CNN Ensembles for Nuclei Segmentation on Histological Images of OED

CNN Ensembles for Nuclei Segmentation on Histological Images of OED

Adriano B. Silva*, Guilherme B. Rozendo†, Thaína A. A. Tosta‡, Alessandro S. Martins§, Adriano M. Loyola¶, Sérgio V. Cardoso¶, Alessandra Lumini ‖, Leandro A. Neves†, Paulo R. de Faria** and Marcelo Z. do Nascimento*
*Faculty of Computer Science - Federal University of Uberlândia
†Department of Computer Science and Statistics (DCCE), São Paulo State University
‡Science and Technology Institute, Federal University of São Paulo
§Federal Institute of Triângulo Mineiro
¶Area of Oral Pathology, School of Dentistry, Federal University of Uberlândia
‖Department of Computer Science and Engineering, University of Bologna
**Department of Histology and Morphology, Institute of Biomedical Science, Federal University of Uberlândia
E-mail: *adriano.barbosa@ufu.br*

***Abstract*—Early diagnosis of potentially malignant disorders, such as oral epithelial dysplasia (OED), is the most reliable way to prevent oral cancer. Computational algorithms have been used as a tool to aid specialists in this process. In recent years, CNN-based methods are getting more attention due to its improved results in nuclei segmentation tasks. Despite these relevant results, achieving high segmentation accuracy remains a challenging task. In this paper, we propose an ensemble of segmentation models to improve the performance of nuclei segmentation in OED histopathology images. The proposed ensemble consists of seven CNN segmentation models, which were combined using three ensemble strategies: simple averaging, weighted averaging, and majority voting, achieved accuracy of 92.14%, 91.21% and 90.67%, respectively, when applied in OED images. The model's performance was also evaluated on three publicly available datasets and achieved comparable performance to the state-of-the-art segmentation methods. These values indicate that the proposed ensemble methods can improve segmentation results and be used in medical image analysis applications.**

***Index Terms*—Oral Epihtelial Dysplasia, Nuclei Segmentation, Histological Image Processing, CNN Ensemble**

I. Introduction

Oral epithelial dysplasia (OED) refers to a disordered growth in the epithelium of the oral cavity that leads to changes in cellular properties such as size and shape. It is a common type of precancerous lesion and can be categorized as mild, moderate, or severe [1]. Several studies in the literature have reported OED to have a transformation rate to oral squamous cell carcinoma ranging from 6-36%. The diagnosis for these lesions is often performed through microscopic analysis of the lesion size and the intensity of morphological alterations in tissue nuclei. However, the irregular aspect of cell nuclei poses a challenge to this analysis [2].

The development of digital applications for histological images has enabled specialists to obtain relevant data for a thorough investigation, assisting them in decision-making [3]. These computer-aided diagnosis (CAD) systems provide quantitative analysis of a large number of data and features [4]. The CAD systems involve a series of steps ranging from improving signal-to-noise ratio, segmentation, feature extraction, and classification. Segmentation is a crucial step, as it allows the identification of lesion structures that will be analyzed in subsequent steps using feature descriptors and classifiers [4]. In the case of OED images, this can be a challenging process due to the growth of the connective tissue, which can invade the epithelial tissue and make nuclei segmentation complex [5].

Recently, deep learning methods have shown significant progress in nuclei segmentation. Convolutional neural network models (CNNs) are among the most popular deep learning approaches due to their ability to achieve good results in image analysis [6]. In Dos Santos et al.'s study [7], a U-Net based approach was employed along with data augmentation to segment cell nuclei in a set of 120 OED histological images. This methodology achieved results with accuracy, precision, Dice coefficient, and Jaccard index values of 0.879, 0.793, 0.820, and 0.699, respectively. Shephard et al. [8] employed the HoVer-Net CNN model along with post processing operations to segment cell nuclei in 43 whole-slide OED images, including 14 mild, 13 moderate, 11 severe, and 5 healthy tissues. The method yielded high segmentation accuracy, with a Dice coefficient and Jaccard index of 0.839 and 0.694, respectively, for cell nuclei in OED lesions. Silva et al. [9] presented an alternative method for OED nuclei segmentation, employing the Mask R-CNN in combination with the ResNet50 backbone to perform the segmentation task. The methodology was applied to an image set of 296 OED images, with 74 images for each class, achieving a nuclei segmentation accuracy of 89.31%.

