Computers in Biology and Medicine 138 (2021) 104932

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

A novel self-learning framework for bladder cancer grading using histopathological images

Gabriel García [a,*], Anna Esteve [a,b], Adrián Colomer [a], David Ramos [b], Valery Naranjo [a]

[a] Instituto de Investigación e Innovación en Bioingeniería, Universitat Politècnica de València, 46022, Valencia, Spain
[b] Hospital Universitario y Politécnico La Fe, Avinguda de Fernando Abril Martorell, 106, 46026, Valencia, Spain

ARTICLE INFO

Keywords:
Bladder cancer
Tumour budding
Unsupervised learning
Deep clustering
Histopathological images
Self-learning
Immunohistochemical staining

ABSTRACT

In recent times, bladder cancer has increased significantly in terms of incidence and mortality. Currently, two subtypes are known based on tumour growth: non-muscle invasive (NMIBC) and muscle-invasive bladder cancer (MIBC). In this work, we focus on the MIBC subtype because it has the worst prognosis and can spread to adjacent organs. We present a self-learning framework to grade bladder cancer from histological images stained by immunohistochemical techniques. Specifically, we propose a novel Deep Convolutional Embedded Attention Clustering (DCEAC) which allows for the classification of histological patches into different levels of disease severity, according to established patterns in the literature. The proposed DCEAC model follows a fully unsupervised two-step learning methodology to discern between non-tumour, mild and infiltrative patterns from high-resolution 512 × 512 pixel samples. Our system outperforms previous clustering-based methods by including a convolutional attention module, which enables the refinement of the features of the latent space prior to the classification stage. The proposed network surpasses state-of-the-art approaches by 2–3% across different metrics, reaching a final average accuracy of 0.9034 in a multi-class scenario. Furthermore, the reported class activation maps evidence that our model is able to learn by itself the same patterns that clinicians consider relevant, without requiring previous annotation steps. This represents a breakthrough in MIBC grading that bridges the gap with respect to training the model on labelled data.

1. Introduction

Bladder cancer arises from uncontrolled proliferation of the urothelial bladder cells, which leads to tumour development. A significant increase in adult incidence and mortality has been observed during the last several years in relation to this condition. Recent studies state that bladder cancer is the second most common urinary tract cancer and the fifth most prevalent among men in developed countries [1,2]. Nowadays, the diagnostic procedure for bladder cancer involves several time-consuming tests. First, urine cytology is performed to determine the presence of cancer cells [3]. Subsequently, vesico-prostatic and renal ultrasound are employed to locate the tumour and assess the type of growth, which can be used to determine the grade and prognosis of the patient. If the tumour cannot be located at the previous stage, MRI urography is carried out to analyse possible local spread [4]. If there is evidence of bladder cancer, the urologist usually performs a cystoscopy based on the transurethral resection technique [5], which allows for the extraction of a sample of abnormal bladder tissue to determine the type of tumour growth. After the preparation process, the biopsied tissue is usually stained with hematoxylin and eosin (H&E) to enhance its histological properties. Finally, an additional staining process can be adopted to highlight special structures associated with the problem under study. The immunohistochemical CK AE1/3 technique was applied on the histological images used in this work to highlight the cancer cells by providing a brown hue when the antigen-antibody binding occurs. The two kinds of bladder cancer, non-muscle invasive (NMIBC) and muscle-invasive (MIBC), are distinguished depending on the level of invasion of tumour growth within the bladder wall. Currently, 75% and 25% of bladder cancer cases correspond to NMIBC and MIBC, respectively [2]. In this study, we focus on the MIBC category as it has the worst prognosis and favours tumour dissemination to adjacent organs. According to Ref. [6], MIBC does not usually present low-grade malignancy, but rather high-grade urothelial carcinomas. Following the classification criteria proposed by the World Health Organisation (WHO) [7], these can be classified as grade 2 or 3. Jimenez et al. [8] described three different histological patterns which correlate

* Corresponding author.
E-mail address: jogarpa7@i3b.upv.es (G. García).

https://doi.org/10.1016/j.compbiomed.2021.104932
Received 15 June 2021; Received in revised form 7 October 2021; Accepted 7 October 2021
Available online 12 October 2021

with the patient outcome. Specifically, histopathological images stained with CK AE1/3 were annotated by a pathologist with more than 20 years of expertise considering nodular, trabecular and infiltrative patterns, as shown in Fig. 1. The nodular pattern (yellow box) is defined by the presence of well-delineated, circular nests of tumour cells. The trabecular pattern is characterised by the presence of tumour cells arranged in interconnected bands. The infiltrative pattern, also known as tumour budding, is composed of cords of tumour cells (red box) or a small cluster of isolated cells called *buds* (blue box). The infiltrative pattern represents the most aggressive scenario and the worst prognosis for the patient [9–12]. Therefore, we combined nodular and trabecular structures into a single specific class (mild pattern) to grade the severity of MIBC according to the prognosis of the disease. We also considered a non-tumour pattern (pink box) to cover cases where the patient shows no signs of tumour. Thus, a multi-class scenario is conducted throughout the paper to classify bladder cancer into non-tumour (NT), mild (M) and infiltrative (I) patterns.

1.1. Related work

Accurate diagnosis of bladder cancer is a time-consuming task for expert pathologists and lacks reproducibility, leading to significant differences in histological interpretation [13,14]. Many state-of-the-art studies have proposed artificial intelligence algorithms to assist pathologists in terms of cost-effectiveness and subjectivity ratio. Most of these approaches focused on machine learning techniques applied to H&E-stained histological images for segmentation [15–17] and classification [13,14,18–23].

Beginning with the segmentation-based studies, Lucas et al. [15] used the popular U-Net architecture to segment normal and malignant cases of bladder images. They then used the common VGG16 network [24] as a backbone to extract histological features from patches of 224 × 224 pixels. The resulting features were combined with other clinical data to carry out a classification step using bidirectional GRU networks [25]. The proposed algorithm reported an accuracy of 0.67 for 5-year survival prediction. In Ref. [16], the authors carried out an end-to-end approach to discern between MIBC and NMIBC categories from H&E images. First, they performed a segmentation process to distinguish tissue from image background. Patches of 700 × 700 pixels were used to perform both manual and automatic feature extraction. The hand-crafted learning was conducted via contextual features such as nuclear size distribution, crack edge, sample ratio, etc., whereas the data-driven learning was conducted using the VGG16 and VGG19 architectures. During the classification stage, different machine learning classifiers such as support vector machine (SVM), logistic regression (LR) and random forest (RF), among others, were used to determine the bladder tissue type. Manual approaches showed superior performance over deep-learning models, reaching an accuracy of 91–96% depending

