity of the dataset, where we observe a noticeable difference between the inference time in [93] on the ICPR12 dataset and the challenging TUPAC16 auxiliary dataset. The considerable inference time obtained by Wollmann et al. [92] is explained by the end to end classification on the WSIs (50000 px × 50000 px) instead of HPFs.

Table 7. Computational time and the used materiel in the proposed methods for mitosis detection.

GPU	Reference	Training time	Inference time
GPU	[18]	One day for each network	8 min
	[49]	–	6.93 s
Without GPU computation	[88]	11.4 h	1.5 min
	ICPR12 [87]	15 h	48 s
Nvidia GeForce GTX 750M	[4]	–	–
Nvidia Getforce GTX 970	[93]	2.5 days	2.5 s
Nvidia GeForce GTX titan X	[16]	-	15 s
	[48]	–	0.4 s to 0.7 s
	[15]	–	0.5 s 0.3 s
	TAUPAC16 [87]	30 h	1 min
Nvidia Quadro K4200 graphics processor	[21]	–	16 s
4 Nvidia Tesla M40 GPUs	[94]	–	0.375 s
	[43]	–	0.388 s
5 NVIDIA tesla K80 GPUs for training AND a single GPU for testing	[64]		0.5 s
–	[92]	–	5 min (WSI)
–	[71]	–	0.3 s

3 Open Problems and Future Outlook

As discussed in the previous section, different DL methods have been proposed to solve various problems related to the mitosis detection task, where a variety of datasets have been published (ICPR12, AMIDA13, MITOS-ATYPIA-14, TAUPAC16) to encourage the research within the mitosis detection task. Although, the main problem lies in the limited number of mitotic figures (1552 as a maximum), which restricts the capacity of the trained DL models to distinguish between the complex morphology of mitosis and other structures. Moreover, these datasets are collected from a maximum of three different pathology centers. Therefore, this limits the generalization ability of the generated models. The solution is to enhance the model's capacity by learning the whole variance. Some studies have suggested the use of stain standardization techniques, whereas many others have ignored this important step on multi-center datasets [87,92]. The main reason of the restricted number of samples in the medical image datasets is not related to their availability, but rather to the considerable workload by expert pathologists for their annotation.

The crowdsourcing is among the proposed solutions to the lack of annotated data. The study presented by Albarqouni et al. [4] is one of the first investigations to the exploitation of crowdsourcing in the mitosis detection task. Despite the efficiency of these techniques in the other domains, their use in the medical field is critical because of the noisy labels by non-expert participants. More researches using controlled and validated labels by experts is needed to obtain more robust results.

Another solution to the lack of annotated data is semi-supervised learning, which has been previously used to train models on both labeled and unlabelled samples. This technique presents an alternative method to supervised learning in case of a limited amount of labeled data. Up to now, the research has tended to focus on supervised rather than semi-supervised learning for mitosis classification by DL techniques [2]. Therefore, the exploitation of these methods for future works can present a good perspective.

Fine-tuning and transfer learning techniques have been proposed to overcome the overfitting problem related to the lack of annotated data. These techniques present 21% of the selected papers, where the trained models on the ImageNet dataset have been reused and fine-tuned on the mitosis datasets. This is explained by the similarity of low-level features (edges and corners). However, there is no theoretically principles and much uncertainty still exists about the relationship between these two heterogeneous domains. Consequently, sharing models within the same domain could be more helpful due to the similar appearance of the histopathological images compared to the other fields.

Other investigations propose the exploitation of shallower networks to prevent overfitting. These networks are characterized by a limited number of layers, whereas there has been no research based on deep architectures with parameters reduction techniques such as inceptions [78], MobileNet [37,73], and suffleNet [101].

The literature has highlighted the importance of the obtained results especially on the ICPR12 and MITOS-ATYPIA-14 datasets. Where several samples from training and testing sets have been collected from the same source. Thus, the efficiency of the generated models on different sourced samples is not guaranteed. The analysis by Kaman et al. [85] provides important insights on the worst agreement between the automated method and pathologists when evaluated on a new dataset from two pathology labs. Consequently, those models must be validated on new samples and accepted by pathologists. The effectiveness of the pathologist's results was justified by their top-down analysis strategy to include the contextual information, which is related to the size of the patch in the automated methods. The multi-scale learning was employed to solve the lack of sufficient context [94], whereas the majority of studies are based on a single scale learning. Hence, the exploitation of multi-scale contextual networks [89] could be promising in future works.

