Deep Learning based HEp-2 Image Classification: A Comprehensive Review

Saimunur Rahman$^{1,2}$, Lei Wang$^1*$, Changming Sun$^2*$, Luping Zhou$^3$

$^1$VILA, School of Computing and Information Technology, University of Wollongong, NSW 2522, Australia
$^2$CSIRO Data61, PO Box 76, Epping, NSW 1710, Australia
$^3$School of Electrical and Information Engineering, University of Sydney, NSW 2006, Australia

Abstract

Classification of HEp-2 cell patterns plays a significant role in the indirect immunofluorescence test for identifying autoimmune diseases in the human body. Many automatic HEp-2 cell classification methods have been proposed in recent years, amongst which deep learning based methods have shown impressive performance. This paper provides a comprehensive review of the existing deep learning based HEp-2 cell image classification methods. These methods perform HEp-2 image classification in two levels, namely, cell-level and specimen-level. Both levels are covered in this review. In each level, the methods are organized with a deep network usage based taxonomy. The core idea, notable achievements, and key advantages and weakness of each method are critically analyzed. Furthermore, a concise review of the existing HEp-2 datasets that are commonly used in the literature is given. The paper ends with an overview of the current state-of-the-arts and a discussion on novel opportunities and future research directions in this field. It is hoped that this paper would give readers a comprehensive reference of this novel, challenging, and thriving field.

Keywords — HEp-2 Classification, Cell Classification, Deep Learning, Review.

1 Introduction

Indirect immunofluorescence (IIF) is widely recognized as the gold standard test for the characterization of autoimmune diseases such as rheumatoid arthritis, pulmonary fibrosis, Sjogren’s syndrome, and Addison disease in the human body [1][2][3]. IIF is applied to the blood serum, and auto-antibodies are spotted from the fluorescence patterns present in the humane elliptical 2 (HEp-2) cells [2]. HEp-2 cells represent more than thirty cytoplasmic and nuclear patterns existing in almost one hundred types of auto-antibodies [4]. However, only a few types of staining patterns are useful for diagnostic purposes, namely, homogeneous, nucleolar, centromere seen, cytoplasmatic, fine speckled, coarse speckled, cytoplasmatic, nucleolar membrane, and golgi. The classification of these patterns from HEp-2 cells remains a challenging task, primarily due to the subtle category differences and the lack of image acquisition standardization [5]. In the past, the classification of the patterns was carried out manually by specialist physicians or pathologists. Specifically, they used to observe every cell in slide under the microscope, and

*Co-corresponding authors
recognize the patterns based on their experience in the field. Nevertheless, despite the long
time requirement for the test, the results were not consistent among laboratories due to inter-
observer disagreements [6]. Therefore, to mitigate these problems and standardize the manual
IIF test practice, the design of reliable and automated HEp-2 image classification systems has
become an active area of research.

As an established and challenging problem in the field of medical image analysis, HEp-2
image classification (HEP2IC) has been a growing area of research since 2002 [4]. Two areas of
HEP2IC have received attention of researchers, namely, individual HEp-2 cell classification and
HEp-2 specimen classification. Notable progresses [7, 8, 9, 10, 11, 12, 6, 13, 14, 15, 16, 17, 18,
19, 20, 21, 22, 23, 24, 25] have been made in individual HEp-2 cell classification (also known
as cell-level HEP2IC). The HEp-2 specimen classification [26] (also known as specimen-level
HEP2IC) is still a relatively new area of research. Three international IIF image classification
competitions were organized in 2012 [7], 2014 [27], and 2016 [28], and these competitions play
a great role in this progress. As a classical image classification problem, machine learning (ML)
techniques have been widely applied to HEP2IC. The advantage of using ML is that it has
the capability of learning from data, while non-learning based techniques rely on rules which
significantly depend on domain knowledge. However, traditional ML techniques do not directly
learn from raw data but rely on some predefined feature representations, which are a critical
task and require complex engineering and a substantial amount of domain experience. A few
successful data representations (also known as feature representations) used by traditional ML
techniques for HEP2IC include: rotation invariant local binary pattern (CoLBP) [29], local
binary pattern (LBP) with Bag-of-Words [30], multiple linear projection descriptors [5], and
Root-SIFT features & multi-resolution local patterns [6].

Deep learning [31, 32], a representation learning approach that automatically learns feature
representation from raw data, has received considerable attention from researchers in recent
years. It is a very powerful approach that has shown its excellent performance in many areas
of medical imaging [33], including HEP2IC. Among many popular deep neural network (DNN)
aricharces, the convolutional neural network (CNN) [34] has been widely used in HEP2IC.
A CNN is a supervised learning algorithm that takes labeled images as input and learns robust
hierarchial representations which are then used for the classification task. Unsupervised learn-
ing algorithms are also employed to learn feature representations. For example, convolutional
autoencoder [35] is an unsupervised neural network that has been successfully used for feature
learning [13] from HEp-2 images. The main benefit of using unsupervised learning algorithms
is that they do not require image labels.

The use of deep learning in HEP2IC first began in competitions, then conferences, and
very recently in journals. The transition from handcrafted methods to deep learning methods
happened during 2013-2014. The size of datasets also emerges from small to medium-scale
during that period. Figure 1 shows the evolution of HEP2IC with regard to methods and
datasets. In 2016 and 2017, the highest number of deep HEP2IC methods were published.
Three dedicated reviews on HEP2IC methods [7, 36, 28] were published during 2013-2016.
These reviews were published as a part of the IIF image classification competitions, and mostly
focused on providing the overview of the methods participated in the competitions. There were
only a few DL methods involved in these competitions, amongst which Gao et al.’s method [2]
ranked high. The majority of the participating and winning methods in these competitions were
based on traditional handcrafted features (as shown in Figure 1). It is worth mentioning that
there were no new HEP2IC competitions organized after 2016. Two recent reviews [33, 37] on
the broader application of deep learning in microscopic image analysis mentioned a few HEP2IC
methods as shown in Tables 2 and 3 (on pages 9 and 18, respectively) as part of their review
and discussed the achievements and the pros and cons of these methods. To the best of our
knowledge, none of the above reviews has fully covered the publications related to deep learning
based HEP2IC methods.
Motivated by the analysis above, this paper aims to provide a comprehensive review of existing deep learning based HEP2IC methods that have not been discussed by existing reviews. In particular, the state-of-the-art HEP2IC methods published from 2013 to October 2019 (in peer-reviewed conferences and journals, and arXiv) are the focus of this review. The methods are critically reviewed by highlighting the core ideas, pros and cons, and key achievements. For quick reference, the summary on existing methods with key information is also presented in a tabular format (on pages 9 and 18). Since datasets are an integral part of HEP2IC methods, a thorough review of the existing HEp-2 datasets is also given. Based on our experience in DL and study on HEP2IC methods, a dedicated section that discusses the key features of current high-performing methods, the open challenges, and the future trends of HEP2IC research is presented at the end of this review.

To summarize, the objective of this review is to:

1. show the recent progresses of HEP2IC based on DL,
2. discover the key challenges to DL based HEP2IC systems,
3. manifest the contributions made by researchers to solve these challenges, and
4. highlight the novel opportunities in HEP2IC and the experience gained through recent research.

The rest of the paper is organized as follows. Section 2 provides a comparison of this review with the existing related papers and discusses the scope of this review. Section 3 gives a comprehensive review of existing deep learning based HEP2IC methods. For ease of understanding, the methods are organized in deep network usage based taxonomy. The core idea, key achievements, and pros and cons of each method in the taxonomy are critically analyzed. Section 4 gives a concise discussion of the existing public HEp-2 datasets with their evaluation.
2 Existing Papers and the Scope of This Review

There have been a large number of independent studies conducted for both the DL and the HEP2IC. Studies combining both areas have emerged in the recent years. A few independent reviews covering selected important aspects of both areas are available in the literature. Briefly, the existing DL reviews have two primary focuses, namely, (1) covering generic DL architectures and discussing their potential applications, and (2) covering DL techniques for specific application domains such as generic image classification. On the other hand, the HEP2IC reviews were published as a part of HEP2IC competitions [7, 36, 28] mentioned earlier and mostly focus on reviewing the handcrafted methods since the DL methods were limited at that time. Meanwhile, motivated by the excellent results of DL in image classification, a significant number of research on DL has been conducted in the last four years for HEP2IC and many have obtained
significantly better results than the handcrafted methods proposed during the competitions. Although there is no independent review to cover these methods, a few surveys [33, 37] reviewed some of these methods as a part of their broader scope. Table 1 shows a list of existing survey papers involving HEP2IC. Unlike the previous surveys, this paper mainly focuses on reviewing the DL based HEP2IC methods, organizes them in a usage based taxonomy, and discusses their core ideas and key achievements.

Table 1: Summary of existing review papers involving HEp-2 image classification and deep learning. The symbols ✓ and x indicate that the survey covers and does not cover the papers from a particular domain, respectively.

| Reference   | Year | Summary of paper                                      | Focused feature domain |
|-------------|------|-------------------------------------------------------|------------------------|
| 1. Foggia et al. [7] | 2013 | ICPR 2013 IIF image competition methods.              | ✓                      |
| 2. Foggia et al. [36] | 2014 | ICPR 2014 IIF image competition methods.              | ✓                      |
| 3. Lovell et al. [28] | 2016 | ICPR 2016 IIF image competition methods.              | ✓                      |
| 4. Litjens et al. [33] | 2017 | DL methods for medial image analysis.                 | x                      |
| 5. Xing et al. [37] | 2017 | DL methods for microscopic image analysis.            | ✓                      |
| 6. This work | 2019 | DL methods for HEp-2 image classification.           | ✓                      |

Scope of this review. There are two kinds of DL based HEP2IC methods available in the literature. Figure 3 shows a high-level overview of both kinds of methods. The first kind takes a single HEp-2 cell image and predicts its class label. The second kind takes a HEp-2 specimen image and predicts its class label. Note that a specimen image contains many single cells. For simplicity, in the following discussions, we refer the single cell image based HEP2IC methods as the cell-level HEP2IC (CL-HEP2IC) methods and the HEp-2 specimen based methods as the specimen-level HEP2IC (SL-HEP2IC) methods. The issues of both CL-HEP2IC and SL-HEP2IC were introduced in ICPR 2013 and ICPR 2014 IIF image classification competitions, respectively. The CL-HEP2IC is a well researched area and more than 20 papers have been published on this task till October 2019. On the other hand, the SL-HEP2IC is a relatively new area of research and the number of published papers is still low (i.e., less than 10 papers till October 2019). In light of the larger number of published papers, this review will primarily focus on reviewing the CL-HEP2IC methods. Meanwhile, the SL-HEP2IC methods will be discussed briefly. In short, Section 3 will provide a review of both the CL-HEP2IC and SL-HEP2IC methods.

3 Deep Learning for HEp-2 Image Classification

This section aims to review the existing state-of-the-art DL based methods for HEP2IC by highlighting the challenges and contributions from recent publications. At first, it gives an overview of the development of the DL based HEP2IC methods in the recent years. Then, based on the classification task, the existing methods are grouped into two main categories, namely, CL-HEP2IC and SL-HEP2IC. The key motivations, main ideas, and achievements on the benchmark datasets of the methods from both categories are thoroughly discussed.