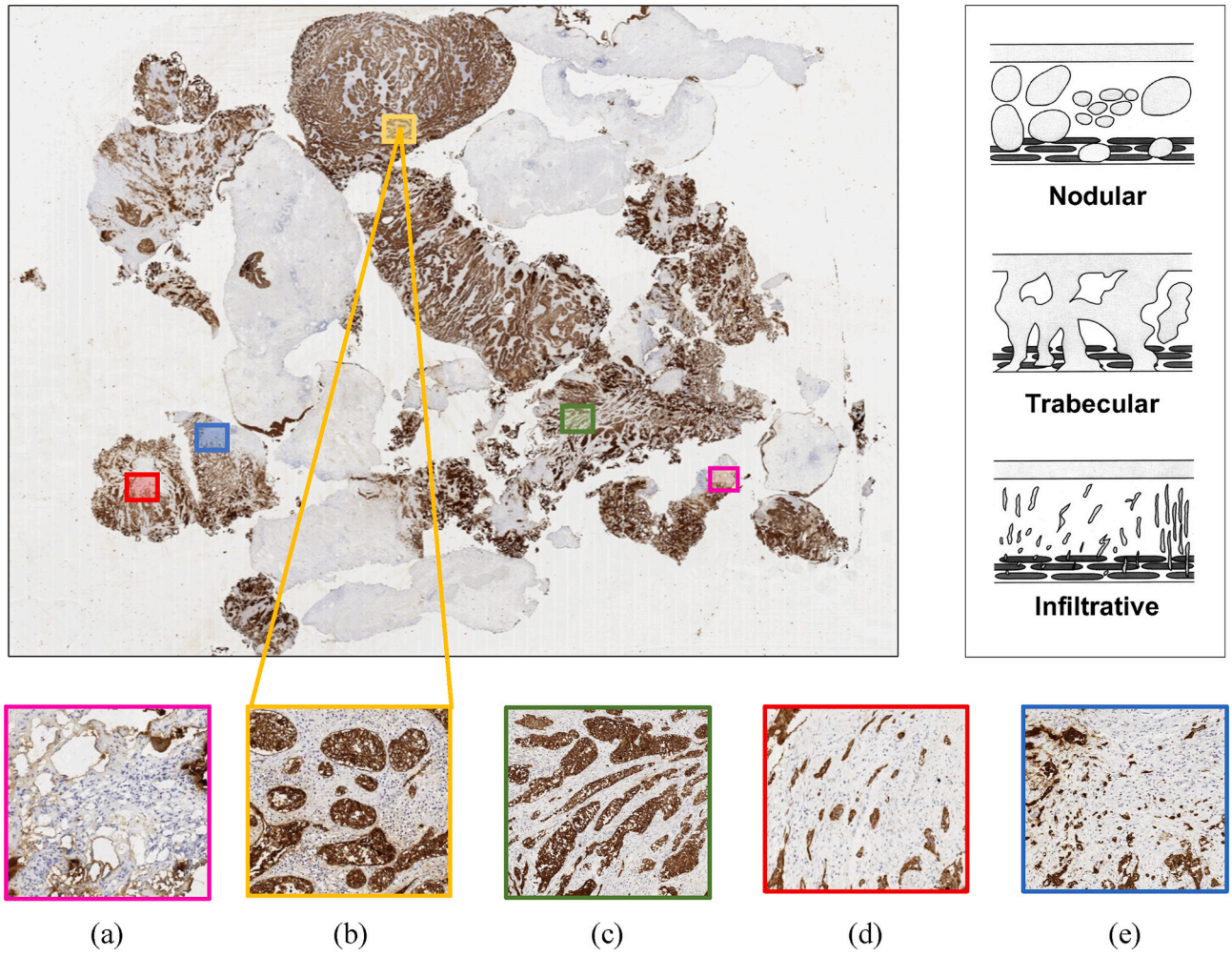

Fig. 1. The larger image corresponds to a Whole-Slide Image (WSI) of a patient suffering from muscle-invasive bladder cancer (MIBC). The top right of the figure (modified from Ref. [8]) is a diagrammatic representation of the theoretical arrangement of the patterns. The patches marked with colours denote different growth patterns. Specifically: (a) non-tumour pattern, (b) nodular arrangement (mild pattern), (c) trabecular arrangement (mild pattern), (d) tumour cell cords of an infiltrative pattern and (e) isolated tumour cells corresponding to an infiltrative pattern.

on the classifier.

Most of the classification-intended studies also focused on H&E-stained histological images, as in the segmentation frameworks. In Ref. [18], researchers proposed a multi-class scenario to detect the molecular subtype in MIBC cases. They applied the ResNet architecture to patches of 512 × 512 pixels, achieving results for the area under the ROC curve (AUC) of 0.89 and 0.87 in terms of micro- and macro-average, respectively. In Ref. [19], the authors made use of the Xception network as a feature extractor from H&E-stained patches of 256 × 256 pixels. An SVM classifier was then implemented to discern between high and low mutational burden, reaching values of 0.73 and 0.75 for accuracy and AUC, respectively. Harmon et al. [20] proposed a classification scenario to detect lymph node metastases from H&E patches of 100 × 100 pixels. A combination of the ResNet-101 architecture with AdaBoost classifiers reported an AUC of 0.678 at test time. Another study [13] carried out a classification approach to categorise tissue type into six different classes: urothelium, stroma, damaged, muscle, blood and background. To this end, the authors combined supervised and unsupervised deep learning techniques on patches of 128 × 128 pixels stained with H&E. Specifically, they trained an autoencoder (AE) from the unlabelled images and used the encoder network to address the classification through features extracted from the labelled samples. They achieved multi-class scores of 0.936, 0.935 and 0.934% for precision, recall and F1-score metrics, respectively. One of the most prominent state-of-the-art studies (focusing on H&E-stained histological images of bladder cancer) was conducted in Ref. [14]. In this study, Zhang et al. compiled a large database of Whole-Slide Images (WSIs) with the aim of discerning between low and high grades of disease. They used an autoencoder network to identify possible areas with cancer. They then fed 1024 × 1024 regions of interest (ROIs) into a Convolutional Neural Network (CNN) for classification into low and high classes. The proposed system obtained an average accuracy of 94%, compared to 84.3% achieved by pathologists. The findings from this study reveal that there exists a significant subjectivity among experts in diagnosing from histological images of bladder cancer, as discussed in Ref. [13].

In addition to histopathological samples, other imaging modalities are also considered in the literature for bladder cancer analysis, e.g. magnetic resonance imaging (MRI) [17], cystoscopy [22,23] or computed tomography (CT) [21]. In particular, Dolz et al. [17] applied deep learning algorithms to detect bladder walls and tumour regions from MRI samples. In Refs. [22,23], different deep learning architectures were implemented to distinguish between healthy and bladder cancer patients using cystoscopy samples. Yang et al. [21] outlined a classification between NMIBC and MIBC categories from CT images. Although immunohistochemical techniques are widely used in the literature to detect tumour budding, most state-of-the-art works applied them on colorectal cancer imaging [26–29]. However, there are relatively few immunohistochemistry-based studies for the diagnosis of bladder cancer in the literature. As far as we are aware, only the study conducted in Ref. [30] proposed the use of immunofluorescence-stained samples to quantify tumour budding for the prognosis of MIBC via machine learning algorithms. Specifically, the authors aimed to establish a relationship between tumour budding and assessed survival in patients with MIBC. To do this, they carried out learning methods based on detecting nuclei and segmenting the tumour into stroma regions to count the isolated tumour budding cells. The authors proposed a survival decision function based on random forest, reporting a hazard ratio of 5.44.

1.2. Contribution of this work

To the best of our knowledge, no previous works have been conducted to analyse the severity of bladder cancer using histological images stained with cytokeratin AE1/AE3 immunohistochemistry. Moreover, all the state-of-the-art studies focused on supervised learning methods to find dependencies between the inputs and the predicted class [13,14,19,20]. Some of them [13,14] also considered using unsupervised techniques in the first methodological steps to find possible ROIs with cancer, but required labelled data to build the definitive predictive models. In addition, pattern recognition tasks aimed at grading bladder cancer have not been addressed in previous studies.

To fill these gaps in the literature, we present in this paper a self-learning framework for bladder cancer growth patterns, which focuses on fully unsupervised learning strategies applied on CK AE3/1-stained WSIs. We propose a Deep Convolutional Embedded Attention Clustering (DCEAC) that boosts the performance of the classification model without incurring the cost of labelled data. In the literature, deep-clustering algorithms have demonstrated a high rate of performance for image classification [31–33], image segmentation [34], speech separation [35,36] and data analysis [37], among other tasks. Inspired by Ref. [31], we propose a tailored algorithm capable of competing with the state-of-the-art results achieved by supervised approaches. As a novelty, we include a convolutional attention module to refine the features embedded in the latent space. Additionally, we are the first to focus on the arrangement of histological structures contained in the high-resolution patches to classify them into non-tumour (NT), mild (M) and infiltrative (I) patterns, according to the criteria proposed in Ref. [8]. We also computed a class activation map (CAM) algorithm [?] to evidence that the proposed network focuses on specific structures that match with the clinical patterns associated with bladder cancer aggressiveness.

The proposed end-to-end framework provides a reliable benchmark for making diagnostic suggestions without involving a pathologist, which adds significant value to the body of knowledge. In summary, the main contributions of this work are listed below:

- For the first time, we make use of CK AE3/1-stained images to enable the automatic diagnosis of bladder cancer using machine learning algorithms.
- We base on advanced unsupervised deep learning techniques to address bladder cancer classification without the need for prior annotation steps.
- We propose a new deep-clustering architecture capable of improving the representation space via convolutional attention modules, resulting in better unsupervised classification.
- We focus on high-resolution histological patches to learn specific-bladder cancer patterns and stratify different levels of disease severity according to the literature.

We include heat maps highlighting decisive areas to incorporate an explainable component for network prediction. This provides an interpretability perspective that matches the clinicians' criteria.

2. Material

This study made use of a private database of 136 WSIs (one per patient) from the Hospital Universitario y Politécnico La Fe (Valencia, Spain). The WSIs were stained by immunohistochemistry, and were digitised using an intelligent scanner (LEICA BIOSYSTEMS – Aperio CS2) providing optical magnifications of 20 × (0.5 μm/pixel) and 40 × (0.25 μm/pixel) with a fast network interface of 1 GB/s. Specifically, the 40 × resolution was selected to take advantage of the inherent structure of the bladder patterns associated with each grade of disease, as high image resolution is necessary in order to achieve an accurate diagnosis of bladder cancer. This is because the class dependencies are only evident in the high frequency of the image, especially the details of the tumour budding.