Although deep learning methods have shown good results in various applications, overfitting and high variations in the results generated by CNNs remain recurrent problems [2], [10]. To mitigate these potential errors, ensembles have been explored as a solution. In this study, we propose a ensemble of CNN models for the segmentation of cell nuclei in histological images of OED. Our study has the following main contributions:

- Evaluation of the performance of CNN models for the segmentation of cell nuclei;
- Study of ensemble segmentation strategies to improve the results achieved by the individual models;
- Analysis of the results obtained by the CNN models and ensemble systems using evaluation metrics;
- Investigation of the proposed method and analysis of other studies proposed in literature on public image datasets.

II. Methodology

The proposed methodology involved a segmentation process that employed CNNs, post processing and ensemble strategies, as presented in Fig. 1. The system was developed using the Python and MATLAB® programming languages. The experiments were performed on a computer equipped with a Ryzen 5 CPU, 64 GB of RAM, and an Nvidia RTX 2070 GPU with 12 GB of VRAM.

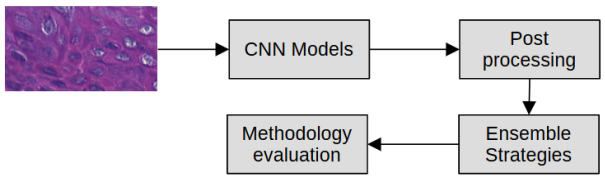


Fig. 1: Flowchart of the main stages employed to segment nuclear structures on OED tissues.

A. Image Dataset

The image dataset was built from 30 H&E-stained mice tongue tissue sections that were previously exposed to a carcinogen during two experiments conducted between 2009 and 2010. These experiments were approved by the Ethics Committee on the Use of Animals under protocol numbers 038/09 and A016/21 at the Federal University of Uberlândia, Brazil.

The histological slides were digitized using the Leica DM500 optical microscope with a magnification of 400×. A total of 456 region of interest (ROI) images were stored in the TIFF format using the RGB colour model, 8-bit channel depth, and dimensions of 450 × 250 pixels. This resulted in 114 raw images for each class. Following the methodology described by Lumerman et al. [11], the images were classified into healthy mucosa, mild, moderate, and severe OED by a specialist. The cell nuclei were manually marked by the specialist, and the resulting labels were evaluated and validated by a pathologist, defining the gold-standard annotations used to evaluate the methodology.

In order to evaluate the proposed methodology in relation to the computational model, this study also employed the following datasets: CryoNuSeg [12], MoNuSeg 2018 [13], and GlaS [14]. CryoNuSeg includes 30 images of size 512 × 512 pixels of 10 human organs. The MoNuSeg 2018 dataset contains 51 tumors images from different organs and with size of 1000 × 1000 pixels. GlaS is composed of 165 images of size 775 × 522 pixels from colorrectal cancer tissues. These databases were chosen to evaluate the proposed method on public domain databases as well as to analyze its performance in relation to values obtained in the literature.

B. Segmentation

On this study, the CNN models chosen for the segmentation task were FPN, Mask R-CNN, U-Net and PSPNet, based on relevant segmentation results in literature [15]–[17]. The FPN model takes a single-scale image as input and outputs proportionally sized feature maps at multiple levels in a fully convolutional fashion [18]. The construction of the pyramid involves a bottom-up pathway, a top-down pathway, and lateral connections. The bottom-up pathway uses the feature maps generated by a backbone to create hierarchical maps at different scales. The top-down pathway generates higher resolution features by upsampling feature maps from higher pyramid levels. These features are then enhanced via lateral connections that merge feature maps of the same spatial size from the bottom-up pathway and the top-down pathway. The network architecture is depicted in Figure 2. An FPN takes advantage of the hierarchical and multi-scale nature of convolutional neural networks to generate useful features for object segmentation at different scales [19].

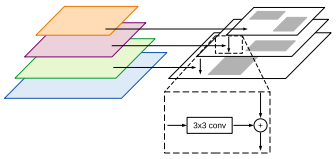


Fig. 2: Illustration model of the FPN architecture.

The Mask R-CNN model [20] is a framework designed for object detection and instance segmentation, which has shown promising results in medical image segmentation [21], [22]. The model is composed of three main components: backbone, feature pyramid network (FPN) and region proposal network (RPN), as depicted in Fig. 3.

First, convolutional layers from ResNet50 were used to build a FPN structure. The features obtained by the FPN are given as the input to a RPN, which identifies candidate objects. A convolutional layer of size 3×3 was employed in each layer of the FPN and the resulting values were employed in two fully connected layers, generating the bounding boxes for each region. Then, a fully connected layer was used on the object feature maps to define the binary masks for each nucleus in the histological image.