We begin with an introduction to DL. DL is sub-area of machine learning that deals with algorithms inspired by the working principle of the human brain called artificial neural networks (ANN). The deep neural networks (DNNs) are the feed-forward ANNs with multiple hidden layers. DNNs are capable of learning from the data in an automatic manner. Due to the powerful feature learning capability of DNN, it has been widely used in many computer vision applications, including HEP2IC. In HEP2IC, the spatial structure-preserving variants of DNN such as CNN are used. CNN has three basic components, convolution, pooling, and output layers. The convolution layers compute the output of locally connected neurons. The pooling
(a) Cell-level HEp-2 image classification pipeline

(b) Specimen-level HEp-2 image classification pipeline

(c) A close view of HEp-2 specimen image. The area bounded by the red bounding box shows a single HEp-2 cell

Figure 3: (a) The pipeline of cell-level HEp-2 classification (CL-HEP2IC) methods which take a single HEp-2 cell image as input and output its label. (b) The pipeline of specimen-level HEp-2 classification (SL-HEP2IC) methods which take an HEp-2 specimen image as input and output its label. (c) An HEp-2 specimen image with zoom-in to a single cell. Better viewed in color.

layers perform downsampling of convolution layer output. And, an output layer computes the class scores. Many state-of-the-art CNN architectures have been proposed in recent years. For surveys on the CNN models, please refer to [38, 39].

While it is common to use DNN as an automatic classification tool in general image classification such as ImageNet [32], some application domains such as medical image analysis consider to use the DNN for feature extraction. In HEP2IC, DNN is used as both feature extractor and classifier. When used as a feature extractor, for every test image, a DNN generates a range of representations organized in a hierarchical structure which are used for classification. In a CNN consisting of \( n \) layers that has been trained with supervised learning, the last layer is specified as a multi-class softmax function based on the number of target classes. To use the CNN as a feature extractor, the feature maps at the \((n - 1)\)th layer are usually extracted to form the image-level feature representation. CNN and convolutional autoencoder (CAE) are two spatial structure-preserving variants of DNNs that have been widely used as a feature extractor for HEP2IC. On the other hand, to use the CNN as a classifier, the output class probability scores of the softmax function is regarded as the classification result. DNN generally requires large training samples, and it will suffer from over-fitting otherwise [10]. However, unlike other areas of medical imaging, HEp-2 datasets are limited in sample size. Collection of large-scale datasets may not be impossible, but given the fact that the datasets must be compiled with accurate labelling by expert physicians, it will be a very challenging and time-consuming task. In order to mitigate this gap, most of the HEP2IC use data augmentation (DA) methods such as rotations, flipping, and cropping. DA methods are simple but very effective in training the DNNs. Apart from DA, there are other strategies such as dropout and batch normalization which are often used with DNN to prevent overfitting.

Figure 4 shows the breakdown of the existing deep learning based HEP2IC literature. In particular, it scrutinizes the exiting literature as per the followings: (a) number of CL-HEP2IC and SL-HEP2IC papers by year, (b) number of papers per datasets, (c) number papers by the type of DNN usage, (d) number of papers by the deepness of DNN architectures, and (e) number of papers by the type of DNN architecture (i.e., generic or customized). They are described in order.

As shown in Figure 4(a), DL based HEP2IC has remained an active area since 2013. A large number of methods have been published, especially in the last four years. This could be
Figure 4: Breakdown of existing deep HEP2IC methods in terms of (a) number of publications per year, (b) yearwise number of publications per datasets, (c) number of publications based on the use of deep networks, (d) number of publications based on the deepness of deep models (models that have layers less than the VGG-Net models are considered as shallow models and models that have layers more or equal to VGG-Net are considered as deep models), (e) number of publications based on generic and customised networks used for HEP2IC. Better viewed in color.

attributed to the release of the ICPR 2014 dataset [36] which has a relatively larger number of image samples (with approx. 14k samples) than the early HEp-2 dataset [7], and a top-ranked deep learning based method [9] in IIF image classification competition in 2014 [36]. The existing HEp-2 datasets will be discussed in Section 4 in detail. In the early years, the research was conducted only for the CL-HEP2IC methods, however, SL-HEP2IC methods have also started to appear recently.

Figure 4(b) shows the growth of DL based CL-HEP2IC and SL-HEP2IC methods with respect to two most common HEp-2 datasets, namely, the ICPR 2012 dataset and the ICPR 2014 dataset. The ICPR 2012 dataset was one of the earliest datasets for HEP2IC. It was released in a public HEP2IC competition in 2012 and has only 1,455 samples. Back then, training of high-performing DNNs with such a small dataset was challenging due to the problems such as over-fitting. The ICPR 2014 dataset which is nine times larger than the earlier dataset was first released in 2014 and re-released in 2016, and since then the HEP2IC research community started to take interest in classification methods with DL. Although developing larger HEp-2 datasets with sufficient classification difficulty is an expensive process, it will surely accelerate the DL based HEP2IC development.

Figure 4(c) segregates the existing DL based HEP2IC methods based on their usage of DNN. It demonstrates that most of the existing methods use DNN as the classifier to perform the classification of HEp-2 samples. Only a small group of methods uses DNN purely as a feature extractor. As shown in Figure 4(d), the majority of the existing HEP2IC methods employ deeper DNN architectures for better feature learning. The earliest HEP2IC methods use shallow DNNs such as LetNet-5 [31] and AlexNet [32], but most of the recent methods employ deeper DNNs such as VGG-16 [41] and ResNet-50 [42].

Figure 4(e) shows the type of DNN architectures used in the existing HEP2IC methods. As illustrated, most of the existing methods use generic DNNs for HEP2IC. Generic DNNs are the
popular DNN architectures that have been originally designed for general image classification tasks but successfully applied to other application domains such as medical imaging. Examples of generic DNN include LeNet \[31\], VGG-Nets \[41\], and ResNet \[42\], among others. However, some methods also use specially designed training schemes, new layers into generic DNNs, and fully customized DNN architectures. Figure 4(e) shows that such methods are only a few.

Motivated by the analysis above, in the following sections, the existing DL based HEP2IC methods are further discussed in details. For simplicity, a DNN usage-based taxonomy is defined below (also shown in Figure 4 on page 4) and used to review both the CL-HEP2IC and SL-HEP2IC methods.

3.1 Cell-level HEp-2 image classification (CL-HEP2IC) methods

As aforementioned, considering the nature of DNN usage, there are two types of methods available, namely, (1) methods that use DNN as a feature extractor, and (2) methods that use DNN as a classifier. Table 2 provides a comprehensive summary of both types of methods. It organizes the existing methods based on their year of publication in ascending order. Furthermore, for each method, it lists down the following key information: pre-processing techniques used, basic DNN architectures, performance on the benchmark datasets, and remarks. The table could be useful for the quick referencing of existing CL-HEP2IC methods. Sections 3.1.1 and 3.1.2 give a thorough discussion on these methods.

3.1.1 CL-HEP2IC methods that use DNN as a feature extractor

Feature extraction is a common step in many medical imaging applications including HEP2IC. It is often used in the handcrafted feature based HEP2IC methods, e.g., \[29\], \[30\], and \[5\]. As the name suggests, the feature extraction step extracts necessary features from the input image to perform the classification, i.e., labelling. The state-of-the-art classifiers such as support vector machines (SVM) and k-nearest neighbours (k-NN) are the popular choices in the literature for classification. While the feature extraction and classification steps are treated separately in the handcrafted feature based methods, the DNN based methods combine them as one integral part. As mentioned previously, a DNN could be considered as a hierarchical feature extractor, which is organized in a bottom-up fashion. The hierarchies are usually defined by the layers and each layer takes an input data, performs some operations on it, and then delivers the result as output. The output of each layer could also be treated as the feature and can be used for training the state-of-the-art classifiers. Overall, depending on the training status of DNN, there are two types of feature extraction in the existing HEP2IC methods, namely, feature extraction from a pre-trained DNN model and feature extraction from a fine-tuned DNN model. Both types are discussed below:

Feature extraction from a pre-trained DNN model. Methods of this kind use the basic form of transfer learning, where features from the DNN trained on a large-scale dataset such as ImageNet \[32\] are extracted \[48\] and used for the training of state-of-the-art classifiers. The methods proposed in \[11\] and \[15\] are the examples of this kind. Both of them use CNN features trained on the ImageNet dataset. The method proposed by Lu et al. \[15\] is relatively straightforward. They directly replaced the handcrafted feature descriptors such as SIFT \[49\] used in the traditional image classification pipeline \[50\] with the CNN features. Specifically, the last convolution features (i.e., conv5) of a VGG-16 network are extracted and a multi-class SVM with a radial basis function (RBF) kernel is trained with them. Even though the ImageNet data are considerably different from the HEp-2 samples, this method achieved a very good performance on both the ICPR 2012 and the ICPR 2014 datasets.

On the other hand, Phan et al. \[11\] also uses ImageNet pre-trained features but adopted
| Ref. | Authors | Year | Dataset | Pre-processing | EN | DA | Classifier | Remarks |
|------|---------|------|---------|----------------|----|----|------------|---------|
| 1    | Minut et al. | 2013 | ICPR 2012 | Custom CNN | X | × | Custom CNN | × |
| 2    | Gao et al. | 2014 | CS, HE, MVN | R S F I | 72.0 | × | LeNet-5 | 98.76 |
| 3    | Baynaghaj et al. | 2015 | ICPR 2012 | Custom CNN | X | × | Custom CNN | × |
| 4    | Khan et al. | 2016 | ICPR 2012 | R | 80.3 | × | VGG-Bi-SVM | × |
| 5    | Li et al. | 2016 | ICPR 2012 | Custom CNN | X | × | MFC-ELM-CNN | × |
| 6    | Han et al. | 2016 | ICPR 2014 | CS | 98.49 | × | ResNet-50 | X |
| 7    | Liu et al. | 2017 | ICPR 2014 | CS | 98.37 | × | VGG-16,GoogleNet | X |
| 8    | Gao et al. | 2017 | ICPR 2014 | M | 98.17 | × | Custom CNN | X |
| 9    | Liu et al. | 2017 | ICPR 2014 | AS | 98.7 | × | LeNet-5 | × |
| 10   | Rodrigues et al. | 2017 | ICPR 2014 | CS, AS | 98.14 | × | VGG-16+SVM | X |
| 11   | Li et al. | 2017 | ICPR 2014 | CS | 98.42 | × | ResNet-50 | × |
| 12   | Rodrigues et al. | 2017 | ICPR 2014 | AS | 98.26 | × | VGG-16+SVM | X |
| 13   | Menon et al. | 2018 | ICPR 2014 | CS | 98.29 | × | LeNet-5, AlexNet, and GoogleNet | X |
| 14   | Li et al. | 2018 | ICPR 2014 | CS | 98.50 | × | Ensemble Net | X |
| 15   | Ebrahim et al. | 2019 | ICPR 2014 | CS | 98.71 | × | HEp-Net | X |
| 16   | Shen et al. | 2019 | ICPR 2014 | CS | 98.60 | × | Simple transfer learning | X |
| 17   | Vununu et al. | 2019 | ICPR 2014 | CS | 98.66 | × | HEp-Net | X |
| 18   | Nguyen et al. | 2019 | ICPR 2014 | CS | 98.27 | × | HEp-Net | X |
| 19   | Vannual | 2019 | ICPR 2014 | CS | 98.27 | × | HEp-Net | X |

Table 2: Summary of the existing methods for cell-level HEp-2 image classification. The widely accepted evaluation protocol for ICPR 2012 and SNPHEp-2 is ACA, and for ICPR 2014 is MCA (please refer to Section 4 for details). EN=Image enhancement method; CS = Contrast stretching; HE=Histogram Equalization; ZM = Zero mean; MVN = Zero mean and unit variance normalization; DA = Data augmentation; R = Rotation; S = Shifting; F = Flipping; I = Intensity variations; C = Cropping; AOD = Adoption from other datasets; IDS = Import from related datasets; WS = Specimen-level classification with leave-one-specimen-out protocol.
a more complex classification framework. Unlike the above method, they propose an intensity-aware two-step classification framework with late feature fusion. Furthermore, a feature selection approach [51] is applied to extract the most dominant features for the training of class-specific SVMs in different stages of the framework. However, their performance is not as high as that of the Lu et al.'s method [15].