In the first step of the database preparation (Fig. 2), an expert from the Pathological Anatomy Department performed a manual segmentation to indicate possible areas of interest. At this point, it is important to highlight that the segmentation was carried out in a very rough manner, as observed in the green areas of Fig. 2, in order to reduce the expert's

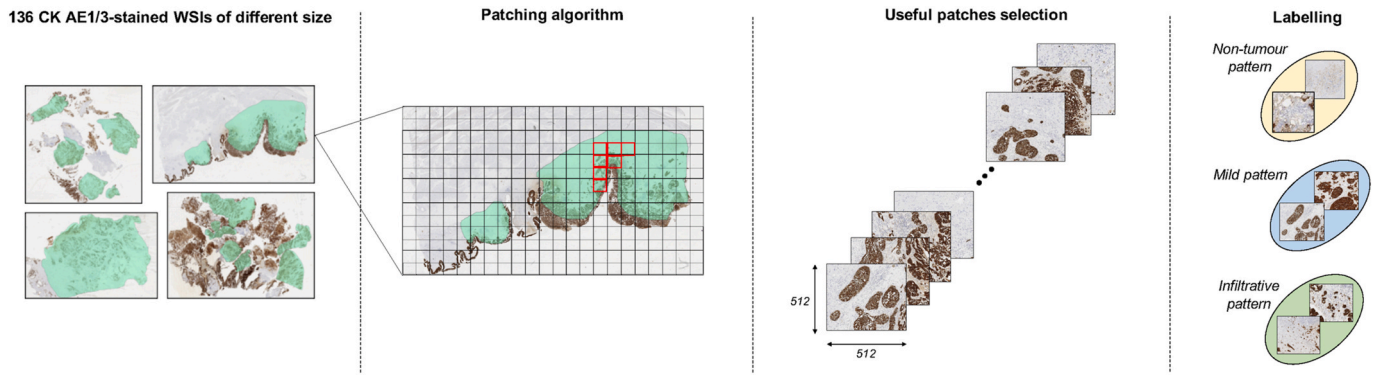

Fig. 2. Database preparation process. On the left-hand side, the rough segmentation (in green) carried out by the pathologist can be seen. Note that each WSI has a different size. In the next section, a patching algorithm was applied on the 136 WSIs stained with CK AE1/3 to extract sub-images (patches) of 512 × 512 pixels. The red rectangles correspond to some examples of useful patches. Finally, a labelling step was conducted for validation purposes. The resulting 2995 sub-images were classified by the expert as non-tumour (NT), mild (M) or infiltrative (I) pattern to give rise to a multi-class scenario for bladder cancer grading.

annotation time as much as possible. The software used to perform the rough annotations was GIGAVISION: a system for labelling tumour regions in gigapixel histological images [38]. A patching algorithm was then applied to extract cropped images with an optimal block size in terms of computational efficiency and structural content. Specifically, patches of dimensions 512 × 512 pixels were extracted, according to some of the most recent studies focusing on histopathological images [18,39,40]. Next, useless regions (WSI background) were discarded by selecting only those patches that contained more than 75% annotated tissue. After this, a total of 2995 representative patches composed the unsupervised framework. For validation purposes, an expert pathologist with more than 20 years of experience manually labelled each patch as non-tumour (NT), mild (M) or infiltrative (I) classes, according to the pattern criteria previously detailed in Section 1. The labelling process resulted in a dataset of 763 non-tumour, 1470 mild and 762 infiltrative cases, as reported in Fig. 2. It is essential to remark that we did not have access to the labelled data during the training phase, as we propose a fully unsupervised strategy to achieve self-learning of the patterns. The labels were only considered at test time to evaluate the models' performance.

Concerning the software and hardware aspects, all models were developed using TensorFlow 2.3.1 on Python 3.6. The experiments were performed on an Intel(R) Core(TM) i7-9700 CPU @3.00 GHz machine with 16 GB of RAM. For deep learning algorithms, a single NVIDIA DGX-A100 Tensor Core with cuDNN 7.6.5 and CUDA Toolkit 10.1 was used.

3. Methods

Recently, deep-clustering algorithms have risen to the forefront of image-based unsupervised techniques, as they are able to enhance feature learning while improving the clustering performance in a unified framework [31]. In this work, we address a fully unsupervised self-learning strategy to cluster a large collection of unlabelled images into $k = 3$ groups corresponding to three different severity levels of MIBC. Inspired by Ref. [31], we propose a novel Deep Convolutional Embedded Attention Clustering (*DCEAC*) in which the feature space is updated in an end-to-end manner to learn stable representations for the clustering stage. Unlike conventional approaches [33], the proposed *DCEAC* algorithm optimises the latent space by preserving the local data structure, which helps stabilise the clustering-learning process without distorting the embedding properties [31] (see Section 3.2).

Self-learning methods aim at learning useful representations by leveraging the domain-specific knowledge from unlabelled data to accomplish downstream tasks. This training procedure is usually tackled by solving pretext tasks [41], relational reasoning [42] or contrastive learning [43] approaches. In our bladder cancer scenario, we advocate for a sequential strategy that uses image reconstruction as an unsupervised prior task. Specifically, we carry out a two-step learning methodology. First, a convolutional autoencoder (CAE) is trained to incorporate information about the properties of the histological domain (Section 3.1). Second (Section 3.2), a clustering branch is included at the output of the CAE bottleneck to provide the class information from the embedded features, which are updated by re-training the CAE on a combined network. In the following sections, we detail both learning steps.

3.1. CAE pre-training

Autoencoder (AE) is one of the most common techniques for data representation and aims to minimise the reconstruction error between X inputs and R outputs. AE architectures consist of two training stages: the encoder $f_\varphi(\cdot)$ and the decoder $g_\theta(\cdot)$, where φ and θ are learnable parameters. The encoder network applies a non-linear mapping function to extract a feature space Z from the input samples X, such that f: $X \rightarrow Z$. The decoder structure is intended to reconstruct the input data from the embedded representations; $R = g_\theta(Z)$. The learning procedure is carried out by minimising a reconstruction loss function.

AE architectures are typically defined by fully connected layers aimed at reducing the dimensionality of the feature space [33,37], or by convolutional layers acting to extract features from 2D or 3D input data [31]. Like [31], we adopted a CAE architecture to address the reconstruction of the histological patches as a pretext task. However, our CAE differs from the current literature in a specific aspect of the network: the bottleneck. Unlike Guo et al. [31], who combined flatten operations with fully connected layers at the central part of the CAE, we introduced a convolutional attention module through a residual connection to improve the latent space for the subsequent clustering task. As seen in Fig. 3, the proposed CAE consists of three main structures: encoder, bottleneck and decoder. The encoder is composed of three stacked convolutional layers with a 3 × 3 receptive field (blue boxes). At the bottleneck, we defined an attention block that allows the embedded features to be refined in the spatial dimension. Specifically, the proposed module combines 1 × 1 convolutions (green boxes) with a sigmoid function (purple layer) intended to re-calibrate the inputs. The inclusion of an identity shortcut forces the network to stabilise the feature space by propagating larger gradients to previous layers via skip connections. An additional 1 × 1 convolutional layer was included at the end of the bottleneck to extract the latent space (z_i) without affecting the dimensions of the feature maps. In the decoding stage, we applied regularisation operations between the transposed convolutional layers (yellow-contour boxes) throughout Batch Normalisation (BN) to avoid the internal covariate shift [44]. Notably, no pooling or up-sampling layers were used to adapt the dimensions of the feature maps after each convolutional step. Instead, we worked with a *stride* > 1 in both the

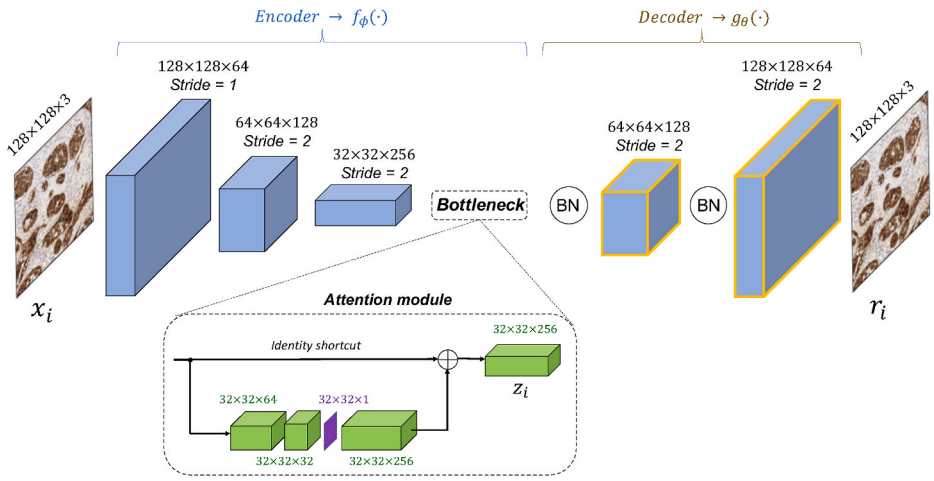

Fig. 3. Architecture of the proposed CAE used for image reconstruction as a pretext task during the learning process.

encoder and decoder structures to provide a more transformable network by learning spatial sub-sampling [31].