In the majority of the proposed DL methods, the architectures were presented as a black box, where there has been no clear strategy on the choice of layers and hyperparameters. The progressive visual analytic system (Deepeyes) [63] reveals the importance of visualization to identify unnecessary filters or layers, which can present a good tool to analyze the future proposed architectures for the mitosis detection task. Moreover, more attention is needed to make solutions comprehensible and understandable. In this context, explainable AI [36] helps to make results interpretable by medical experts by creating cooperation between humans and algorithms.

4 Conclusion

The developments in computer vision and digital pathology encouraged the proposition of computerized methods to automate several challenging tasks in the medical domain. Mitosis detection is among the laborious tasks for an expert pathologist, which suffers from inter variability and subjectivity. To solve these shortcomings, DL methods are used due to their capacity to learn data representation. However, their main obstacles are resumed in the required computational resources, the limited amount of data and their unbalanced nature in the medical domain.

To conclude, the literature identifies the strength of DL methods for mitosis detection. Nevertheless, they still suffer from several shortcomings, which support their use as a second tool to aid pathologists rather than their direct exploitation for clinical use.

References

1. Agoff, S.N., Lawton, T.J.: Papillary lesions of the breast with and without atypical ductal hyperplasia: can we accurately predict benign behavior from core needle biopsy? Am. J. Clin. Pathol. **122**(3), 440–443 (2004)

2. Akram, S.U., Qaiser, T., Graham, S., Kannala, J., Heikkilä, J., Rajpoot, N.: Leveraging unlabeled whole-slide-images for mitosis detection. In: Stoyanov, D., et al. (eds.) OMIA/COMPAY -2018. LNCS, vol. 11039, pp. 69–77. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00949-6_9
3. Al-Janabi, S., van Slooten, H.J., Visser, M., Van Der Ploeg, T., Van Diest, P.J., Jiwa, M.: Evaluation of mitotic activity index in breast cancer using whole slide digital images. PLoS One **8**(12), e82576 (2013)
4. Albarqouni, S., Baur, C., Achilles, F., Belagiannis, V., Demirci, S., Navab, N.: AggNet: deep learning from crowds for mitosis detection in breast cancer histology images. IEEE Trans. Med. Imaging **35**(5), 1313–1321 (2016)
5. Albayrak, A., Bilgin, G.: Mitosis detection using convolutional neural network based features. In: IEEE 17th International Symposium on Computational Intelligence and Informatics (CINTI), pp. 000335–000340. IEEE (2016)
6. Anwar, S.M., Majid, M., Qayyum, A., Awais, M., Alnowami, M., Khan, M.K.: Medical image analysis using convolutional neural networks: a review. J. Med. Syst. **42**(11), 226 (2018)
7. Beevi, K.S., Nair, M.S., Bindu, G.: Automatic mitosis detection in breast histopathology images using convolutional neural network based deep transfer learning. Biocybern. Biomed. Eng. **39**(1), 214–223 (2019)
8. Beikman, S., Gordon, P., Ferrari, S., Siegel, M., Zalewski, M.A., Rosenzweig, M.Q.: Understanding the implications of the breast cancer pathology report: a case study. J. Adv. Pract. Oncol. **4**(3), 176 (2013)
9. Bejnordi, B.E., et al.: Stain specific standardization of whole-slide histopathological images. IEEE Trans. Med. Imaging **35**(2), 404–415 (2015)
10. Beresford, M.J., Wilson, G.D., Makris, A.: Measuring proliferation in breast cancer: practicalities and applications. Breast Cancer Res. **8**(6), 216 (2006)
11. Bonert, M., Tate, A.J.: Mitotic counts in breast cancer should be standardized with a uniform sample area. Biomed. Eng. Online **16**(1), 28 (2017)
12. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., Jemal, A.: Global cancer statistics 2018: globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. **68**(6), 394–424 (2018)
13. Burt, J.R., Torosdagli, N., Khosravan, N., RaviPrakash, H., Mortazi, A., Tissavirasingham, F., Hussein, S., Bagci, U.: Deep learning beyond cats and dogs: recent advances in diagnosing breast cancer with deep neural networks. Br. J. Radiol. **91**(1089), 20170545 (2018)
14. Cao, C., et al.: Deep learning and its applications in biomedicine. Genom. Proteom. Bioinf. **16**(1), 17–32 (2018)
15. Chen, H., Dou, Q., Wang, X., Qin, J., Heng, P.A.: Mitosis detection in breast cancer histology images via deep cascaded networks. In: Thirtieth AAAI Conference on Artificial Intelligence (2016)
16. Chen, H., Wang, X., Heng, P.A.: Automated mitosis detection with deep regression networks. In: IEEE 13th International Symposium on Biomedical Imaging (ISBI), pp. 1204–1207. IEEE (2016)
17. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Semantic image segmentation with deep convolutional nets and fully connected CRFs. arXiv preprint arXiv:1412.7062 (2014)
18. Cireşan, D.C., Giusti, A., Gambardella, L.M., Schmidhuber, J.: Mitosis detection in breast cancer histology images with deep neural networks. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013. LNCS, vol. 8150, pp. 411–418. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40763-5_51