A more recent work proposed by Cascio et al. [46] also uses a two-step classification framework. Given the CNN features, the proposed framework firstly passes them into a set of class-aware binary SVM’s to generate a compact and discriminative feature representation. Next, a k-NN classifier is used to classify the compact cell features. However, it produces lower performance than the above two methods [15] [11] and requires segmentation masks to obtain a discriminative feature set. Moreover, the idea of training class-specific SVMs is not suitable for datasets with a large number of classes.

Since there exists a domain gap between the ImageNet and HEp-2 data, a few methods in the literature have used fine-tuned CNN models [52] for better HEP2IC. The following paragraphs will focus on the methods that use fine-tuned DNN extracted features.

**Feature extraction from a fine-tuned DNN model.** Fine-tuning of pre-trained DNN models to target datasets generally improves the discriminative capacity of the features, hence, leads to better classification performance. Interested readers are referred to a recent work on this topic [52] for details on fine-tuning. It is worth mentioning that the process of feature extraction from the fine-tuned CNN is similar to that from the pre-trained CNNs. Among the existing methods of this kind, Han et al. [19] used a fine-tuned CaffeNet [32] for feature extraction. While it is common for the CNNs to resize the input image in order to have a fixed-sized input (in other words, size normalized input), the proposed method developed a new pooling strategy called ‘K-spatial pooling’ to support the arbitrary sized HEp-2 cell images. The idea behind K-spatial pooling is to leverage the frequency of neural activation patterns in feature pooling. Given a set of convolutional activation maps (also known as feature maps), the proposed pooling strategy finds the K larger activation values in a defined region from each of feature maps and performs a mean operation on them. Suppose we have N feature maps and M pre-defined regions in each of the feature maps, the K-spatial pooling output will be \( z = [\Psi(F_{M1}^1), \Psi(F_{M2}^2), ..., \Psi(F_{MN}^N)]^T \) where \( \Psi(\cdot) \) calculates the mean values of K larger activations from feature maps \( F \). The feature vector \( z \) can be further used for the training of common classifiers such as SVM. The experimental results in [19] show that the proposed pooling method is able to extract more discriminative and rotation invariant features from the convolutional activation maps than the traditional max-pooling. However, the parameter K is sensitive and determined by empirical evaluations which is a time-consuming task. To improve this situation, instead of using a pre-calculated value of K during the feature extraction, automatically learning it during the fine-tuning of pre-trained CNN model could be a more efficient way.

While the above three methods choose CNN as the feature extractor, in [47] Vununu et al. used the CAE to extract features for CL-HEP2IC. Specifically, they trained two distinct CAEs with the regular (i.e., RGB) and gradient images. While the regular image based CAE learns the geometrical information of HEp-2 cells, the gradient image based CAE learns the local intensity changes in HEp-2 cells. The CAE is based on the VGG-Net architecture. Unlike CNN, the latent space features between the encoder and decoder of CAE are extracted. The decoder part of CAE uses the latent space features to reconstruct the input, hence, it presumed to carry rich information of the HEp-2 cells. The latent space features from the above networks are then combined and classified using a simple neural network based classifier. The joint utilization of cells’ geometrical and local intensity information significantly improves the classification performance, as demonstrated by the state-of-the-art results on both the SNPHEp-2 and the ICPR 2014 datasets.

**Summary of discussion.** Feature extraction based methods use image features that have been automatically learned by the DNN to perform the CL-HEP2IC. Both the pre-trained and
fine-tuned DNN models have been used for feature extraction. However, due to the domain gap between the data used for pre-training and the HEp-2 cell images, the performance of pre-trained model based methods are usually lower than that of the fine-tuned model based methods. Nevertheless, efforts have been made to improve the performance of the former, and they include multi-step and class-specific classification frameworks and discriminative feature selections [11, 46]. On the other hand, features extracted from fine-tuned models are more powerful and do not require multi-step frameworks or explicit feature selection. The early feature extraction based methods are based on CNN, whereas the recent methods also use CAE features.

The common classifiers used in feature extraction based methods are SVM and k-NN. However, for classification tasks, DNNs are usually trained using the softmax classifier [33] and it is possible to directly use it for CL-HEP2IC. In the literature, most of the DNN based HEP2IC methods directly use softmax predictions and avoid external classifier training such as SVM.

3.1.2 Cell-level HEp-2 (CL-HEP2IC) methods that use DNN as a classifier

The methods of this kind treat the feature extraction and classification steps jointly. Given the image, while DNN is used as a classifier, necessary features are automatically extracted, and then classification is performed using the extracted features. Based on the use of network architectures and their training and fine-tuning strategies, the DNN-as-a-classifier based CL-HEP2IC-methods are decomposed into three groups, namely, (1) pure generic DNN based methods, (2) generic DNN based methods with partial changes in layers or training schemes, and (3) customized DNN based methods. They are thoroughly discussed below:

**Pure generic DNN based methods.** Generic DNN based methods use the popular DNN architectures to perform CL-HEP2IC. The methods proposed in [10, 14, 23, 24, 43] are of this kind. Since the generic CNNs are primarily designed for the general image classification tasks such as ImageNet classification [32], pre-processing techniques such as image enhancement and DA (i.e., data augmentation) are carefully considered in most of these papers to achieve high CL-HEP2IC performance. A list of pre-processing techniques used in these methods is given in Table 2.

One of the early and successful DL based HEP2IC methods proposed by Bayramoglu et al. [10] uses the AlexNet CNN architecture. Since AlexNet was originally proposed for the ImageNet classification, the authors have performed a comprehensive study on various pre-processing techniques and used the best performing techniques to train the network. They have managed to achieve the state-of-the-art performance on the ICPR 2012 dataset. One of the interesting strategies they have followed was the adoption of samples from the ICPR 2014 dataset and the SNPHEp-2 dataset for the training. They have showed that their training scheme is very useful for achieving good performance on the smaller HEp-2 datasets.

Two other independent studies [23, 24] conducted by Rodrigues et al. also performed a comprehensive review on various pre-processing techniques for the training of generic CNNs. They considered three well known CNN architectures in their studies, namely, LetNet-5, AlexNet, and GoogleNet [53]. In both of their studies, they demonstrate that without any pre-processing used, GoogleNet outperforms the two other architectures on the ICPR 2014 dataset. This indicates that the deeper CNN model is more robust against illumination changes and is able to generate more robust and discriminative features than the shallow models. Meanwhile, Lei et al. [14] propose a HEP2IC method based on the ResNet architecture. Similar to Rodrigues et al., they use the ICPR 2012 dataset without any pre-processing to train the ResNet-50 CNN model. Their method achieves the state-of-the-art classification performance on the ICPR 2014 dataset. Unlike most of the generic CNN model based CL-HEP2IC methods, Lei et al.’s method first pre-trains the CNN with smaller datasets, and then applies fine-tuning to larger datasets, i.e., the ICPR 2014 dataset. Both studies in [23, 24] and [14] suggest that the deep CNNs are
capable of generating more robust features than the shallow CNNs.

In Rodrigues et al.’s and other similar studies, the authors only considered employing a single CNN stream for HEP2IC. Unlike them, Nguyen et al. study the network ensemble for HEP2IC in [44]. However, their study is only limited to pre-trained models. The motivation behind their study is CNN’s over-fitting tendency with smaller datasets, e.g., ICPR 2012. To avoid the fine-tuning of CNN features, they extracted a wide range of pre-trained features. Specifically, six types of ResNet and GoogleNet based ImageNet pre-trained models are used for the feature extraction. Since it is computationally expensive to combine all of the extracted features, the features are averaged and then used for the softmax classification. The findings of Nguyen et al.’s method are interesting. They showed that combining various model features further enhances the classification performance. At the same time, we presume that their method could benefit from fine-tuning. Also, instead of simply averaging the features, a weighted feature fusion mechanism could be adopted in order to obtain a more effective image representation.

Another interesting work in this group was proposed by Majtner et al. in [43]. They propose to perform generative adversarial network (GAN) [54] based DA. In existing HEP2IC methods, DA is performed using simple image transformations such as rotation and flipping. Unlike the existing methods, Majtner et al. propose to generate synthetic HEp-2 cell images by the GAN. Different from the traditional DA methods such as rotation and flipping, GAN produces images with different compositions than the original images. DA and its advantages are further discussed in Section 3.3.1. The proposed method in [43] trained recent CNN architectures such as GoogleNet with GAN-produced images to investigate their effectiveness. The findings were interesting. It was shown that the traditional DA methods, e.g., rotations, are more effective than that of GAN based DA. Furthermore, the authors mentioned that the GAN [54] is not robust against the large intra-class variation of data and as a result produced HEp-2 synthetic images of poor quality. The authors also emphasized on continuing the research of GAN to solve the problems of datasets with a limited number of annotated samples.

**Generic DNN based methods with partial changes in layers or training schemes.** The methods in this group make minor changes in the generic network architectures or use special training schemes for robust feature learning from HEp-2 images. Two existing methods [20, 16] fall under this category. The summary of both methods is given in Table 2.

The first method is proposed by Li et al. [20]. They propose to add two additional convolutional layers with $1 \times 1$ filters before the classification layers (i.e., FC and Softmax) of LeNet-5 [31] CNN architecture. Their motivation of using these additional layers is to increase the number of feature channels for the classification layers of the LeNet-5 model. The network is trained from scratch with a relatively large dataset compiled from both the ICPR 2014 Task-1 and Task-2 datasets (the details of Task-1 and Task-2 datasets are further discussed in Section 4). Note that this method primarily deals with the specimen classification and uses SL-HEP2IC based evaluation metric. However, due to its individual image based classification strategy and the performance on ICPR 2014 Task-1 dataset, we include this method under CL-HEP2IC. Their experimental results show that the proposed architecture outperforms the baseline LeNet-5 by a significant margin.