As observed in Fig. 3, given an input set of patches $X = \{x_1, x_2, \ldots, x_i, \ldots, x_N\}$, with N samples per batch, the encoder network maps each input $x_i \in \mathbb{R}^{M\times M\times 3}$ into an embedded feature space $z_i = f_\varphi(x_i)$ resulting from the attention module. At the end of the autoencoder network, the decoder function was trained to provide a reconstruction map $r_i = g_\theta(z_i)$ trying to minimise the mean squared error (MSE) between the input x_i and the output r_i, according to Equation (1). Note that the histological patches were resized from $M_0 = 512$ to $M = 128$ to alleviate GPU constraints during model training.

$$L_r = \frac{1}{N}\sum_{i=1}^{N} \|x_i - g_\theta(f_\varphi(x_i))\|^2 \tag{1}$$

3.1.1. Learning details for the CAE pre-training

Given a training set $\mathcal{X} = \{X_1, \ldots, X_b, \ldots, X_B\}$ composed of 2995 histological patches, the proposed CAE was trained during $\varepsilon = 200$ epochs by applying a learning rate of 0.5 on $B = 94$ batches, with $X_b \subset \mathcal{X}$ being a single batch composed of $N = 32$ samples. The Adadelta optimiser [45] was used to update the reconstruction weights by minimising the MSE loss function L_r after each epoch e, as detailed in Algorithm 1. Loading the 2995 histological samples in memory takes 156.59 s at the beginning of the model's training. Then, each epoch takes 6.87 s to train the $B = 94$ batches.

Algorithm 1. CAE training.

Algorithm 1: CAE training.

Data: Unlabelled training dataset
$\mathcal{X} = \{X_1, \ldots, X_b, \ldots, X_B\}$
Results: Trained CAE parameters ϕ and θ.

$\phi, \theta \leftarrow$ random;
for $e \leftarrow 1$ **to** ϵ **do**
  for $b \leftarrow 1$ **to** B **do**
    $X \leftarrow X_b \subset \mathcal{X}$;
    for $i \leftarrow 1$ **to** N **do**
      $r_i \leftarrow g_\theta(f_\phi(x_i))$;
  $\mathcal{L}_r \leftarrow \frac{1}{N}\sum_{i=1}^{N} \|x_i - r_i\|^2$;
  Update ϕ, θ using $\nabla_{\phi,\theta}\mathcal{L}_r$;

3.2. DCEAC training

In the pioneer deep-clustering work [33], the authors proposed a Deep Embedded Clustering (DEC) algorithm in which the decoder structure was discarded during the second stage of clustering training. However, Guo et al. [31] demonstrated that fine-tuning only the encoder network could distort the feature space and hurt the classification performance. Instead, they kept the decoder untouched under the claim that AE architectures can avoid embedding distortion by preserving the local information of the data [46]. We also propose a simultaneous learning process for both reconstruction and clustering branches to avoid feature space corruption, similar to the approach taken in Ref. [31].

Once the CAE was pre-trained in the first stage (Algorithm 1), we incorporated a clustering branch at the output of the CAE bottleneck giving rise to the proposed *DCEAC* model able to provide a soft label of class dependency. From the embedded representations $z_i = \{z_{i,1}, \ldots, z_{i,k}, \ldots, z_{i,C}\}$, with $C = 256$ the number of feature maps $z_{i,k} \in \mathbb{R}^{H\times W}$, we performed a spatial squeeze to obtain a feature vector $z_i' \in \mathbb{R}^C$ leading to a better label assignment. As depicted in Fig. 4, a Global Average Pooling (GAP) layer (faded green) was used to reduce the feature maps $z_{i,k} \in \mathbb{R}^{H\times W}$, with $H = W = 32$, into the feature vector $z_{i,k}' \in \mathbb{R}^{1\times 1}$ (see Equation (2)).

$$z_{i,k}' = \frac{1}{H\times W}\sum_{h=1}^{H}\sum_{w=1}^{W} z_{i,k}(h,w) \tag{2}$$

After the GAP operation, a clustering layer (red box in Fig. 4) was included to map each embedded representation z_i' onto a soft label $q_{i,j}$, which represents the probability of z_i' belonging to cluster j. In accordance with 3, $q_{i,j}$ was calculated via Student's T-distribution [47], keeping the cluster centres $\{\mu_j\}_1^K$ as trainable parameters.

$$q_{i,j} = \frac{\left(1 + \left\|z_i' - \mu_j\right\|^2\right)^{-1}}{\sum_j \left(1 + \left\|z_i' - \mu_j\right\|^2\right)^{-1}} \tag{3}$$

Note that the cluster centres were initialised by running *kmeans* on the embedded features z_i', as detailed in Algorithm 2. From here, a normal target distribution $p_{i,j}$ (defined in Equation (4)) was used as the ground truth during model training.

$$p_{i,j} = \frac{q_{i,j}^2 \Big/ \sum_i q_{i,j}}{\sum_j q_{i,j}^2 \Big/ \sum_i q_{i,j}} \tag{4}$$

The learning framework for the proposed *DCEAC* (Algorithm 2) was carried out by minimising a custom loss function (Equation (5)), where $\mathcal{L}_r$ and $\mathcal{L}_c$ are the reconstruction and clustering losses, respectively. $\gamma > 0$ is a temperature parameter used to prevent the distortion of the feature space, as $\gamma = 0$ would be equivalent to training just the CAE architecture, as detailed in Section 3.1.

$$\mathcal{L} = \mathcal{L}_r + \gamma \mathcal{L}_c \tag{5}$$

Specifically, the clustering loss was defined as Kullback-Leibler divergence ($KL = (P||Q)$) according to Equation (6), whereas the MSE was used as a reconstruction loss function.

$$\mathcal{L}_c = \sum_i \sum_j p_{i_j} log \frac{p_{i,j}}{q_{i,j}} \tag{6}$$

As mentioned above, the autoencoders are responsible for preserving the local structure of the data, so the clustering term must provide only a slight contribution to the updating of the weights in order to avoid latent space corruption. Therefore, we empirically set $\gamma = 0.3$ for all experiments in the training process detailed in Algorithm 2.

Algorithm 2. DCEAC training.

Algorithm 2: DCEAC training.

Data: Unlabelled training dataset
$\mathcal{X} = \{X_1, ..., X_b, ..., X_B\}$
Results: Cluster assignment $\hat{y}_i$ for each histological sample x_i.

Step 1: Cluster centres initialisation
$\phi, \theta \leftarrow$ pre-trained CAE parameters;
$\mathcal{Z} \leftarrow f_\phi(\mathcal{X})$;
$\{\mu_j\}_{j=1}^K \leftarrow kmeans(\mathcal{Z})$;

Step 2: DCEAC training
for $e \leftarrow 1$ **to** ϵ **do**
  for $b \leftarrow 1$ **to** B **do**
    $X \leftarrow X_b \subset \mathcal{X}$;
    for $i \leftarrow 1$ **to** N **do**
      $z_i \leftarrow f_\phi(x_i)$;
      $r_i \leftarrow g_\theta(z_i)$;
      $z_i' \leftarrow GAP(z_i)$;
      $q_{i,j} \leftarrow \frac{(1+\|z_i' - \mu_j\|^2)^{-1}}{\sum_j (1+\|z_i' - \mu_j\|^2)^{-1}}$;
      $p_{i,j} \leftarrow \frac{q_{i,j}^2 / \sum_i q_{i,j}}{\sum_j q_{i,j}^2 / \sum_i q_{i,j}}$;
  $\mathcal{L}_r \leftarrow \frac{1}{N} \sum_{i=1}^N \|x_i - r_i\|^2$;
  $\mathcal{L}_c \leftarrow \sum_i \sum_j p_{i_j} log \frac{p_{i,j}}{q_{i,j}}$;
  $\mathcal{L} \leftarrow \mathcal{L}_r + \gamma \mathcal{L}_c$;
  Update ϕ, θ, μ_j using $\nabla_{\phi,\theta,\mu_j} \mathcal{L}$;

Step 3: Label prediction
for $b \leftarrow 1$ **to** B **do**
  $X \leftarrow X_b \subset \mathcal{X}$;
  for $i \leftarrow 1$ **to** N **do**
    $z_i' \leftarrow GAP(f_\phi(x_i))$;
    $q_{i,j} \leftarrow \frac{(1+\|z_i' - \mu_j\|^2)^{-1}}{\sum_j (1+\|z_i' - \mu_j\|^2)^{-1}}$;
    $\hat{y}_i \leftarrow argmax_j(q_{i,j})$;

3.2.1. Learning details for DCEAC training

As in the previous CAE pre-training, given an input batch X_b of $N = 32$ samples, we made use of the Adadelta optimiser with a learning rate of 0.5 to minimise the custom loss function $\mathcal{L}$ detailed in Algorithm 2. In this case, a single epoch takes 12.42 s to train each batch of $B = 94$ histological samples.

4. Experimental results

4.1. Comparison with other state-of-the-art methods

In this section, we show a comparison between the proposed *DCEAC* model and the most relevant deep clustering-based works in the literature. In particular, we adapted the study carried out in Ref. [33], where the authors proposed a two-step learning strategy based on a Deep Embedded Clustering (DEC) model composed of fully connected layers. In the first step, they trained the autoencoder network to extract domain knowledge from the unlabelled images. In the second step, after encoding the specific image information, Xie et al. [33] discarded the decoder structure to directly address the clustering phase from the learned feature space without considering the reconstruction error. However, later works such as [31] claimed that CAEs are more powerful than fully connected AEs for dealing with images. Thus, we adapted the previous DEC methodology by including convolution operations instead of fully connected layers. To do this, we followed the methodology proposed in Ref. [48], where stacked CAEs were originally proposed for hierarchical feature extraction. To perform a reliable state-of-the-art comparison, we fused both clustering [33] and CAE architectures [48] to provide a refined DEC model (*rDEC*).