19. Cruz-Roa, A., et al.: High-throughput adaptive sampling for whole-slide histopathology image analysis (HASHI) via convolutional neural networks: application to invasive breast cancer detection. PLoS One **13**(5), e0196828 (2018)
20. Dalmış, M.U., et al.: Using deep learning to segment breast and fibroglandular tissue in MRI volumes. Med. Phys. **44**(2), 533–546 (2017)
21. Das, D.K., Dutta, P.K.: Efficient automated detection of mitotic cells from breast histological images using deep convolution neutral network with wavelet decomposed patches. Comput. Biol. Med. **104**, 29–42 (2019)
22. van Diest, P.J., et al.: Reproducibility of mitosis counting in 2,469 breast cancer specimens: results from the multicenter morphometric mammary carcinoma project. Hum. Pathol. **23**(6), 603–607 (1992)
23. Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A.: The pascal visual object classes (VOC) challenge. Int. J. Comput. Vision **88**(2), 303–338 (2010)
24. Farahani, N., Parwani, A.V., Pantanowitz, L.: Whole slide imaging in pathology: advantages, limitations, and emerging perspectives. Pathol. Lab. Med. Int. **7**, 23–33 (2015)
25. Fischer, A.H., Jacobson, K.A., Rose, J., Zeller, R.: Hematoxylin and eosin staining of tissue and cell sections. Cold Spring Harbor Protoc. **2008**(5), pdb-prot4986 (2008)
26. Fishman, J.E., Milikowski, C., Ramsinghani, R., Velasquez, M.V., Aviram, G.: US-guided core-needle biopsy of the breast: how many specimens are necessary? Radiology **226**(3), 779–782 (2003)
27. Frierson Jr., H.F., et al.: Interobserver reproducibility of the nottingham modification of the bloom and richardson histologic grading scheme for infiltrating ductal carcinoma. Am. J. Clin. Pathol. **103**(2), 195–198 (1995)
28. Fukushima, K., Miyake, S.: Neocognitron: a self-organizing neural network model for a mechanism of visual pattern recognition. In: Amari, S., Arbib, M.A. (eds.) Competition and Cooperation in Neural Nets. Lecture Notes in Biomathematics, vol. 45, pp. 267–285. Springer, Heidelberg (1982)
29. Gal, R., Rath-Wolfson, L., Rosenblatt, Y., Halpern, M., Schwartz, A., Koren, R.: An improved technique for mitosis counting. Int. J. Surg. Pathol. **13**(2), 161–165 (2005)
30. Girshick, R.: Fast R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1440–1448 (2015)
31. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Region-based convolutional networks for accurate object detection and segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **38**(1), 142–158 (2015)
32. Hadsell, R., et al.: Learning long-range vision for autonomous off-road driving. J. Field Robot. **26**(2), 120–144 (2009)
33. Hamidinekoo, A., Denton, E., Rampun, A., Honnor, K., Zwiggelaar, R.: Deep learning in mammography and breast histology, an overview and future trends. Med. Image Anal. **47**, 45–67 (2018)
34. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
35. He, X., Deng, L.: Deep learning for image-to-text generation: a technical overview. IEEE Signal Process. Mag. **34**(6), 109–116 (2017)
36. Holzinger, A., Langs, G., Denk, H., Zatloukal, K., Müller, H.: Causability and explainabilty of artificial intelligence in medicine. Wiley Interdisc. Rev. Data Min. Knowl. Disc. **9**, e1312 (2019)