The second method is proposed by Lei et al. [16]. It propose to enhance the training and classification performance of the ResNet-50 architecture by combining the early layer predictions into the final classification layer. In particular, a parametric bridging mechanism is proposed to connect the early layers to the final layers. The experimental result shows that combining the classification decisions of early layers into the final layers has a positive impact on the final classification performance. Their method achieved state-of-the-art result in both the ICPR 2012 and the ICPR 2014 datasets. The major drawback of this method is that it significantly (almost double of the baseline) increases the number of network parameters.

**Customized DNN based methods.** A large number of customized DNN based methods [7, 9, 12, 6, 13, 21, 22, 17, 18, 25, 45] have been proposed in recent years, especially in the
last four years. While only a few of them made partial modifications to the generic DNN architectures, most of the methods use fully customized DNN architectures. Both are discussed below:

Methods that made partial modifications to the generic DNN models. Liu et al.’s method \cite{13} is one of the very few to make partial modifications to the CAE for HEP2IC. They combines an unsupervised CAE with supervised CNN classification branch that takes raw image data and predicts their class labels. It is a multi-task network optimized over the classification and reconstruction loss. Figure 5 shows the network proposed in \cite{13}. Denote $F_a$ as the unsupervised CAE auto-encoder model, $F_c$ as the supervised CNN classification model, and $W_a$ and $W_c$ are their learnable parameters, respectively. The network is optimized by the loss function $L$ as follows:

$$L_T = \lambda L_a(W_a) + L_c(W_c)$$

where $L_T$ is the total loss, $L_a$ is the reconstruction loss, $L_c$ is the classification loss, and $\lambda$ is the trade-off hyper-parameter. The idea of using an encoder where hidden layer features are able to reconstruct the output helps improve the feature representation in the classification layer, as demonstrated by the state-of-the-art classification result achieved by the proposed method on the ICPR 2012 dataset. Figure 6 shows some example inputs and reconstructed images produced by the CAE in the proposed method in \cite{13}. However, based on our experience, there could be some drawbacks of using CAE for classification, such as, (1) compared with CNN, it increases the number of training parameters in the network and (2) it tries to capture the maximum information from the manifold of training data which may be a problem if transfer learning to a slightly different dataset is performed.

Li et al. \cite{21} propose to combine the CNN with fully connected networks and extreme learning machines (ELMs). In the proposed method, the authors replaced the CNN softmax layer with ELMs. Furthermore, they have used multiple fully connected layers to combine the features from the final convolutional layers of a CNN into the ELMs (referred as multi-form feature extraction by the authors). The motivation behind this is to take advantage of the

Figure 5: Deep autoencoding-classification network (DACN) proposed in \cite{13}.

Figure 6: Input and reconstructed images by the CAE in \cite{13}. Top and bottom rows show the input and reconstructed images, respectively.
CNN’s hierarchical features. The obtained features are then used in training the ELMs for label prediction. The proposed method has a better feature generalization capacity than the CNN based methods as shown in an experiment where the proposed network is pre-trained using a grading hepatocellular carcinoma image dataset [21] but still achieves a decent (∼81%) classification performance on the ICPR 2014 dataset. However, it has more trainable parameters than the CNN, and some of the layers are designed carefully to guarantee the best feature representation for the task.

As mentioned by Lei et al. in [14], the training of ResNet hidden layers is very challenging. They often suffer from the vanishing gradient problem during the training with backpropagation. In [18], Shen et al. modify the original residual blocks as in [42] to increase their efficiency and capacity of overcoming the vanishing gradient problem. Similar to the original residual block, cross residual shortcuts between the blocks are used to further improve the features. The modification increases the depth of the residual block by two times at a lower cost with the number of parameters being reduced by 26.9%. Furthermore, the proposed method takes advantage of multi-scale feature extraction and fusion of shallow model predictions for improved feature learning. Figure 7 shows the DRI module proposed in [18].

Figure 7: Deep residual module (DRI) with three cross connections proposed in [18].

Li and Shen [22] propose deep residual inception network (DRI-net) by modifying the basic convolution blocks in the residual module of ResNet [42]. DRI-net integrates the key advantages of two high-performing CNN architectures, namely, Inception-net and ResNet. DRI-net has the following advantages over the ResNet and Inception-net, (1) multi-scale feature utilization in classification from Inception-net, (2) efficient network optimization from ResNet, and (3) improved classification by considering the auxiliary classifier decisions taken with the features from the early, middle, and end layers. The design of the original inception module is also modified for better feature representation and for overcoming the vanishing gradient problem when the number of layers in network increases. Precisely, batch normalization and parametric rectified linear unit (PReLU) were used before all the convolution layers, and two new identity shortcuts are added to the network. Unlike partially modified generic DNN based methods mentioned above, DRI-net is not suitable to train with smaller datasets and often requires longer training duration. The multi-scale convolutional component (MCC) module in DRI-net is primarily responsible for this issue.

In [17], the authors further improve the MCC module by (1) reducing the size of its convolution layers, (2) expanding the convolution from two to three scales, and (3) introducing a new global shortcut as a replacement of two identity shortcuts. The improved version of the MCC module is capable of extracting rich features at a higher speed which significantly improves the network training time. A new deep architecture named ‘HEp-Net’ based on this improved MCC module is proposed in [17]. The HEp-Net is very lightweight, i.e., only 7% of the DRI-net size (the number of parameters of DRI-net is about 1,985M), but is capable of learning rich features from smaller HEp-2 datasets. Figure 8 shows the architecture of the HEp-Net.

**Methods that rely on fully customized DNN models.** Malon et al. first attempted to use CNN for CL-HEP2IC in the ICPR 2012 HEp-2 Image Classification Competition [7]. They
Figure 8: Architecture of the HEp-Net proposed in [17]. The network is comprised of batch normalization, convolution, max-pooling, MCC block, network-in-network block, average pooling, and dropout operations. A joint feature representation of all the MCC blocks are used during training.

have proposed a simple CNN to perform CL-HEP2IC on the ICPR 2012 dataset. To deal with the illumination variations in the cell images, they have used absolute value rectification and subtractive spatial normalization in the CNN. Although the proposed method does not manage to achieve superior performance (i.e., 6th place), it outperforms many specially designed handcrafted methods [7] and later inspires many researchers to use the CNN for HEP2IC. The main issue of Malon et al.’s method is the neglect of the background surrounding the cell contours which is later proved to be important for separating similar cell classes [6] and the insufficient consideration to the common problems such as cell rotations.

Gao et al. [9, 6] first propose the successful DL based method for HEP2IC. They have used a LeNet-5 [31] inspired CNN architecture. The CNN is trained from scratch with the ICPR 2014 dataset and is used to perform image classification on the ICPR 2012 and the ICPR 2014 datasets. In their experiments, they observe the followings, (1) DA plays a crucial role in training the high-performing CNN for CL-HEP2IC. They observe a significant boost of classification performance when DA with rotations is applied during training, (2) the background of the HEp-2 cells is useful for the classification. They train a CNN with the mask of HEp-2 cell images and find that its performance is even lower than that of the CNN trained on HEp-2 images without taking advantage of the masks, and (3) the combined prediction of CNNs trained at different epochs generally gives better classification performance than that of a single CNN.

In the subsequent years, Jia et al. [12] propose a custom CNN model for CL-HEP2IC. The proposed model shares similarity with the VGG-M network [41], but uses more convolution operations. Furthermore, it uses dropout [40] to overcome the effect of over-fitting. The proposed model manages to achieve competitive performance as the previous model used in [9] with a similar cost of DA (i.e., the training images are rotated at a smaller angle, e.g., 18° for the ICPR 2014 dataset.) It is expected that the proposed network could be further improved with the use of more convolution and batch normalizations (BN) [55] operations.

A recent work by Ebrahim et al. [25] also proposes custom models. In particular, they propose two CNN models, one with a few layers and the other with more layers. Their models share similarity with Jia et al.’s model [12], but comparatively use a smaller number of layers. In their experimental investigation on network training, they have performed thorough experiments on various pre-processing techniques. One of the interesting findings in their study is the online DA (also be called as ‘on the fly DA’). Contrary to the traditional DA which performs DA operations on the training set before training, the online DA performs DA operations during training. However, the reported experimental results show that online DA significantly decreases the training performance. The possible reason could be the use of online DA which generates random data samples at each iteration of the training. This could affect the convergence of the optimization of the CNN, unless the training process is carefully managed. Meanwhile,
online DA could be beneficial if used wisely. For example, the training process can begin with a certain online DA and remain unchanged until the network is converged. After that, a new online DA could be implemented to train the network further to enhance the learned features. This process can continue until the validation loss becomes stable.

Figure 9: Positive and negative intensity images of Fine-speckled and Homogeneous HEp-2 image classes in SNPHEp-2 dataset [56]. The positive intensity images contain stronger cellular shape and higher illumination level, and the negative intensity images contain weaker cellular shape and lower illumination level.

Another recent work proposed by Vununu et al. [45] uses a four-stream CNN to learn local intensity and geometric information to deal with the heterogeneity problem occurring in HEp-2 cells. The intensity variations between the HEp-2 images sometimes characterize severe intra-class variations. Figure 9 shows images of two common HEp-2 classes, namely, fine speckled and homogeneous. For each class, positive and negative intensity images are shown. The positive intensity images have stronger cellular shape and higher illumination level. On the other hand, the negative intensity images have weaker cellular shape and lower illumination level. Note that shape and illumination are essential features for accurate classification of cells. Deterioration of shape and illumination in the negative images may significantly lower the intra-class variations which can cause serious confusion during the classification process. The proposed CNN in [45] extracts features from the following discrete wavelet transform (DWT) images, namely, horizontal detail coefficients, vertical detail coefficients, diagonal detail coefficients, and approximation coefficients with its four-streams to deal with the above problem. The detail and approximate coefficients of DWT are useful in minimizing the divergences between the images with positive and negative intensities. The detail coefficients can be regarded as the gradients in different directions. They provide a comprehensive understanding of the cellular shapes regardless of the intensity levels present in the original image. On the other hand, the approximation coefficient provides only a certain homogenization based on the intensity level present in the original image. Learning of the above coefficients in parallel enables the proposed network to extract useful features for the classification of HEp-2 images with different cellular shapes and intensity levels.

Similar to DWT, other image transformation methods such as fast Fourier transform, Haar wavelet transform, and directional gradient images could be used for analyzing the statistical distribution of image pixels. Also, binarized descriptors such as local binary pattern [57] and local phase quantization [58] could be used for this purpose. A further improvement of this method could be to use a shared single CNN stream to process the multiple coefficient images by treating them as a set of feature channels. Using this architecture could effectively reduce the total number of network parameters and this helps improve the generalization capability of the network.