We also replicated the experiments conducted by Guo et al. [31], who proposed a hybrid learning for deep-clustering with convolutional autoencoders. The main difference with respect to the previous *rDEC* is that [31] kept the decoder term untouched during model training, resulting in a hybrid framework that combines reconstruction L_r and clustering L_c losses. The idea behind this is that the feature space embedded in *rDEC* could be distorted if only clustering-oriented loss is used. Therefore, they proposed leveraging the decoder structure to avoid latent space corruption by also considering the reconstruction error. Note that one of the main contributions of Guo et al. [31] lies in the proposed bottleneck, as they forced the dimension of the embedded features to be equal to the number of clusters along the fully connected layers. However, this is not scalable to other classification problems with higher-dimensionality input images or with a reduced number of clusters. Specifically, they applied the algorithms on the MNIST dataset composed of samples $x_i \in \mathcal{R}^{28\times28\times1}$ and provided an embedded space z_i with 10 features, depending on the $k = 10$ number of clusters. In our case, we deal with $128 \times 128 \times 3$ pixel images, where the high resolution is essential for the classification performance, unlike in the MNIST dataset. Furthermore, our goal is to classify the histological samples into $k = 3$ classes, so replicating the architecture of [31] is unfeasible as the decoder term would be unable to reconstruct the images from only three feature values. Therefore, to drive a convincing comparison with [31], we kept the same architectures and training details proposed in this work, but removed the convolutional attention module as it is one of our own main contributions. Henceforth, we will refer to this approach as *rDCEC*.

4.2. Quantitative results

In this section, we report the unsupervised classification performance achieved by the aforementioned *rDEC* [33] and *rDCEC* [31] algorithms in comparison to our proposed *DCEAC* model. Conventional methods based on running the clustering algorithms (*kmeans*, *spectral* and *agglomerative*) on the feature space were also considered to find out the performance difference between the proposed model and traditional techniques. These conventional approaches will be referred to as $AE + kmeans$, $AE + spectral$ and $AE + agg$, respectively. In Table 1, we present the class performance obtained from the conventional clustering methods to show how well the three algorithms classify the 2995 histological patches with non-tumour (NT), mild (M) and infiltrative (I)

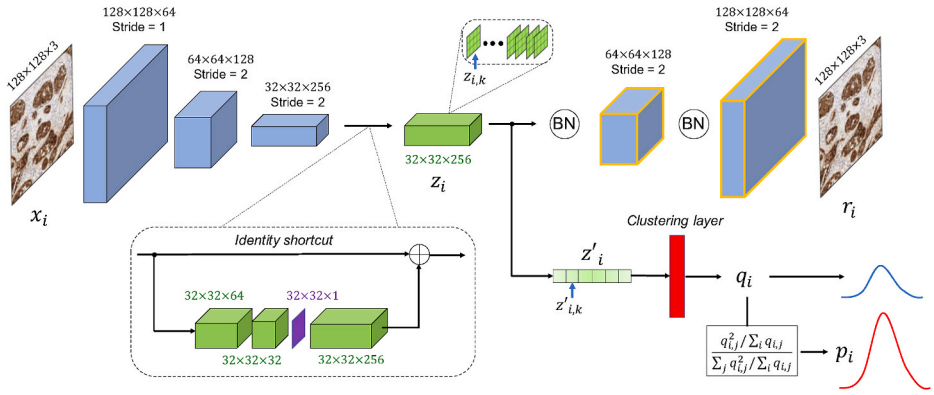

Fig. 4. Architecture of the proposed Deep Convolutional Embedded Attention Clustering (DCEAC). The model is trained in an end-to-end manner by minimising both reconstruction and clustering loss functions. The reconstruction pretext task stabilises the feature space z_i avoiding the embedding distortion, while the clustering term predicts the soft-class assignments q_i.

patterns. Similarly, we also evaluate the per-class behaviour of the deep clustering-based algorithms (*rDEC*, *rDCEC* and *DCEAC*) in Table 2. In addition, the micro- and macro-average classification results are reported in Table 3. Both metrics provide information about the overall average performance of the classification models, but the micro-average takes into account the imbalance between classes, which allows for a truer picture of the models' behaviour than does the macro-average. Comparison among the different methods is handled by means of different figures of merit, such as sensitivity (SN), specificity (SP), F-score (FS), accuracy (ACC) and area under the ROC curve (AUC).

To improve the comparison between the six learning approaches, we represent in Fig. 5 the latent space laid out by each clustering model with its respective confusion matrix. While the confusion matrix provides information about the classification ability of each model, the representation of the embedded features contributes to a more comprehensive clustering scenario for bladder cancer grading. Thus, while the latent space representation would fit better in the qualitative section, the confusion matrix provides a quantitative perspective that aids the interpretation of the embedded feature map, as discussed in Section 5.

4.3. Qualitative results

In an attempt to incorporate an interpretative perspective for the reported quantitative results, we computed the class activation maps (CAMs), which highlight the regions to which the model pays attention in order to predict the class of each sample. This often helps to find hidden patterns associated with a specific class or to determine whether the label prediction is based on the same patterns as the clinicians' findings. In this way, the reported heat maps lead to a better understanding of the embedded feature space by pinpointing areas of the histological patches that are decisive in cluster assignment.

Table 1
Unsupervised classification results per class achieved from conventional clustering methods.

	NON-TUMOUR			MILD			INFILTRATIVE		
	AE + kmeans	*AE + spectral*	*AE + agg*	*AE + kmeans*	*AE + spectral*	*AE + agg*	*AE + kmeans*	*AE + spectral*	*AE + agg*
SN	0.9345	0.9227	**0.9869**	0.5020	0.6327	**0.7415**	0.5105	**0.5827**	0.2533
SP	0.9870	0.9718	**0.9960**	0.7659	**0.8210**	0.6249	0.6556	0.7398	**0.8307**
FS	0.9475	0.9203	**0.9875**	0.5754	0.6958	**0.6960**	0.4052	**0.4969**	0.2896
ACC	0.9736	0.9593	**0.9937**	0.6364	**0.7285**	0.6821	0.6187	**0.6938**	0.6838

Table 2
Unsupervised classification results per class achieved from deep-clustering methods.

	NON-TUMOUR			MILD			INFILTRATIVE		
	rDEC	*rDCEC*	*DCEAC*	*rDEC*	*rDCEC*	*DCEAC*	*rDEC*	*rDCEC*	*DCEAC*
SN	**1**	**1**	0.9987	0.8082	0.8952	**0.9041**	0.4659	0.5105	**0.6168**
SP	0.9319	0.9780	**0.9978**	**0.8262**	0.7862	0.8118	0.8782	0.9319	**0.9364**
FS	0.9094	0.9689	**0.9961**	0.8129	0.8458	**0.8613**	0.5112	0.5971	**0.6841**
ACC	0.9492	0.9836	**0.9980**	0.8174	0.8397	**0.8571**	0.7733	0.8247	**0.8551**

Table 3
Unsupervised classification results in terms of micro- and macro-average achieved from both conventional and deep clustering methods.

	MICRO-AVERAGE						MACRO-AVERAGE					
	AE + kmeans	*AE + spectral*	*AE + agg*	*rDEC*	*rDCEC*	*DCEAC*	*AE + kmeans*	*AE + spectral*	*AE + agg*	*rDEC*	*rDCEC*	*DCEAC*
SN	0.6144	0.6938	0.6798	0.7699	0.8240	**0.8551**	0.6490	0.7127	0.6606	0.7580	0.8019	**0.8399**
SP	0.8072	0.8469	0.8399	0.8850	0.9120	**0.9275**	0.8028	0.8442	0.8172	0.8788	0.8987	**0.9153**
FS	0.6144	0.6938	0.6798	0.7699	0.8240	**0.8551**	0.6427	0.7043	0.6577	0.7445	0.8039	**0.8472**
ACC	0.7429	0.7959	0.7865	0.8466	0.8827	**0.9034**	0.7429	0.7959	0.7865	0.8466	0.8827	**0.9034**
AUC	0.7259	0.7784	0.7389	0.8184	0.8503	**0.8776**	0.7259	0.7784	0.7389	0.8184	0.8503	**0.8776**