U-Net architecture consists of a convolution path and a deconvolution path, where the convolution path follows the typical architecture of a convolutional network [23]. It uses the convolutional layers of the backbone model to extract features from the images. In the deconvolution path, the image is upsampled to its original size, allowing precise localization of high-resolution features. To do this, every step of this path consists of an upsamplig of the feature map, a

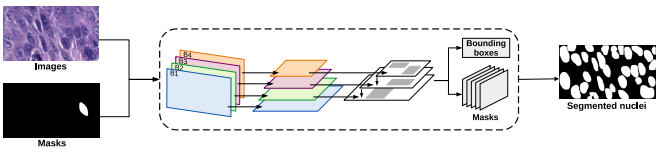


Fig. 3: Architecture illustration of the Mask R-CNN employed in the OEDs segmentation process.

concatenation with the correspondingly cropped feature map from the convolution path, and two kernel convolutions, each followed by a ReLU. The cropping is necessary because of the loss of border pixels in every convolution. At the final layer a 1x1 convolution is used to map each feature vector to the desired number of classes. This networks architecture is shown in Fig. 4. In this study, we employed different backbones to evaluate the U-net.

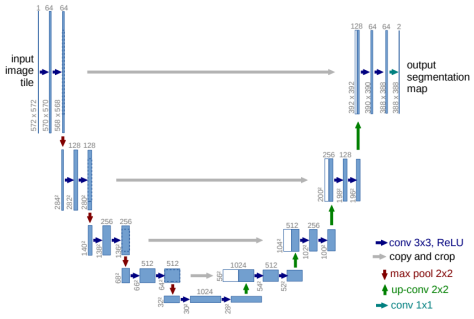


Fig. 4: Illustration model of the U-Net architecture [Source: Ronneberger et al. [23]].

PSPNet is a model that comprises a convolution path and a decoding path [24]. The convolution path extracts image features from different levels by employing layers from the backbone model and creating a feature pyramid at different scales. The decoding path generates a segmentation output by aggregating these features at different scales, and each step in the decoding path expands the feature maps, concatenates them with the corresponding feature map from the convolution path, and performs two convolutions. This concatenation improves segmentation by combining information from different scales, and the feature maps generated by the decoding path are concatenated with the maps from the encoder path, aggregating features from various scales. Finally, the classification layer performs semantic segmentation on the feature maps. Fig. 5 illustrates the architecture of the PSPNet model.

In this study, the CNN models were trained using the ResNet50 and the MobileNetV3 backbone architectures, with the exception of the Mask R-CNN model, which used only the ResNet50. These backbone models were chosen due to promissing characteristics in segmentation tasks, as described in the literature [25], [26]. All models were pre-trained on the ImageNet dataset and fine-tuned to our dataset. The images were split on the proportion of 60% for training set, 10% for validation set and 30% for test set, totalling 272, 46 and 138

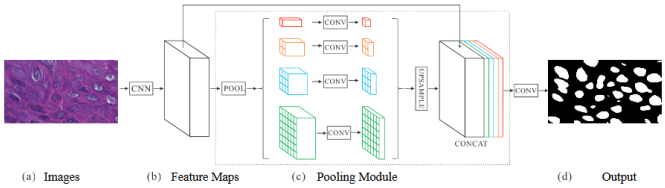


Fig. 5: PSPnet model architecture used for nucleus segmentation. Image adapted from [24].

images for each set, respectively. During the fine-tuning stage, the models were trained with 40 epochs, with 150 iterations for each epoch, learning rate of 0.001, and a momentum value of 0.9, using the Nadam optimizer. All models were trained using these values, to ensure consistency at this stage. Different values were tested during this stage to identify the optimal combination between the structures and background of the investigated images.

C. Post Processing

The resulting images may contain incomplete nuclei regions or small artifacts. To address this issue, morphological operations were applied. In order to close the contour of the nuclei, the dilation operation was performed using a cross-shaped structuring element with a size of 3×3 pixels. Subsequently, a hole-fill function was used on the binary objects. Then, an erosion operation with the structuring element of 3×3 pixels was applied. Finally, objects with an area smaller than 30 pixels were classified as background regions and were removed from the masks. The parameters used in this stage were chosen empirically.

D. CNN Ensembles

CNN ensembles are approaches used to combine models results and have proven to be a promising approach for improving image segmentation [27]–[29]. The combination of different CNN models can generate a more accurate and robust final segmentation. The ensemble strategies used in this study were the simple averaging, weighted averaging, and the majority voting.

Simple averaging is a widely used combination technique in segmentation problems with multiple CNN models. In this strategy, the final decision is obtained by averaging the results of individual models [30]. If the sum result is higher than the average, the pixel is classified as a region of interest; otherwise, it is classified as a background region. Simple averaging is especially useful for binary data. The mathematical formula for the simple averaging is given by Eq. 1.

$$A(x) = \frac{1}{N}\sum_{i=1}^{N} a_i(x), \tag{1}$$

where $A(x)$ is the final result, N is the number of individual models and $a_i(x)$ is the output generated by the i-th model.