**Summary of discussion.** A large number of CL-HEP2IC methods use DNN as a classifier. The initial methods are based on the generic DNNs that were designed for traditional image classification such as ImageNet. Motivated by the excellent results, some researchers have put efforts on the development of customized DNN architectures for CL-HEP2IC. Customized DNNs can extract more discriminative features than the generic DNNs. In terms of the DNN
architectures, CNN is the most popular choice in the literature. However, some recent methods have also used CAE for feature extraction.

Existing DNNs are prone to image rotation, cells structural deterioration, and illumination variation in the image. Although theoretically, pre-alignment, uniform-scaling, and image enhancement of cell samples seem to be the solution to the aforementioned problems, they are not very effective in practice [6]. As an alternative, most of the existing methods use DA to generate additional training data to make the CNN to learn features robust against these variations. However, with the special training mechanism such as that in [16], it is also possible to avoid excessive usage of DA such as rotating images at a very small angle. The DA is further discussed in Section 3.3.1.

From the perspective of DNN architecture, shallow CNN models are computationally efficient, but not so robust to the aforementioned variations. On the other hand, deeper CNNs such as ResNet have been trained from scratch by using limited samples in cell image benchmark datasets, with the minimum use of DA, e.g., [16, 45, 47]. Recent CL-HEP2IC methods are mostly based on the deeper CNN and CAE models. It is expected to see this ongoing trend in the future.

### 3.2 Specimen-level HEp-2 image classification (SL-HEP2IC) methods

Unlike the CL-HEP2IC methods, the SL-HEP2IC methods perform classification of HEp-2 specimen images. Based on how the specimen image is processed to obtain the classification results, existing SL-HEP2IC methods can be further decomposed into two major types, namely, single cell processing based SL-HEP2IC methods (SCP-SL-HEP2IC) and multi-cell processing based methods (MCP-SL-HEP2IC). Table 3 provides a comprehensive summary of both types of methods. It organizes the aforementioned methods based on their year of publication in ascending order. Furthermore, for each method, it lists the following key information: pre-processing techniques, basic DNN architectures, performance on the benchmark datasets, and remarks. Sections 3.2.1 and 3.2.2 provide a thorough discussion on them.

#### 3.2.1 Single cell processing based specimen-level HEp-2 image classification methods (SCP-SL-HEP2IC)

Given a specimen image, SCP-SL-HEP2IC methods decompose it into individual cell images by cell-level ground truth labels, i.e., bounding box annotations and segmentation masks. Next, each individual cell image is classified by a DNN. The classification results of each of the cell images are then accumulated to obtain the specimen-level result (e.g., via the majority voting strategy). SCP-SL-HEP2IC methods can also be regarded as the extension of CL-HEP2IC methods (discussed in Section 3.1) to the SL-HEP2IC. Based on the use of DNN, SCP-SL-HEP2IC methods can be further divided into two types, namely, DNN feature extraction based methods and true DNN based methods.

**Feature extraction based methods.** Similar to the methods discussed in Section 3.1.1, the DNN feature extraction based methods use a two-stage pipeline for the classification of individual cell images obtained from the input specimen image. In the first stage, the DNN feature is extracted from each individual cell image. Following that, in the second stage, the extracted features are classified using a traditional classifier, e.g., SVM. Li et al. [60] propose one of the very few methods to consider the extraction of DNN features for individual cell classification. They have used a modified LeNet-5 CNN model for feature extraction. Specifically, a few 1 \times 1 convolutional blocks have been added to the original LeNet-5 architecture to increase the depth of the feature channels. The modified LeNet-5 CNN model gives better performance than the original LeNet-5 model. Once the classification of individual cell images is done, a
### Table 3: Summary of deep HEp-2 specimen classification methods on the ICPR 2014 Task-2 dataset.

The widely accepted evaluation protocol for the ICPR 2014 dataset is MCA with LOSO protocol (refer to Section 4 for details). EN = Image enhancement method; CS = Contrast stretching; DA = Data augmentation; R = Rotation; F = Flipping; C = Cropping; M = Mirroring; SCP-SL-HEP2IC = Single cell processing based SL-HEP2IC methods; MCP-SL-HEP2IC = Multi-cell processing based SL-HEP2IC methods; * = Results were produced using ICPR 2014 Task-1 dataset.

| Ref. | Authors | Year | Pre-processing | Classifier | Results | Remarks |
|------|---------|------|----------------|------------|---------|---------|
| 1.   | Li et al. | 2016 | X R, M | Modified LeNet-5 | 83.55* | Use simple network, cross-dataset based training, prone to illumination variation. |
| 2.   | Li et al. | 2016 | X R | Modified LeNet-5 + SVM | 85.62* | Simple network, cell population histogram, prone to illumination and intra-class variations. |
| 3.   | Cascio et al. | 2019 | CS R | AlexNet+SVM | 93.75* | Two-stage framework, class-specific feature fusion, multi-stage training, use segmentation mask. |
| 1.   | Li et al. | 2016 | CS R, C | FCN | 90.89 | Single-shot prediction, multi-task network, low-resolution segmentation map. |
| 2.   | Li et al. | 2017 | CS R, M, C | Extended FCRN | 94.67 | Efficient network architecture, generalized features, low resolution segmentation map. |
| 3.   | Oraibi et al. | 2018 | X R | VGG-19+LBP + JML | 92.11 | Hybrid features, efficient framework, handcrafted features, prone to illumination variation. |
| 3.   | Xie et al. | 2019 | X R, M, C | DSFCN | 95.40 | Rich feature map, parametric fusion, prone to illumination variation and blur. |

The majority voting (MV) strategy is applied to the predicted cell labels to obtain the specimen label. The MV strategy selects the most dominant cell label as the specimen label, which works well on all the ICPR 2014 Task-2 dataset, except the Mitotic Spindle class. A majority of the Mitotic Spindle specimens are misclassified as the Golgi and Homogeneous specimens. To solve this issue, the authors of [60] propose to represent each specimen by a population histogram (PH). A PH is a simple histogram that describes the frequency of exiting individual cells in a specimen image. Figure 10 describes the pipeline for PH construction. From the point of view on features, it can be regarded as a high-level feature representation. The experimental results in [60] demonstrate that PH classifies the Mitotic Spindle images more accurately than the MV strategy.

**Figure 10:** Overview of the HEp-2 pattern histogram proposed in [60]. Given a specimen image, the individual HEp-2 cell images are extracted using the bounding box annotations. A CNN is then used to classify the individual cells. Based on the population of the predicted cell labels in the specimen, the cell pattern histogram (PH) is then constructed. Finally, an SVM is used for the classification of PH. Better viewed in color.

Since the PH is based on the individual cell image prediction by the DNN, adoption of a recent DNN model such as ResNet could help improve its discriminative capacity. Furthermore, in the proposed LeNet-5 based CNN model, the authors of [60] assume that it can well handle the
illumination variations, and do not use any image enhancement techniques. However, in other HEP2IC methods such as Gao et al. [6], image enhancement is used on LeNet-5 to obtain better invariance across the illumination variations. Hence, it is believed that the proposed LeNet-5 based CNN model in [60] could be further improved with the use of image enhancement.

In a recent work by Cascio et al. [46], the authors use a simpler approach to classify individual cell images extracted from a specimen sample. The basic working principle of this method is discussed as a part of CL-HEP2IC, and interested readers are referred to Section 3.1.1. As a side note, unlike Li et al.’s method [60], the reported performance of this method is based on the ICPR 2014 Task-1 dataset which is relatively small. The main drawback of this method is that it uses pre-trained features from an early CNN model [32]. The use of fine-tuned features from a more recent CNN model such as ResNet may further improve the performance of this method.

True DNN based methods. True or pure DNN based SCP-SL-HEP2IC employs the DNN as a classifier. They are very similar to the methods discussed in Section 3.1.2. An early approach is proposed by Li et al. [59]. They have employed a modified LeNet-5 CNN model to classify the individual cell images extracted from specimen images. Their modified LeNet-5 CNN model is similar to the model proposed in [60]. Even though the proposed CNN is simple, it achieved a good performance on the ICPR 2014 Task-1 dataset. The key to their good performance is the use of cross-dataset samples for training. Readers are referred to Section 4.5 for the details of cross-dataset training.

Summary of discussion. Feature extraction based methods are more popular for SCP-SL-HEP2IC. The key disadvantage of SCP-SL-HEP2IC method is their requirement of thorough classification of single cells in the specimen image to obtain the specimen label. SCP-SL-HEP2IC methods are more suitable for the cases where only a small number of single cells are present in a specimen image.

3.2.2 Multi-cell processing based specimen-level HEp-2 image classification methods (MCP-SL-HEP2IC)

Unlike the SCP-SL-HEP2IC methods described in Section 3.2.1, MCP-SL-HEP2IC based methods process an entire specimen image at a time. Hence, they are computationally more efficient. Based on their nature, they can be further decomposed into two types, namely, pixel-wise prediction based methods and image-wise prediction based methods.

Pixel-wise prediction based methods. Pixel-wise prediction based methods perform dense prediction on a specimen image such that each pixel of the specimen image is given a class label. The classification of the specimen is performed via a MV (i.e., majority voting) strategy on the predicted pixel labels. Due to the dense predictions, pixel-wise prediction based methods are capable of solving multiple tasks such as HEp-2 image classification and segmentation. However, since the focus of our review is HEP2IC, the concentration will be given only to discuss the methods that are associated with HEP2IC. The fully convolutional network (FCN) [64] is often used as a backbone in pixel-wise prediction based methods. The FCN is a variant of CNN that replaces the fully connected layers with convolution layers. Figure 11 shows the architecture of FCN used for the task of MCP-SL-HEP2IC.

Li et al. [61] are the first to consider FCN for SL-HEP2IC. They have simply adopted the FCN model proposed by Long et al. [64] for the task and managed to achieve good classification performance on the ICPR 2014 Task-2 dataset. FCN has the following key advantages, (1) it can process a specimen image in a single-shot, which saves a significant amount of computation time during both training and testing, (2) as opposed to the CNN, it can operate on arbitrary input sizes; hence, fixation of input specimen image size during training and testing is not necessary, and (3) it can make dense predictions which could be further used for cell segmentation. However, FCN suffers from low-resolution feature maps. Due to the propagation of an input
image through a stack of layers composed of convolution and pooling operations, FCN cannot provide high resolution (i.e., very precise object boundaries) feature maps. Low-resolution feature maps may not well represent cells captured at a smaller scale and can cause fuzzy object boundaries.

Following the success of FCN, in a subsequent of work by Li et al. [26], the authors propose the fully convolutional residual network (FCRN) [65] to further improve the classification performance. The authors extend the original FCRN by increasing its depth to nearly double (1.76 times) by replacing the bottleneck module with the residual in residual (RiR) module. The RiR module has more layers and identity shortcuts than the bottleneck module. Typically, the depth of FCRN is increased by adding more residual modules which are computationally expensive, i.e., more trainable parameters. Unlike that, the RiR modules allow the FCRN to increase its depth without requiring more residual modules. The proposed extended FCRN outperforms the baseline FCRN-50 and FCN, and achieves a good performance (i.e., 94.67%) on the ICPR 2014 dataset. Figure 12 shows some example feature maps learned by the RiR module in the proposed network. The extended FCRN is an effective and efficient approach for specimen classification. However, there are some cell classes such as homogeneous and mitotic spindle for which the proposed network suffers to achieve better classification accuracy. This could be due to their higher class similarities which confuse the classification.