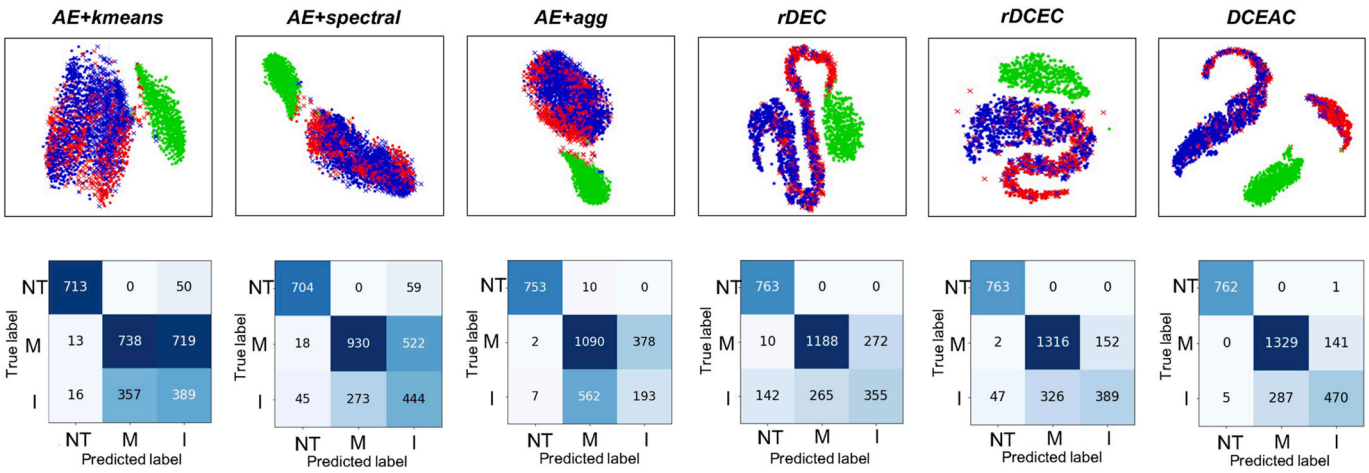

Fig. 5. The top row with the scatter graphics corresponds to the representation of the latent space from the clustering classification achieved by each method. The T-distributed Stochastic Neighbour Embedding (TSNE) tool was employed to illustrate the feature space in a 2D map. Well- and missclassified embedded features are represented by spots and crosses, respectively. The green, blue and red colours refer to the non-tumour (NT), mild (M) and infiltrative (I) patterns. In the bottom row, a confusion matrix per method is shown to elucidate the performance of each one in discerning the three different levels of MIBC disease severity.

As can be deduced from Fig. 5, the major challenge in the classification of MIBC lies in distinguishing mild (M) and infiltrative (I) cancerous patterns, as expected. For this reason, in Fig. 6 we present several examples of heat maps corresponding to missclassified samples to elucidate why the proposed model is flawed. We also show examples of well-predicted CAMs to evidence the relevant structures to which the

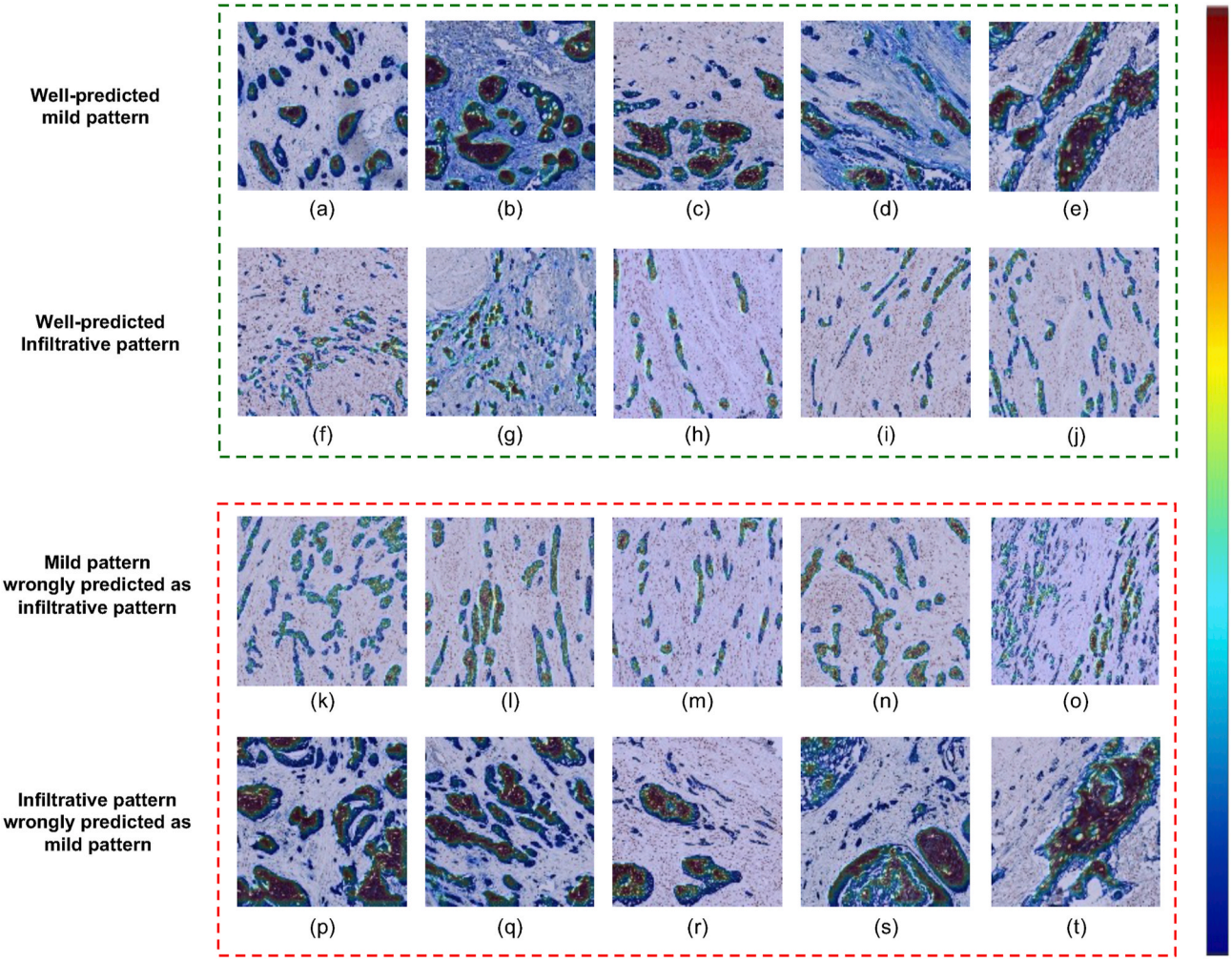

Fig. 6. Class activation maps highlighting the regions that the proposed DCEAC model considers relevant for the class prediction. The green frame refers to well-predicted images with mild (M) and infiltrative (I) patterns, whereas the red frame corresponds to missclassified samples where disease aggressiveness has been confused. The more important the areas, the warmer the colour in which they are represented in the heat maps, so that the blue tones denote less important regions, and red tones refer to more important ones.

network pays attention when predicting correctly. Specifically, we show five examples per case to make clear the criteria followed by the proposed model to determine the class. In the green frame of Fig. 6, we illustrate well-classified mild (a-e) and infiltrative (f-j) histological patterns. Additionally, in the red frame, we show bladder cancer samples with a mild pattern missclassified as tumour budding (k-o), and vice versa (p-t). The findings from the class activation maps will be discussed in Section 5.

5. Discussion

5.1. About quantitative results

From Tables 1 and 2, we can observe that all the contrasted models work well for detecting the non-tumour class. In the conventional approaches (Table 1), the *AE + agg* algorithm shows slightly better performance, but at the cost of greatly compromising detection of the rest of the classes, as discussed below. In contrast, the proposed *DCEAC* model (Table 2) achieves the highest performance for all metrics except for sensitivity, as the model missclassifies a non-tumour sample as an infiltrative case, as reported in the confusion matrix of *DCEAC* in Fig. 5. With respect to the mild (M) and infiltrative (I) patterns, it is appreciated that the deep-clustering models notably improve the success of unsupervised classification compared to conventional approaches. Regarding the mild class, *AE + spectral* and *AE + agg* clustering methods show similar behaviour, but the proposed model provides the highest performance with an increase in accuracy of 12.85% with respect to the best conventional approach. Only *rDEC* surpasses it in any metric (by 1% in specificity), but in exchange for a 10% drop in sensitivity relative to the proposed *DCEAC*. During the evaluation of the infiltrative class, the *AE + agg* method drops strongly, which places *AE + kmeans* and *AE + spectral* as much more reliable conventional clustering algorithms. Comparing all algorithms side-by-side, the proposed *DCEAC* method shows the best results for all metrics, especially the F-score, in which *DCEAC* outperforms the other approaches by more than 10%.

Table 3 reports the overall performance of the models, in terms of micro- and macro-average. As mentioned above, the micro-average results take into account the imbalance between classes, which is an important aspect in this study, as the samples with mild pattern are oversampled. Nevertheless, the proposed *DCEAC* model consistently outperforms the other clustering methods by 2–3% across both micro- and macro-averaging, as can be seen in Table 3. As a final remark on the quantitative results, it is worth noting that the expert's decision coincides with the proposed artificial intelligence system in 90.34% of the cases, according to the average accuracy.