37. Howard, A.G., et al.: MobileNets: efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861 (2017)
38. Humphrey, A., Sur, M., Ulrich, D., Sherman, S.: Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. J. Comp. Neurol. **233**, 159–189 (1985)
39. Irshad, H., Veillard, A., Roux, L., Racoceanu, D.: Methods for nuclei detection, segmentation, and classification in digital histopathology: a review–current status and future potential. IEEE Rev. Biomed. Eng. **7**, 97–114 (2013)
40. Janowczyk, A., Madabhushi, A.: Deep learning for digital pathology image analysis: a comprehensive tutorial with selected use cases. J. Pathol. Inf. **7**, 29 (2016)
41. Kallenberg, M., et al.: Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring. IEEE Trans. Med. Imaging **35**(5), 1322–1331 (2016)
42. Kaman, E., Smeulders, A., Verbeek, P., Young, I., Baak, J.: Image processing for mitoses in sections of breast cancer: a feasibility study. Cytometry J. Int. Soc. Anal. Cytol. **5**(3), 244–249 (1984)
43. Kausar, T., Wang, M., Wu, B., Idrees, M., Kanwal, B.: Multi-scale deep neural network for mitosis detection in histological images. In: International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS), vol. 3, pp. 47–51. IEEE (2018)
44. Khan, A.M., Rajpoot, N., Treanor, D., Magee, D.: A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution. IEEE Trans. Biomed. Eng. **61**(6), 1729–1738 (2014)
45. Klein, G., Kim, Y., Deng, Y., Senellart, J., Rush, A.M.: OpenNMT: open-source toolkit for neural machine translation. arXiv preprint arXiv:1701.02810 (2017)
46. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp. 1097–1105 (2012)
47. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., et al.: Gradient-based learning applied to document recognition. Proc. IEEE **86**(11), 2278–2324 (1998)
48. Li, C., Wang, X., Liu, W., Latecki, L.J.: Deepmitosis: mitosis detection via deep detection, verification and segmentation networks. Med. Image Anal. **45**, 121–133 (2018)
49. Li, Y., Mercan, E., Knezevitch, S., Elmore, J.G., Shapiro, L.G.: Efficient and accurate mitosis detection-a lightweight RCNN approach. In: ICPRAM, pp. 69–77 (2018)
50. Litjens, G., et al.: A survey on deep learning in medical image analysis. Med. Image Anal. **42**, 60–88 (2017)
51. Liu, A., Li, K., Kanade, T.: Mitosis sequence detection using hidden conditional random fields. In: IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 580–583. IEEE (2010)
52. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440 (2015)
53. Ma, M., Shi, Y., Li, W., Gao, Y., Xu, J.: A novel two-stage deep method for mitosis detection in breast cancer histology images. In: 24th International Conference on Pattern Recognition (ICPR), pp. 3892–3897. IEEE (2018)
54. Macenko, M., et al.: A method for normalizing histology slides for quantitative analysis. In: IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 1107–1110. IEEE (2009)