A more recent work by Xie et al. [63] proposes a simple yet effective method to generate high-resolution feature maps by the FCN for SL-HEP2IC. They propose to combine the downsampled feature maps in the convolution layers with the upsampled feature maps in the deconvolution layers. However, unlike in the original FCN, they propose to use the skip connections between intermediate layers to preserve the rich object boundary details. Furthermore, a parametric feature fusion strategy is also proposed to better fuse the intermediate layer features. Experimental results on the ICPR 2014 Task-2 dataset demonstrate that the proposed method outperforms the above two methods by Li et al. [26]. However, the authors [63] agree that the proposed method is not robust against illumination variations and blurry image conditions.
**Image-wise prediction based methods.** Given a specimen image, the image-wise prediction based methods return a single label for the whole input image. A recent method proposed by Oraibi et al. [62] considers a simple approach to SL-HEP2IC. Unlike the pixel-wise prediction based SL-HEP2IC methods, the proposed method resizes the specimen image and classifies it with a VGG-19 CNN model [41]. Since the resizing operation deteriorates the local cell information, the authors propose to use handcrafted features such as LBP [58] and joint motif labels [66] to strengthen the discriminative capacity of CNN features. The experimental result shows that the proposed method surpasses the one proposed by Li et al. [61]. The followings could be used to improve the performance of the proposed method, (1) more discriminative CNN features such as ResNet-based features instead of VGG-based features and (2) gradient-based local shape features such as SIFT [49] instead of motif labels.

**Summary of discussion.** MCP-SL-HEP2IC methods are computationally more efficient than the SCP-SL-HEP2IC methods due to their multiple cell processing capacity. They are also more practical and scalable. While the pixel-wise prediction based methods are the most researched MCP-SL-HEP2IC methods, image-wise prediction based methods are also finding their way to SL-HEP2IC. However, due to the multi-tasking capability of pixel-wise prediction based methods, i.e., segmentation and classification, it is expected that most SL-HEP2IC research will be in this direction.

### 3.3 Common Discussion

This section will present two common discussions on the CL-HEP2IC and SL-HEP2IC methods discussed in Sections 3.1 and 3.2 respectively.

#### 3.3.1 Pre-processing techniques used in CL-HEP2IC and SL-HEP2IC methods

Pre-processing techniques have been widely used in the existing deep learning based HEP2IC methods. A list of pre-processing techniques that have been employed is available in Tables 2 and 3. Based on the type of operations, these techniques can be divided into two groups, namely, image enhancement techniques and data augmentation techniques. Both are discussed below:

**Image enhancement:** Image enhancement (IE) is a common pre-processing technique used in many computer vision tasks. In the existing deep learning based HEP2IC methods, IE has been used as an image normalization technique. Among many IE techniques, contrast stretching (CS) which scales the range of intensity values of an input image to a desired range of values is widely used. There are some papers such as [23, 24, 10] that consider other types of IE techniques such as histogram equalization. It is worth mentioning that the use of IE is popular among the early HEP2IC methods. Recent HEP2IC methods use more deep models and do not require IE to handle image intensity variations.

**Data augmentation:** DA is widely used in the existing deep learning based HEP2IC methods. Rotations (R), mirroring (M), flipping (F), and cropping (C) remain the most popular DA operations. These operations produce transformed images while keeping the semantic information same as in the original images. To obtain images with different compositions from the original image, generative adversarial network (GAN) [67] can be used. GAN produces synthesized images [43]. However, the use of GAN for DA in HEP2IC is still in its primary stage (to the best of our knowledge, only one work in this area has used GAN based DA [43]). DA can be applied to the HEp-2 datasets in two ways, (1) conventional way, and (2) class-aware way. In the conventional way, the DA operation is performed on every image classes of the HEp-2 datasets without looking at its distribution. The conventional way is not very effective for datasets that have small tails in its distribution. On the other hand, the class-aware way is
more effective for datasets with imbalanced class distributions such as the ICPR 2014 dataset. Specifically, it applies additional DA operations to the classes at the tail of the distribution to make the class sizes more balanced. The class-aware DA is widely used in the existing HEP2IC methods.

### 3.3.2 Evaluation strategies used in CL-HEP2IC and SL-HEP2IC methods

The common evaluation strategies used for the CL-HEP2IC and SL-HEP2IC methods are average class accuracy (ACA) and mean class accuracy (MCA), respectively. Both of these strategies are discussed in details in Section 4. In the case of CL-HEP2IC, all the datasets are split into training and test sets (optionally, validation set). Differently, the SL-HEP2IC commonly uses leave-one-specimen-out (LOSO) evaluation protocol for training and testing (i.e., possibly due to a limited number of samples available in the existing datasets). In addition, a few SL-HEP2IC methods also report their performance using the k-fold cross validation.

### 4 HEp-2 Public Datasets

Similar to other cell image datasets, the compilation of a dataset that contains HEp-2 images is costly, requiring special image acquisition equipment and expert judgments on data annotations. Due to this, at this moment, there are only a few datasets available publicly for the development of HEP2IC approaches. Figure 13 shows a sample of various HEp-2 image classes in existing datasets. The performance of the existing HEP2IC methods on various HEp-2 public datasets is given in Table 2 and Table 3. This section provides a review on the existing public HEp-2 datasets and their evaluation strategies widely used by the community.

![Sample categories available in existing HEp-2 public image datasets.](image)

Figure 13: Sample categories available in existing HEp-2 public image datasets.
4.1 ICPR 2012 dataset

The ICPR 2012 dataset[^1] is also known as the MIVIA HEp-2 images dataset, named after the MIVIA laboratory of the University of Salerno as the first public dataset released during ‘Contest on HEp-2 Cells Classification’ at ICPR 2012 [7, 68]. It has a total of 1,455 individual cell images distilled from twenty-eight 1,388×1,038 pixels color specimen samples and partitioned into 721 and 734 images for training and testing, respectively. Immunology experts annotate each specimen at cell-level. Based on the annotations, each cell image is classified into one of the following six categories, namely, Homogeneous, Coarse speckled, Fine speckled, Nucleolar, Centromere, and Cytoplasmic. Figure 13 shows sample images from these categories. For performance evaluation of this dataset, average classification accuracy (ACA) proposed by the authors in [68] is widely used in current HEP2IC methods. The ACA is calculated as follows:

\[
ACA = \frac{\text{Number of correctly predicted samples}}{\text{Number of total samples}}
\]

4.2 SNHEp-2 dataset

The Sullivan Nicolaides Pathology HEp-2 dataset[^2], in short, the SNPEp-2 dataset [56], is another public dataset used in the literature. In terms of the image classes, it has a higher similarity with the ICPR 2012 dataset and the AIDA dataset than the ICPR 2014 dataset. The only difference is that it does not have the Cytoplasmic and Nuclear dots class images. There are in total 1,884 individual monochrome cell images available in this dataset, from which 905 images are selected for training and 979 images are selected for testing purposes. A five-fold cross-validation strategy is applied to evaluate the performance of this dataset. Figure 13 shows sample images of this dataset. In comparison with the ICPR 2012 dataset, the SNPHEp-2 dataset is a less popular benchmarking dataset in the existing HEP2IC methods.

4.3 ICPR 2014 dataset

The ICPR 2014 dataset[^3] is also known as indirect immunofluorescence images (I3A) competition dataset. It was introduced in IIF image classification competition held at ICPR 2014 [27]. There exists two versions of this dataset, namely, Task-1 and Task-2 datasets. Task-1 is primarily designed for the CL-HEP2IC task and the Task-2 is designed for SL-HEP2IC. The Task-1 dataset has in total 13,596 monochrome single cell images extracted from 83 specimens using the bounding box annotations by experts. The images are divided into six classes, namely, Homogeneous, Speckled, Nucleolar, Centromere, Nuclear membrane (NuMem), and Golgi. Figure 13 shows sample images of these categories. Unlike the ICPR 2012 and the SNPHEp-2 datasets, the Coarse speckled and Fine speckled classes are represented using a single class called Speckled class. The NuMem and Golgi are new and not available in the above datasets. Furthermore, unlike the ICPR 2012 dataset, the test set is reserved by the organizer and is not publicly available. Since there is no test set available, current HEP2IC methods usually split the training set into 64% (8,701 images), 16% (2,175 images), and 20% (2,720 images) of the samples and use them for training, validation, and test purposes, respectively [6].

The Task-2 dataset has a total of 1,008 images, and each image is taken from four different locations of the 252 specimen samples, i.e., 252×4 = 1,008. The images in this dataset are distributed across seven classes, namely, Speckled (208 images), Nucleolar (200 images), Homogeneous (212 images), Mitotic spindle (60 images), Golgi (40 images), Nuclear membrane (64 images), and Centromere (204 images). Figure 14 shows the sample specimen images from

[^1]: Download link for the ICPR 2012 dataset: https://mivia.unisa.it/contest-hep-2/
[^2]: Download link for the SNPHEp-2 dataset: http://staff.itee.uq.edu.au/lovell/snphep2
[^3]: Download link for the ICPR 2014 dataset: https://hep2.unisa.it/dbtools.html
each class. In the ICPR 2014 competition, the participants trained their algorithms with the provided training set and submitted them to the organizers for testing. The test set was never released to the public. As previously mentioned, existing SL-HEP2IC methods reviewed in this paper widely apply the leave-one-specimen-out (LOSO) protocol to the training set for benchmarking.

![Sample specimen images of various classes](image)

Figure 14: Sample specimen images of various classes in the ICPR 2014 Task-2 dataset.