A reinforcement of the quantitative results is reported in Fig. 5. In the confusion matrices, it is clear from the range of colours that all models tend to confuse mild cancerous and infiltrative patterns. Conventional algorithms demonstrate a very low ability to discern between cancerous samples as most of the images are predicted as a mild pattern due to oversampling of that class. This changes when deep-clustering algorithms are profiled. Specifically, the *rDEC* model improves the classification of carcinogenic images but greatly compromises that of the non-tumour class by missclassifying samples with infiltrative pattern. In contrast, the *rDCEC* model improves the results by significantly decreasing the instances of tumour budding samples erroneously predicted as non-tumour cases. In addition, *rDCEC* increases the number of true positives for tumour samples. However, this model presents a major shortcoming in predicting infiltrative patterns, as a large number of them are wrongly labelled as mild. Unequivocally, the proposed *DCEAC* model provides the best classification results. The number of samples with tumour budding missclassified as non-tumour cases decreases to a minimum, in contrast to the aforementioned methods. Moreover, the number of true positives increases for both mild and infiltrative patterns compared to the results of the other methods, while false positives and false negatives are reduced.

The representation of the embedded feature space offers a visual perspective of the quantitative results. We can observe that the conventional approaches are able to roughly discern between non-tumour and carcinogenic histological samples. However, the point clouds are too fuzzy to separate mild from infiltrative classes. Contrarily, the *rDEC* model shows a better distribution of the embedded data, although the features relative to each class are still close together in the latent space. This improves in the case of the *rDCEC* model, where independent clusters start to become apparent. The non-tumour features (shown in green) become unmarked in the representation space and the embedded tumour samples start to disperse into different classes of cluster. Indisputably, the *DCEAC* algorithm provides the best embedding representation as the features are distributed throughout the latent space, forming independent clusters according to a specific class. This further strengthens our confidence in the ability of the proposed model to discern between non-tumour, mild and infiltrative histological patterns.

From the above in-depth analysis of the quantitative and interpretative results, several conclusions can be drawn. The first is that the use of deep learning techniques improves classification performance compared to conventional clustering approaches. As expected, all deep clustering-based methods, i.e. *rDEC*, *DCEC* and *DCEAC*, outperform the baseline based on the traditional *kmeans*, *spectral* and *agglomerative* algorithms. This is because the deep-clustering models allow for a more extensive learning stage in which the embedded features conform to a target distribution, unlike conventional algorithms, which modify the clusters iteratively without updating the feature learning. Additionally, we can observe that models with both the reconstruction and clustering branches integrated into a unified framework provide better results than the *rDEC* model, which carries out the learning process in two independent stages. The reason behind *rDCEC* and *DCEAC* outperforming *rDEC* lies in the preservation of the local structure of the embedded data. Since *rDCEC* and *DCEAC* models have a connected output between the clustering and reconstruction stages, the clustering term can transfer class information to the reconstruction term, which is responsible for updating the weights of the encoder network. In this way, the embedded features can be optimised by incorporating the class prediction without distorting the latent space, thanks to the decoder structure. Finally, the proposed *DCEAC* model shows substantial performance improvements over the rest of the approaches. This is due to the inclusion of the convolutional attention block, which allows for the refinement of the latent space, in order to provide more suitable features for the clustering phase.

5.2. About qualitative results

As observed in the CAMs shown in Fig. 6, the proposed *DCEAC* model focuses on tumour cell nests (Fig. 6, a-c) and interconnected tumour bands (Fig. 6, d-e) when predicting samples with a mild pattern. This implies that the proposed network has learnt by itself to associate nodular and trabecular structures with a mild pattern of disease. Furthermore, the *DCEAC* model recognises small clusters of isolated buds (Fig. 6, f-g) or tumour cell cords (Fig. 6, h-j) as structures characteristic of the infiltrative pattern. These findings are evidenced in the green frame of the heat maps corresponding to well-predicted samples.

In the case of the wrong predictions (red frame in Fig. 6), we can observe that the proposed network maintains consistency in determining the class of each sample. The histological patches in Fig. 6(k–o), in which the network highlights small filaments reminiscent of tumour budding structures, are similar in appearance to infiltrative patterns. However, the true label assigned by the expert for these samples was a mild pattern, as trabecular structures are often difficult to distinguish from the cell cords of the infiltrative pattern. The qualitative results thus present an opportunity here for the model's suggestions to serve a role similar to that of a second opinion, and lead pathologists to reconsider their diagnoses. In addition, the human eye is susceptible to fatigue, so the proposed system could help in cases where some patterns have gone unnoticed, in order to avoid a biased diagnosis.

In contrast, in the cases of Fig. 6(p–t), samples with an infiltrative pattern are erroneously predicted by the model as mild cases. In these histological patches, the proposed network focuses on larger structures related to nodular or trabecular patterns, but ignores small, isolated tumour cells that lead to increased severity of bladder cancer. It follows that, although the final prediction could be wrong, the pattern recognition accomplished by the model maintains consistency. Note that the model is wrong because patterns that belong to a different class coexist in the same histological patch, so we will face this problem in future research lines.

In summary, the proposed *DCEAC* model demonstrates, through heat maps, a high-confidence prediction as it is able to focus on the same patterns as the clinicians, without having prior information from them. As mentioned above, the expert's opinion and the proposed model coincide in most cases (specifically, 90.34% of cases). Thus, the artificial intelligence system could help as a computer-aided system for process review, which would lead to an improvement in the quality of diagnosis without the need to involve other experts. In addition, the proposed system could be used as a competent tool to help inexperienced pathologists by suggesting annotations for specific areas of interest.

6. Conclusion

In this paper, we have proposed a novel self-learning framework based on deep-clustering techniques to grade the severity of bladder cancer through histological samples. Immunohistochemistry staining methods have been considered to enhance the non-tumour, mild and infiltrative patterns, according to the literature. The proposed DCEAC outperforms other conventional and deep clustering-based methods, achieving an average accuracy of 0.9034 for grading the aggressiveness of MIBC. Furthermore, the reported CAMs show that the proposed system is able to self-learn the same structures as clinicians to associate patterns with the correct disease severity grade, without incurring prior annotation steps. In this line, our fully unsupervised approach bridges the gap with respect to other supervised algorithms, as the proposed system does not require the involvement of experts for model training.

In future research lines, we will work on improving the accuracy of tumour sample classification when structures of different growth patterns appear in the same image. We will propose the use of convolutional variational autoencoders considering probabilistic and deterministic attention modules. We will use more powerful hardware systems to process entire high-resolution WSIs to provide a diagnosis per biopsy, instead of per patch. Finally, we will pursue an end-to-end system in which no prior raw annotations are necessary.

Funding

This work has been partially funded by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska Curie grant agreement No 860627 (CLARIFY Project), the State Research Spanish Agency under the AI4SKIN project (PID2019-105142RB-C21) and GVA through project PROMETEO/2019/109. The work of Gabriel García has been supported by the State Research Spanish Agency PTA2017-14610-I. The equipment used for this research has been funded by the European Union within the operating Program ERDF of the Valencian Community 2014–2020 with the grant number IDIFEDER/2020/030.

References

[1] S. Antoni, J. Ferlay, I. Soerjomataram, A. Znaor, A. Jemal, F. Bray, Bladder cancer incidence and mortality: a global overview and recent trends, Eur. Urol. 71 (1) (2017) 96–108.

[2] L. Lorenzo, Valor pronóstico de la presencia de un componente tumoral indiferenciado ("tumor budding") en pacientes con carcinoma vesical músculo-invasivo, Ph.D. thesis, Universitat de València, 2018.

[3] G. Feil, A. Stenzl, Pruebas de marcadores tumorales en el cáncer de vejiga, Actas Urol. Esp. 30 (1) (2006) 38–45.

[4] S. Sharma, P. Ksheersagar, P. Sharma, Diagnosis and treatment of bladder cancer, Am. Fam. Physician 80 (7) (2009) 717–723.

[5] K.A. Richards, N.D. Smith, G.D. Steinberg, The importance of transurethral resection of bladder tumor in the management of nonmuscle invasive bladder cancer: a systematic review of novel technologies, J. Urol. 191 (6) (2014) 1655–1664.

[6] A. Stenzl, N. Cowan, M. De Santis, M. Kuczyk, A. Merseburger, M. Ribal, A. Sherif, J. Witjes, Guía clínica sobre el cáncer de vejiga con invasión muscular y metastásico, European Association of Urology 1 (2010) 1–72.

[7] C. Busch, F. Algaba, The WHO/ISUP 1998 and WHO 1999 systems for malignancy grading of bladder cancer. Scientific foundation and translation to one another and previous systems, Virchows Arch. 441 (2) (2002) 105–108.

[8] R.E. Jimenez, E. Gheiler, P. Oskanian, R. Tiguert, W. Sakr, D.P. Wood Jr., J. E. Pontes, D.J. Grignon, Grading the invasive component of urothelial carcinoma of the bladder and its relationship with progression-free survival, Am. J. Surg. Pathol. 24 (7) (2000) 980–987.

[9] A. Almangush, M. Karhunen, S. Hautaniemi, T. Salo, I. Leivo, Prognostic value of tumour budding in oesophageal cancer: a meta-analysis, Histopathology 68 (2) (2016) 173–182.

[10] E. Karamitopoulou, I. Zlobec, D. Born, A. Kondi-Pafiti, P. Lykoudis, A. Mellou, K. Gennatas, B. Gloor, A. Lugli, Tumour budding is a strong and independent prognostic factor in pancreatic cancer, Eur. J. Cancer 49 (5) (2013) 1032–1039.