55. Madabhushi, A., Lee, G.: Image analysis and machine learning in digital pathology: challenges and opportunities. Med. Image Anal. **33**, 170–175 (2016)
56. Makki, J.: Diversity of breast carcinoma: histological subtypes and clinical relevance. Clin. Med. Insights Pathol. **8**, 23–31 (2015). CPath-S31563
57. Malon, C., Miller, M., Burger, H.C., Cosatto, E., Graf, H.P.: Identifying histological elements with convolutional neural networks. In: Proceedings of the 5th International Conference on Soft Computing as Transdisciplinary Science and Technology, pp. 450–456. ACM (2008)
58. Malon, C.D., Cosatto, E.: Classification of mitotic figures with convolutional neural networks and seeded blob features. J. Pathol. Inform. **4**, 9 (2013)
59. Miotto, R., Wang, F., Wang, S., Jiang, X., Dudley, J.T.: Deep learning for healthcare: review, opportunities and challenges. Briefings Bioinform. **19**(6), 1236–1246 (2017)
60. Orchid, N.N., Puthanpurayil, S.: Factors affecting the assessment of mitotic count in histopathological sections of tumors: a study of interobserver and intraobserver variability. Int. J. Res. Med. Sci. **4**(3), 762 (2016)
61. Oyama, T., Koibuchi, Y., McKee, G.: Core needle biopsy (CNB) as a diagnostic method for breast lesions: comparison with fine needle aspiration cytology (FNA). Breast Cancer **11**(4), 339–342 (2004)
62. Paeng, K., Hwang, S., Park, S., Kim, M.: A unified framework for tumor proliferation score prediction in breast histopathology. In: Cardoso, M.J., et al. (eds.) DLMIA/ML-CDS -2017. LNCS, vol. 10553, pp. 231–239. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67558-9_27
63. Pezzotti, N., Höllt, T., Van Gemert, J., Lelieveldt, B.P., Eisemann, E., Vilanova, A.: DeepEyes: progressive visual analytics for designing deep neural networks. IEEE Trans. Visual Comput. Graphics **24**(1), 98–108 (2017)
64. Rao, S.: MITOS-RCNN: a novel approach to mitotic figure detection in breast cancer histopathology images using region based convolutional neural networks. arXiv preprint arXiv:1807.01788 (2018)
65. Ravì, D., et al.: Deep learning for health informatics. IEEE J. Biomed. Health Inform. **21**(1), 4–21 (2016)
66. Reinhard, E., Adhikhmin, M., Gooch, B., Shirley, P.: Color transfer between images. IEEE Comput. Graphics Appl. **21**(5), 34–41 (2001)
67. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems, pp. 91–99 (2015)
68. Roux, L., et al.: Mitos & atypia. Image Pervasive Access Lab (IPAL), Agency for Science and Technology & Research Institute for Infocom Research, Singapore, Technical report, vol. 1, pp. 1–8 (2014)
69. Roux, L., et al.: Mitosis detection in breast cancer histological images an ICPR 2012 contest. J. Pathol. Inform. **4**, 8 (2013)
70. Saafin, W., Schaefer, G.: Pre-processing techniques for colour digital pathology image analysis. In: Valdés Hernández, M., González-Castro, V. (eds.) MIUA 2017. CCIS, vol. 723, pp. 551–560. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-60964-5_48
71. Saha, M., Chakraborty, C., Racoceanu, D.: Efficient deep learning model for mitosis detection using breast histopathology images. Comput. Med. Imaging Graph. **64**, 29–40 (2018)
72. Samala, R.K., Chan, H.P., Hadjiiski, L., Helvie, M.A., Wei, J., Cha, K.: Mass detection in digital breast tomosynthesis: deep convolutional neural network with transfer learning from mammography. Med. Phys. **43**(12), 6654–6666 (2016)

73. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C.: MobileNetV2: inverted residuals and linear bottlenecks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4510–4520 (2018)
74. Shah, M., Wang, D., Rubadue, C., Suster, D., Beck, A.: Deep learning assessment of tumor proliferation in breast cancer histological images. In: IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 600–603. IEEE (2017)
75. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
76. Su, Y.T., Lu, Y., Chen, M., Liu, A.A.: Spatiotemporal joint mitosis detection using CNN-LSTM network in time-lapse phase contrast microscopy images. IEEE Access **5**, 18033–18041 (2017)
77. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A.: Inception-v4, Inception-ResNet and the impact of residual connections on learning. In: Thirty-First AAAI Conference on Artificial Intelligence (2017)
78. Szegedy, C., et al.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9 (2015)
79. Tajbakhsh, N., et al.: Convolutional neural networks for medical image analysis: full training or fine tuning? IEEE Trans. Med. Imaging **35**(5), 1299–1312 (2016)
80. Tellez, D., et al.: Whole-slide mitosis detection in H&E breast histology using PHH3 as a reference to train distilled stain-invariant convolutional networks. IEEE Trans. Med. Imaging **37**(9), 2126–2136 (2018)
81. Ten Kate, T., Belien, J., Smeulders, A., Baak, J.: Method for counting mitoses by image processing in Feulgen stained breast cancer sections. Cytometry J. Int. Soc. Anal. Cytol. **14**(3), 241–250 (1993)
82. Jimenez-del Toro, O., et al.: Analysis of histopathology images: from traditional machine learning to deep learning. In: Biomedical Texture Analysis, pp. 281–314. Elsevier (2017)
83. Veta, M., et al.: Predicting breast tumor proliferation from whole-slide images: the TUPAC16 challenge. Med. Image Anal. **54**, 111–121 (2019)
84. Veta, M., Pluim, J.P., Van Diest, P.J., Viergever, M.A.: Breast cancer histopathology image analysis: a review. IEEE Trans. Biomed. Eng. **61**(5), 1400–1411 (2014)
85. Veta, M., Van Diest, P.J., Jiwa, M., Al-Janabi, S., Pluim, J.P.: Mitosis counting in breast cancer: object-level interobserver agreement and comparison to an automatic method. PLoS One **11**(8), e0161286 (2016)
86. Veta, M., et al.: Assessment of algorithms for mitosis detection in breast cancer histopathology images. Med. Image Anal. **20**(1), 237–248 (2015)
87. Wahab, N., Khan, A., Lee, Y.S.: Two-phase deep convolutional neural network for reducing class skewness in histopathological images based breast cancer detection. Comput. Biol. Med. **85**, 86–97 (2017)
88. Wang, H., et al.: Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features. J. Med. Imaging **1**(3), 034003 (2014)
89. Wang, X., Ma, H., Chen, X., You, S.: Edge preserving and multi-scale contextual neural network for salient object detection. IEEE Trans. Image Process. **27**(1), 121–134 (2017)
90. Ward, E.M., et al.: Cancer statistics: breast cancer in situ. CA Cancer J. Clin. **65**(6), 481–495 (2015)
91. Willems, S.M., Van Deurzen, C., Van Diest, P.: Diagnosis of breast lesions: fine-needle aspiration cytology or core needle biopsy? A review. J. Clin. Pathol. **65**(4), 287–292 (2012)

92. Wollmann, T., Rohr, K.: Automatic grading of breast cancer whole-slide histopathology images. Bildverarbeitung für die Medizin 2017. I, pp. 249–253. Springer, Heidelberg (2017). https://doi.org/10.1007/978-3-662-54345-0_56
93. Wollmann, T., Rohr, K.: Deep residual Hough voting for mitotic cell detection in histopathology images. In: IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), pp. 341–344. IEEE (2017)
94. Wu, B., et al.: FF-CNN: an efficient deep neural network for mitosis detection in breast cancer histological images. In: Valdés Hernández, M., González-Castro, V. (eds.) MIUA 2017. CCIS, vol. 723, pp. 249–260. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-60964-5_22
95. Xing, F., Yang, L.: Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: a comprehensive review. IEEE Rev. Biomed. Eng. **9**, 234–263 (2016)
96. Xu, J., Luo, X., Wang, G., Gilmore, H., Madabhushi, A.: A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images. Neurocomputing **191**, 214–223 (2016)
97. Xu, J., Zhou, C., Lang, B., Liu, Q.: Deep learning for histopathological image analysis: towards computerized diagnosis on cancers. In: Lu, L., Zheng, Y., Carneiro, G., Yang, L. (eds.) Deep Learning and Convolutional Neural Networks for Medical Image Computing. ACVPR, pp. 73–95. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-42999-1_6
98. Yap, M.H., et al.: End-to-end breast ultrasound lesions recognition with a deep learning approach. In: Medical Imaging 2018: Biomedical Applications in Molecular, Structural, and Functional Imaging, vol. 10578, p. 1057819. International Society for Optics and Photonics (2018)
99. Zerhouni, E., Lányi, D., Viana, M., Gabrani, M.: Wide residual networks for mitosis detection. In: IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), pp. 924–928. IEEE (2017)
100. Zhang, R., et al.: Real-time user-guided image colorization with learned deep priors. arXiv preprint arXiv:1705.02999 (2017)
101. Zhang, X., Zhou, X., Lin, M., Sun, J.: ShuffleNet: an extremely efficient convolutional neural network for mobile devices. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6848–6856 (2018)
102. Zhi, H., Ou, B., Luo, B.M., Feng, X., Wen, Y.L., Yang, H.Y.: Comparison of ultrasound elastography, mammography, and sonography in the diagnosis of solid breast lesions. J. Ultrasound Med. **26**(6), 807–815 (2007)