Mean class accuracy (MCA) is generally used for performance measure of both the Task-1 and Task-2 datasets. MCA is calculated by taking the mean of all class accuracies:

$$MCA = \frac{1}{C} \sum_{i=1}^{C} CA_i$$

(2)

where $CA_i$ is the classification accuracy of class $i$ and $C$ is the total number of image classes. The CA is defined as follows:

$$CA = \frac{\#\text{Number of correctly predicted samples}}{\#\text{Number of total samples in a class}}$$

4.4 AIDA dataset

The autoimmunity: diagnosis assisted by computer (AIDA) dataset\[69\] is a large-scale HEp-2 image dataset proposed as part of the AIDA project. The AIDA dataset is divided into two parts, namely, private and public. The AIDA private dataset has 20,000 samples, but it is only available to AIDA project partners. The AIDA public dataset has 2,080 specimen images and each specimen has (single or multiple) HEp-2 cells from 22 types of antinuclear antibody patterns including those in the previous datasets. The image size is not fixed and only specimen-level annotation by human experts is provided. One major difference of AIDA from the previous datasets is that it has negative intensity samples while others have only positive and weak positive samples. The positive, weak positive, and negative intensity images have higher, medium, and lower intensity levels. Figure 9 shows the examples of positive and negative intensity images. Specifically, it has 582 negative intensity samples and 1,498 positive intensity samples. Until now, this dataset has only been used for the HEp-2 specimen fluorescence intensity classification \[70\]. As a result, no evaluation protocol (e.g., the training, validation, and test image sets) exists for the HEp-2 pattern classification on this dataset until now. However, following the data partitioning scheme used in the ICPR 2014 dataset, the AIDA dataset may be divided into training, validation, and test splits for the HEP2IC task. Figure 13 shows some examples of the HEp-2 cell patterns in the AIDA public dataset.

\[\text{Download link for the AIDA dataset: http://www.aidaproject.net/index.php/it/downloads}\]
4.5 Cross-HEp-2 dataset recognition

Designing a high-performing classifier is challenging. In the conventional classification tasks, the test samples are always expected to have similar characteristics as training samples. However, in the case of cross-dataset classification, the classifier is trained with one dataset and tested on another dataset that has a slightly different data distribution. The classifier is expected to handle the change in data distribution and give a good classification performance. Existing HEp-2 datasets are not produced under the same laboratory settings, so the sample varies in terms of contrast, illumination variation, scale, and rotation. To perform cross-dataset classification, a classification system should well handle these issues. An easy way to deal with various image issues such as contrast and cellular shapes is to combine all the samples from the existing datasets to create a super training set, and use it to train the classifier [10, 20]. Nevertheless, this classifier would only perform well for the test samples that have similar characteristics with the training samples, which is not the goal of cross-dataset classification. One of the most suggested alternatives to the previous approaches is to train a classifier with a larger sized source dataset and then fine-tune it with a target dataset [13, 60, 26, 18, 17, 60, 59]. Note that the above method is limited to the datasets that share the similar set of class labels. A moderately deep network such as VGG-16 can be used for feature learning for this approach. Usually, the ICPR 2014 dataset is used for pre-training due to its larger size than the other datasets. The disadvantage of this approach is that it requires a larger source dataset for the training. However, customized training strategies such as the cross-modal transfer learning [16] could be used to partially resolve the larger source dataset issue.

5 Discussions and Future Trends

5.1 Overview

This paper has reviewed the literature of HEP2IC until October 2019. There is a total of 31 papers currently available that use DL for HEP2IC. Although the use of DL began for HEP2IC in 2013, most of the papers were published in the last four years (2016-2019). At the early stage, the use of DL was only limited to CL-HEP2IC. However, in the recent years, DL has been successfully applied to SL-HEP2IC. The networks used in the existing CL-HEP2IC and SL-HEP2IC methods are covered. Only a few methods (see Section 3.1.1) currently use pre-trained deep networks as feature extractor, which can be regarded as the extension of handcrafted methods. The majority of the methods (see Tables 2 and 3), especially those published in the last five years, use end-to-end trainable deep architectures. It can be confidently said that the use of such networks has become an established practice.

5.2 Key features of most successful deep HEP2IC methods

After conducting the review of the existing literature of CL-HEP2IC and SL-HEP2IC, we have noticed that designing a perfect network architecture may not be the research focus in achieving higher HEp-2 image classification performance. It has been seen that even with the established general-purpose deep networks, some authors have managed to achieve the state-of-the-art results [14]. Scrutinizing the methods that obtain good classification performance, we can see that they mostly go beyond the culture of just deepening the network by adding more layers. In particular, they look into the areas of deep learning such as data augmentation and data enhancement [23]. At the same time, data augmentation and enhancement are not the only keys to achieve good performance. Some authors have suggested to use special mechanisms such as multi-scale feature extraction [22, 17], knowledge sharing between layers [16], and training with transformed images [47, 45].

25
5.3 Unique challenges of HEp-2 image classification

HEP2IC using DL faces some special challenges. In the existing literature, the limitation of sample size in the current HEp-2 datasets is often mentioned as a great challenge. However, the recent studies [17, 16] show that it is only partially valid. With the efficient design of deep architecture or a little tweak in generic DNN models, one could easily train a high-performing HEP2IC system with smaller datasets. Instead, the greater challenge with the current datasets is dealing with various imaging conditions. Although the current datasets are collected in laboratories under controlled conditions, their images present significant challenges for deep networks. The acquisition of HEp-2 datasets is carried out in several stages, and usually, in batches. Due to the manual staining process, the samples among different batches vary in terms of lighting conditions and staining pattern strengths (the strengths of HEp-2 pattern depend on the serum sample which is subjective). Also, the HEp-2 datasets are often collected in collaborative environments. Many research groups from different geographic locations actively participate in the image acquisition, where each group collects distinctive blood serum samples for the HEp-2 staining. The HEp-2 staining patterns collected from these samples may significantly increase the inter-class variability of the dataset. Most of the existing methods deal with the various image acquisition conditions by applying image enhancement techniques such as contrast stretching. However, some recent methods have considered other techniques such as gradient image [47] and DWT transformation [45].

Another challenging issue in existing HEp-2 datasets lies at that they may not contain a similar number of samples for all classes. Some classes have a smaller number of samples, i.e., tail classes, than most of the other classes. Training DNNs with such datasets could focus too much on learning robust features for the dominant classes. The minor classes receive less attention during training and this may adversely affect their classification. One of the effective solutions proposed by the existing methods to this problem is class-balanced data augmentation [22]. It is expected that more efficient solutions would be developed with the research progress on HEP2IC.

Lastly, the labeling of HEp-2 images is another challenging task. There are two types of labeling currently used in the existing datasets, namely, cell-level labeling and specimen-level labeling. Usually, several immunology experts take part in the HEp-2 image labeling process to achieve observer-independent annotations since it is very common to label the same HEp-2 image differently by different experts [7, 8]. This is a time-consuming process and usually takes 2-3 years for a dataset [7, 58, 28]. The ICPR 2012, SNPHEp-2, and ICPR 2014 datasets have cell-level annotations, whereas the AIDA dataset only has specimen-level annotations. Although the labeling challenge is primarily concerned with HEp-2 dataset collection and compilation, labeling could have significant impact on the HEP2IC.

5.4 Future trends of deep learning in HEp-2 image classification

Current HEP2IC methods tackle the challenges from the existing datasets mentioned above. However, there are still some limitations that would be possible to overcome in the future following the ongoing progress of DL research. Currently, there are only three datasets available publicly for HEP2IC. Although the existing HEP2IC methods use data augmentation to increase the samples in the datasets, we must agree that these data augmentations are largely simple image transformations and they are not efficient in exploring the true distribution of HEp-2 cell images. Synthesized image generation (e.g., with the GAN technique) could be a better solution to this issue. However, synthesized image generation for HEP2IC is still in its beginning stage. To the best of our knowledge, at present, only one work has explored the use of GAN based DA for HEP2IC [43]. It is expected that in the future more datasets could be released for public use. For example, following the success of recent HEP2IC competitions [7, 8], new competitions using large-scale HEp-2 image datasets could be further organized to encourage researchers to
continue to focus on the HEP2IC issues.

From the network design perspective, the methods published between 2017 and 2019 are slightly different from the earlier ones. They are more focused on reducing the number of network parameters [22], designing the networks to be trainable with smaller datasets [17], and avoiding the excessive use of data pre-processing [10]. It is expected that the trend of designing computationally efficient HEP2IC methods will continue to grow in the future. As aforementioned, labeling images is one of the most challenging tasks during the compilation of the current HEp-2 datasets [7, 8]. Weakly-supervised, self-supervised, and unsupervised learning are the three promising areas of deep learning which could help overcoming this labeling challenge. While weakly-supervised methods require some initial label information, self-supervised and unsupervised methods could be conducted without using label information. One of the unsupervised learning methods, CAE, has been successfully used in HEP2IC for unsupervised feature learning, and it shows promising results [13, 47]. In the coming years, it is expected that more research will be conducted in these areas to address the labeling limitation.

Lastly, at this moment, deep learning methods are largely treated as ‘black boxes’ in the literature. As indicated in [33], having just a good classification system is not sufficient in medicine. There are substantial risks of legal issues involved in the whole process. Repeatability is another issue which would make practitioners of the field to rely on deep networks results. Among the existing HEP2IC methods, Lei et al. [16] and Xie et al. [63] have tried to understand what the intermediate layers of CNN and FCN were learning, respectively, and proposed to integrate their features into the final layers. It is expected that in the future more methods will be developed to understand the response of various parts of deep networks with respect to HEp-2 cell images. From the above discussions, we can conclude that deep learning has dramatically impacted the progress of the HEP2IC and more exciting results are waiting ahead.

References

[1] M. Fritzler, Immunofluorescent antinuclear antibody tests, Manual of Clinical Laboratory Immunology. ASM, Washington (1986) 733–739.

[2] R. Hiemann, N. Hilger, J. Michel, J. Nitschke, A. Böhm, U. Anderer, M. Weigert, U. Sack, Automatic analysis of immunofluorescence patterns of HEp-2 cells, Annals of the New York Academy of Sciences 1109 (1) (2007) 358–371.

[3] S. Manivannan, W. Li, S. Akbar, R. Wang, J. Zhang, S. J. McKenna, An automated pattern recognition system for classifying indirect immunofluorescence images of HEp-2 cells and specimens, Pattern Recognition 51 (2016) 12–26.

[4] P. Perner, H. Perner, B. Müller, Mining knowledge for HEp-2 cell image classification, Artificial Intelligence in Medicine 26 (1-2) (2002) 161–173.

[5] L. Liu, L. Wang, HEp-2 cell image classification with multiple linear descriptors, Pattern Recognition 47 (7) (2014) 2400–2408.

[6] Z. Gao, L. Wang, L. Zhou, J. Zhang, HEp-2 cell image classification with deep convolutional neural networks, IEEE Journal of Biomedical and Health Informatics 21 (2) (2017) 416–428.

[7] P. Foggia, G. Percannella, P. Soda, M. Vento, Benchmarking HEp-2 cells classification methods, IEEE Transactions on Medical Imaging 32 (10) (2013) 1878–1889.

[8] P. Hobson, B. C. Lovell, G. Percannella, M. Vento, A. Wiliem, Benchmarking human epithelial type 2 interphase cells classification methods on a very large dataset, Artificial Intelligence in Medicine 65 (3) (2015) 239–250.
[9] Z. Gao, J. Zhang, L. Zhou, L. Wang, HEp-2 cell image classification with convolutional neural networks, in: 1st Workshop on Pattern Recognition Techniques for Indirect Immunofluorescence Images (I3A), IEEE, 2014, pp. 24–28.

[10] N. Bayramoglu, J. Kannala, J. Heikkilä, Human epithelial type 2 cell classification with convolutional neural networks, in: 15th International Conference on Bioinformatics and Bioengineering (BIBE), IEEE, 2015, pp. 1–6.

[11] H. T. H. Phan, A. Kumar, J. Kim, D. Feng, Transfer learning of a convolutional neural network for HEp-2 cell image classification, in: 13th International Symposium on Biomedical Imaging (ISBI), IEEE, 2016, pp. 1208–1211.