[11] R. Masuda, H. Kijima, N. Imamura, N. Aruga, Y. Nakamura, D. Masuda, H. Takeichi, N. Kato, T. Nakagawa, M. Tanaka, et al., Tumor budding is a significant indicator of a poor prognosis in lung squamous cell carcinoma patients, Mol. Med. Rep. 6 (5) (2012) 937–943.

[12] K. Fukumoto, E. Kikuchi, S. Mikami, K. Ogihara, K. Matsumoto, A. Miyajima, M. Oya, Tumor budding, a novel prognostic indicator for predicting stage progression in t1 bladder cancers, Cancer Sci. 107 (9) (2016) 1338–1344.

[13] R. Wetteland, K. Engan, T. Eftestøl, V. Kvikstad, E.A. Janssen, Multiclass tissue classification of whole-slide histological images using convolutional neural networks, ICPRAM 1 (2019) 320–327.

[14] Z. Zhang, P. Chen, M. McGough, F. Xing, C. Wang, M. Bui, Y. Xie, M. Sapkota, L. Cui, J. Dhillon, et al., Pathologist-level interpretable whole-slide cancer diagnosis with deep learning, Nature Machine Intelligence 1 (5) (2019) 236–245.

[15] M. Lucas, I. Jansen, T.G. van Leeuwen, J.R. Oddens, D.M. de Bruin, H. A. Marquering, Deep learning-based recurrence prediction in patients with non-muscle-invasive bladder cancer, Eur. Urol.Focus 1 (2020) 1–8.

[16] P.-N. Yin, K. Kishan, S. Wei, Q. Yu, R. Li, A.R. Haake, H. Miyamoto, F. Cui, Histopathological distinction of non-invasive and invasive bladder cancers using machine learning approaches, BMC Med. Inf. Decis. Making 20 (1) (2020) 1–11.

[17] J. Dolz, X. Xu, J. Rony, J. Yuan, Y. Liu, E. Granger, C. Desrosiers, X. Zhang, I. Ben Ayed, H. Lu, Multiregion segmentation of bladder cancer structures in MRI with progressive dilated convolutional networks, Med. Phys. 45 (12) (2018) 5482–5493.

[18] A.-C. Woerl, M. Eckstein, J. Geiger, D.C. Wagner, T. Daher, P. Stenzel, A. Fernandez, A. Hartmann, M. Wand, W. Roth, et al., Deep learning predicts molecular subtype of muscle-invasive bladder cancer from conventional histopathological slides, Eur. Urol. 78 (2) (2020) 256–264.

[19] H. Xu, S. Park, S.H. Lee, T.H. Hwang, Using Transfer Learning on Whole Slide Images to Predict Tumor Mutational Burden in Bladder Cancer Patients, bioRxiv, 2019, 554527.

[20] S.A. Harmon, T.H. Sanford, G.T. Brown, C. Yang, S. Mehralivand, J.M. Jacob, V. A. Valera, J.H. Shih, P.K. Agarwal, P.L. Choyke, et al., Multiresolution application of artificial intelligence in digital pathology for prediction of positive lymph nodes from primary tumors in bladder cancer, JCO.Clin.Cancer .Inf. 4 (2020) 367–382.

[21] Y. Yang, X. Zou, Y. Wang, X. Ma, Application of deep learning as a noninvasive tool to differentiate muscle-invasive bladder cancer and non-muscle-invasive bladder cancer with CT, Eur. J. Radiol. (2021), 109666.

[22] A. Ikeda, H. Nosato, Y. Kochi, T. Kojima, K. Kawai, H. Sakanashi, M. Murakawa, H. Nishiyama, Support system of cystoscopic diagnosis for bladder cancer based on artificial intelligence, J. Endourol. 34 (3) (2020) 352–358.

[23] R. Yang, Y. Du, X. Weng, Z. Chen, S. Wang, X. Liu, Automatic recognition of bladder tumours using deep learning technology and its clinical application, Int. J. Med. Robot. Comput. Assist. Surg. (2020), e2194.

[24] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Adv. Neural Inf. Process. Syst. 25 (2012) 1097–1105.

[25] K. Cho, B. Van Merriënboer, D. Bahdanau, Y. Bengio, On the Properties of Neural Machine Translation: Encoder-Decoder Approaches, arXiv preprint arXiv: 1409.1259, 2014.

[26] F. Prall, H. Nizze, M. Barten, Tumour budding as prognostic factor in stage i/ii colorectal carcinoma, Histopathology 47 (1) (2005) 17–24.

[27] A. Lugli, E. Karamitopoulou, I. Panayiotides, P. Karakitsos, G. Rallis, G. Peros, G. Iezzi, G. Spagnoli, M. Bihl, L. Terracciano, et al., Cd8+ lymphocytes/tumour-budding index: an independent prognostic factor representing a 'pro-/anti-tumour' approach to tumour host interaction in colorectal cancer, Br. J. Cancer 101 (8) (2009) 1382–1392.

[28] T. Ogawa, T. Yoshida, T. Tsuruta, W. Tokuyama, S. Adachi, M. Kikuchi, T. Mikami, K. Saigenji, I. Okayasu, Tumor budding is predictive of lymphatic involvement and lymph node metastases in submucosal invasive colorectal adenocarcinomas and in non-polypoid compared with polypoid growths, Scand. J. Gastroenterol. 44 (5) (2009) 605–614.

[29] I. Zlobec, M.P. Bihl, A. Foerster, A. Rufle, A. Lugli, The impact of CpG island methylator phenotype and microsatellite instability on tumour budding in colorectal cancer, Histopathology 61 (5) (2012) 777–787.
[30] N. Brieu, C.G. Gavriel, I.P. Nearchou, D.J. Harrison, G. Schmidt, P.D. Caie, Automated tumour budding quantification by machine learning augments TNM staging in muscle-invasive bladder cancer prognosis, Sci. Rep. 9 (1) (2019) 1–11.
[31] X. Guo, X. Liu, E. Zhu, J. Yin, Deep clustering with convolutional autoencoders, in: International Conference on Neural Information Processing, Springer, 2017, pp. 373–382.
[32] X. Guo, E. Zhu, X. Liu, J. Yin, Deep embedded clustering with data augmentation, in: Asian Conference on Machine Learning, PMLR, 2018, pp. 550–565.
[33] J. Xie, R. Girshick, A. Farhadi, Unsupervised deep embedding for clustering analysis, in: International Conference on Machine Learning, PMLR, 2016, pp. 478–487.
[34] J. Enguehard, P. O'Halloran, A. Gholipour, Semi-supervised learning with deep embedded clustering for image classification and segmentation, IEEE Access 7 (2019) 11093–11104.
[35] J.R. Hershey, Z. Chen, J. Le Roux, S. Watanabe, Deep clustering: discriminative embeddings for segmentation and separation, in: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2016, pp. 31–35.
[36] B.H. Prasetio, H. Tamura, K. Tanno, A deep time-delay embedded algorithm for unsupervised stress speech clustering, in: 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), IEEE, 2019, pp. 1193–1198.
[37] R. del Amor, A. Colomer, C. Monteagudo, V. Naranjo, A Deep Embedded Refined Clustering Approach for Breast Cancer Distinction Based on DNA Methylation, arXiv preprint arXiv:2102, 2021, 09563.
[38] A. Colomer, V. Naranjo, F. Fuentes, Gigavision, Sistema para el marcado de regiones tumorales en imágnes histológicas gigapixel, 2016.
[39] R. del Amor, L. Launet, A. Colomer, A. Moscardó, A. Mosquera-Zamudio, C. Monteagudo, V. Naranjo, An Attention-Based Weakly Supervised Framework for Spitzoid Melanocytic Lesion Diagnosis in WSI, arXiv preprint arXiv:2104.09878, 2021.
[40] J. Silva-Rodriguez, A. Colomer, J. Dolz, V. Naranjo, Self-learning for weakly supervised gleason grading of local patterns, IEEE. J. Biomed.Health Inf. 25 (2021) 3094–3104.
[41] S. Gidaris, P. Singh, N. Komodakis, Unsupervised Representation Learning by Predicting Image Rotations, arXiv preprint arXiv:1803.07728, 2018.
[42] M. Patacchiola, A. Storkey, Self-supervised Relational Reasoning for Representation Learning, arXiv preprint arXiv:2006, 2020, 05849.
[43] T. Chen, S. Kornblith, M. Norouzi, G. Hinton, A simple framework for contrastive learning of visual representations, in: International Conference on Machine Learning, PMLR, 2020, pp. 1597–1607.
[44] S. Ioffe, C. Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift, in: International Conference on Machine Learning, PMLR, 2015, pp. 448–456.
[45] M.D. Zeiler, Adadelta: an Adaptive Learning Rate Method, arXiv preprint arXiv: 1212.5701, 2012.
[46] X. Peng, S. Xiao, J. Feng, W.-Y. Yau, Z. Yi, Deep subspace clustering with sparsity prior, in: IJCAI, 2016, pp. 1925–1931.
[47] L. Van der Maaten, G. Hinton, Visualizing data using t-sne, J. Mach. Learn. Res. 9 (11) (2008).
[48] 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.