[12] X. Jia, L. Shen, X. Zhou, S. Yu, Deep convolutional neural network based HEp-2 cell classification, in: 23rd International Conference on Pattern Recognition (ICPR), IEEE, 2016, pp. 77–80.

[13] J. Liu, B. Xu, L. Shen, J. Garibaldi, G. Qiu, HEp-2 cell classification based on a deep autoencoding-classification convolutional neural network, in: 14th International Symposium on Biomedical Imaging (ISBI), IEEE, 2017, pp. 1019–1023.

[14] H. Lei, T. Han, W. Huang, J. Y. Kuo, Z. Yu, X. He, B. Lei, Cross-modal transfer learning for HEp-2 cell classification based on deep residual network, in: IEEE International Symposium on Multimedia (ISM), IEEE, 2017, pp. 465–468.

[15] M. Lu, L. Gao, X. Guo, Q. Liu, J. Yin, HEp-2 cell image classification method based on very deep convolutional networks with small datasets, in: Ninth International Conference on Digital Image Processing (ICDIP), Vol. 10420, International Society for Optics and Photonics, 2017, p. 1042040.

[16] H. Lei, T. Han, F. Zhou, Z. Yu, J. Qin, A. Elazab, B. Lei, A deeply supervised residual network for HEp-2 cell classification via cross-modal transfer learning, Pattern Recognition 79 (2018) 290–302.

[17] Y. Li, L. Shen, HEp-Net: A smaller and better deep-learning network for HEp-2 cell classification, Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization (2018) 1–7.

[18] L. Shen, X. Jia, Y. Li, Deep cross residual network for HEp-2 cell staining pattern classification, Pattern Recognition 82 (2018) 68–78.

[19] X.-H. Han, J. Lei, Y.-W. Chen, HEp-2 cell classification using K-support spatial pooling in deep CNNs, in: Deep Learning and Data Labeling for Medical Applications, Springer, 2016, pp. 3–11.

[20] H. Li, J. Zhang, W.-S. Zheng, Deep CNNs for HEp-2 cells classification: A cross-specimen analysis, arXiv preprint arXiv:1604.05816.

[21] S. Li, H. Jiang, W. Pang, Joint multiple fully connected convolutional neural network with extreme learning machine for hepatocellular carcinoma nuclei grading, Computers in Biology and Medicine 84 (2017) 156–167.

[22] Y. Li, L. Shen, A deep residual inception network for HEp-2 cell classification, in: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Springer, 2017, pp. 12–20.
[23] L. F. Rodrigues, M. C. Naldi, J. F. Mari, Exploiting Convolutional Neural Networks and preprocessing techniques for HEp-2 cell classification in immunofluorescence images, in: 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), IEEE, 2017, pp. 170–177.

[24] L. F. Rodrigues, M. C. Naldi, et al., HEp-2 cell image classification based on convolutional neural networks, in: Workshop of Computer Vision (WVC), IEEE, 2017, pp. 13–18.

[25] M. Ebrahim, M. Alsmirat, M. Al-Ayyoub, Performance study of augmentation techniques for HEp-2 CNN classification, in: 9th International Conference on Information and Communication Systems (ICICS), IEEE, 2018, pp. 163–168.

[26] Y. Li, L. Shen, S. Yu, HEp-2 specimen image segmentation and classification using very deep fully convolutional network, IEEE Transactions on Medical Imaging 36 (7) (2017) 1561–1572.

[27] B. C. Lovell, G. Percannella, M. Vento, A. Wiliem, Performance evaluation of indirect immunofluorescence image analysis systems, in: ICPR 2014, 2014.

[28] B. C. Lovell, G. Percannella, A. Saggese, M. Vento, A. Wiliem, International contest on pattern recognition techniques for indirect immunofluorescence images analysis, in: 2016 23rd International Conference on Pattern Recognition (ICPR), IEEE, 2016, pp. 74–76.

[29] R. Nosaka, K. Fukui, HEp-2 cell classification using rotation invariant co-occurrence among local binary patterns, Pattern Recognition 47 (7) (2014) 2428–2436.

[30] L. Shen, J. Lin, S. Wu, S. Yu, HEp-2 image classification using intensity order pooling based features and bag of words, Pattern Recognition 47 (7) (2014) 2419–2427.

[31] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proceedings of the IEEE 86 (11) (1998) 2278–2324.

[32] A. Krizhevsky, I. Sutskever, G. E. Hinton, ImageNet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.

[33] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, C. I. Sánchez, A survey on deep learning in medical image analysis, Medical Image Analysis 42 (2017) 60–88.

[34] Y. LeCun, K. Kavukcuoglu, C. Farabet, Convolutional networks and applications in vision, in: Proceedings of IEEE International Symposium on Circuits and Systems, IEEE, 2010, pp. 253–256.

[35] J. Masci, U. Meier, D. Cireşan, J. Schmidhuber, Stacked convolutional auto-encoders for hierarchical feature extraction, in: International Conference on Artificial Neural Networks, Springer, 2011, pp. 52–59.

[36] P. Foggia, G. Percannella, A. Saggese, M. Vento, Pattern recognition in stained HEp-2 cells: Where are we now?, Pattern Recognition 47 (7) (2014) 2305–2314.

[37] F. Xing, Y. Xie, H. Su, F. Liu, L. Yang, Deep learning in microscopy image analysis: A survey, IEEE Transactions on Neural Networks and Learning Systems (99) (2017) 1–19.

[38] W. Rawat, Z. Wang, Deep convolutional neural networks for image classification: A comprehensive review, Neural Computation 29 (9) (2017) 2352–2449.
[39] Z. Zhao, P. Zheng, S. Xu, X. Wu, Object Detection With Deep Learning: A Review, IEEE Transactions on Neural Networks and Learning Systems (2019) 1–21.

[40] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, The Journal of Machine Learning Research 15 (1) (2014) 1929–1958.

[41] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556.

[42] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[43] T. Majtner, B. Bajić, J. Lindblad, N. Sladoje, V. Blanes-Vidal, E. S. Nadimi, On the Effectiveness of Generative Adversarial Networks as HEp-2 Image Augmentation Tool, in: Scandinavian Conference on Image Analysis, Springer, 2019, pp. 439–451.

[44] L. D. Nguyen, R. Gao, D. Lin, Z. Lin, Biomedical image classification based on a feature concatenation and ensemble of deep CNNs, Journal of Ambient Intelligence and Humanized Computing (2019) 1–13.

[45] C. Vununu, S.-H. Lee, O.-J. Kwon, K.-R. Kwon, A Dynamic Learning Method for the Classification of the HEp-2 Cell Images, Electronics 8 (8) (2019) 850.

[46] D. Cascio, V. Taormina, G. Raso, Deep CNN for IIF Images Classification in Autoimmune Diagnostics, Applied Sciences 9 (8) (2019) 1618.

[47] C. Vununu, S.-H. Lee, K.-R. Kwon, A Deep Feature Extraction Method for HEp-2 Cell Image Classification, Electronics 8 (1) (2019) 20.

[48] A. Sharif Razavian, H. Azizpour, J. Sullivan, S. Carlsson, CNN features off-the-shelf: An astounding baseline for recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2014, pp. 806–813.

[49] D. G. Lowe, et al., Object recognition from local scale-invariant features, in: ICCV, Vol. 99, 1999, pp. 1150–1157.

[50] J. Yang, Y.-G. Jiang, A. G. Hauptmann, C.-W. Ngo, Evaluating bag-of-visual-words representations in scene classification, in: Proceedings of the International Workshop on Multimedia Information Retrieval, ACM, 2007, pp. 197–206.

[51] H. Peng, F. Long, C. Ding, Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy, IEEE Transactions on Pattern Analysis & Machine Intelligence (8) (2005) 1226–1238.

[52] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, J. Liang, Convolutional neural networks for medical image analysis: Full training or fine tuning?, IEEE Transactions on Medical Imaging 35 (5) (2016) 1299–1312.

[53] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.

[54] A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, arXiv preprint arXiv:1511.06434.
[55] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, arXiv preprint arXiv:1502.03167.

[56] A. Wiliem, Y. Wong, C. Sanderson, P. Hobson, S. Chen, B. C. Lovell, Classification of human epithelial type 2 cell indirect immunofluorescence images via codebook based descriptors, in: IEEE Workshop on Applications of Computer Vision (WACV), IEEE, 2013, pp. 95–102.

[57] T. Ahonen, A. Hadid, M. Pietikainen, Face description with local binary patterns: Application to face recognition, IEEE Transactions on Pattern Analysis & Machine Intelligence (12) (2006) 2037–2041.

[58] V. Ojansivu, J. Heikkilä, Blur insensitive texture classification using local phase quantization, in: International Conference on Image and Signal Processing, Springer, 2008, pp. 236–243.

[59] H. Li, J. Zhang, W.-S. Zheng, Deep CNNs for HEp-2 cells classification: A cross-specimen analysis, arXiv preprint arXiv:1604.05816.

[60] H. Li, H. Huang, W.-S. Zheng, X. Xie, J. Zhang, HEp-2 specimen classification via deep CNNs and pattern histogram, in: 23rd International Conference on Pattern Recognition (ICPR), 2016, pp. 2145–2149.

[61] Y. Li, L. Shen, X. Zhou, S. Yu, HEp-2 specimen classification with fully convolutional network, in: 23rd International Conference on Pattern Recognition (ICPR), IEEE, 2016, pp. 96–100.

[62] Z. A. Oraibi, H. Yousif, A. Hafiane, G. Seetharaman, K. Palaniappan, Learning local and deep features for efficient cell image classification using random forests, in: 25th IEEE International Conference on Image Processing (ICIP), IEEE, 2018, pp. 2446–2450.

[63] H. Xie, H. Lei, Y. He, B. Lei, Deeply supervised full convolution network for HEp-2 specimen image segmentation, Neurocomputing 351 (2019) 77–86.

[64] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.

[65] Z. Wu, C. Shen, A. v. d. Hengel, High-performance semantic segmentation using very deep fully convolutional networks, arXiv preprint arXiv:1604.04339.

[66] V. S. Prasath, Y. M. Kassim, Z. A. Oraibi, J.-B. Guiriec, A. Hafiane, G. Seetharaman, K. Palaniappan, HEp-2 cell classification and segmentation using motif texture patterns and spatial features with random forests, in: 23rd International Conference on Pattern Recognition (ICPR), IEEE, 2016, pp. 90–95.

[67] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Advances in Neural Information Processing Systems, 2014, pp. 2672–2680.

[68] Contest on HEp-2 Cells Classification, URL http://mivia-web.diem.unisa.it/hep2contest

[69] A. Benammar Elgaaied, D. Cascio, S. Bruno, M. C. Ciaccio, M. Cipolla, A. Fauci, R. Morgante, V. Taormina, Y. Gorgi, R. Marrakchi Triki, et al., Computer-assisted classification patterns in autoimmune diagnostics: the AIDA project, BioMed Research International 2016, Art. ID 2073076.
[70] D. Cascio, V. Taormina, G. Raso, Deep Convolutional Neural Network for HEP-2 Fluorescence Intensity Classification, Applied Sciences 9 (3) (2019) 408.