Weighted averaging is an extension of the simple averaging strategy, in which the results of individual models are averaged

with different weights [31]. The weights can be assigned according to the model's performance on a validation set or the model's complexity. This ensemble method is defined by Eq. 2.

$$W_A(x) = \sum_{i=1}^{N} w_i a_i(x), \tag{2}$$

where $W_A(x)$ is the final result, $a_i(x)$ is the output generated by the i-th model and w_i is the weight assigned to model a_i.

Majority voting takes into account the most voted class by each model. In this method, binary images are combined by counting votes for each class at each pixel, resulting in a final binary image where each pixel is assigned to the class with the majority of votes [32]. This voting rule is given by Eq. 3.

$$M(x) = \begin{cases} 1 & \text{if } \sum_{i=1}^{N} a_i(x) > N \\ 0 & \text{otherwise} \end{cases}, \tag{3}$$

where $M(x)$ is the final output, N is the number of models and $a_i(x)$ is the binary image generated by model i.

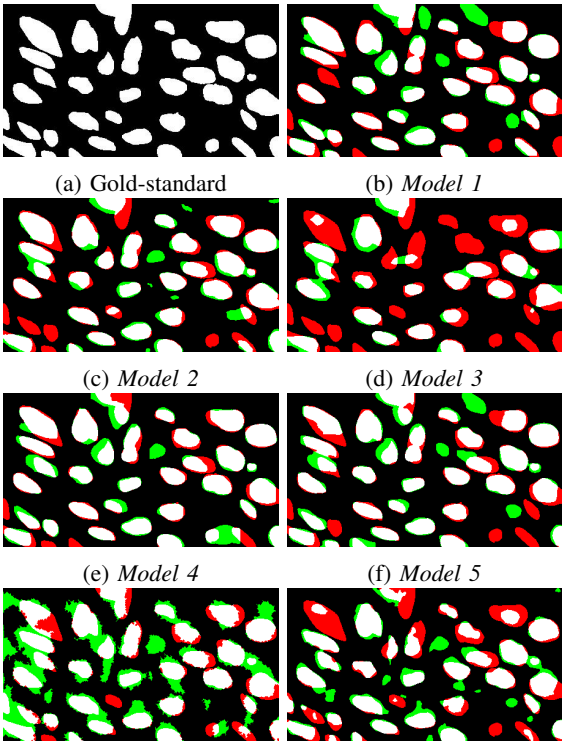
E. Evaluation Metrics

The segmentation method was evaluated by comparing the result image with the gold-standard and calculating the overlapping regions [33]. The evaluation metrics employed were accuracy (A_{CC}), sensitivity (S_E), specificity (S_P) and Dice coefficient(D_C) [34], [35].

The metric of A_{CC} was used to quantify the amount of pixels correctly segmented as nuclei and background in relation to all pixels in the image. The S_E measures how well the method segmented the nuclei regions, while the S_P measures the proportion of pixels correctly identified as the background region. The D_C measure was used to evaluate the overlapping regions between the mask outputs and the gold-standard.

III. Experimental Results

Fig. 6 displays the masks generated by the different models on an randomly chosen mild OED image. On this image, the green regions indicate FP areas and red regions indicate FN. The CNN architectures and backbones were combined as follows: Mask R-CNN - ResNet50 (*Model 1*), PSPnet - ResNet50 (*Model 2*), PSPnet - MobileNetV3 (*Model 3*), FPN - ResNet50 (*Model 4*), FPN - MobileNetV3 (*Model 5*), U-Net - ResNet50 (*Model 6*) and U-Net - MobileNetV3 (*Model 7*). *Model 1* was able to identify nuclear structures and generated results similar to the gold-standard. However, it presented regions of FP in the form of artifacts in the image and regions of FN in the form of unidentified nuclei. *Models 2* and *3* produced FN regions, resulting in the degradation of nuclei edges and the limited segmentation of nuclei in the case of *Model 3*. Similarly to *Model 1*, *Models 4* and *5* identified and segmented cellular nuclei, with results close to the gold-standard, but with some FN regions and degraded contours on some structures. *Models 6* and *7* generated numerous FP regions, segmenting regions near nuclei as areas of interest and showing artifacts. It is important to note that *Models 1*, *4*, and *5* yielded visually similar outcomes.



(a) Gold-standard (b) *Model 1* (c) *Model 2* (d) *Model 3* (e) *Model 4* (f) *Model 5* (g) *Model 6* (h) *Model 7*

Fig. 6: Visual comparison between the gold-standard and nuclei masks generated by CNN models. Green region indicate FP and red regions indicate FN.

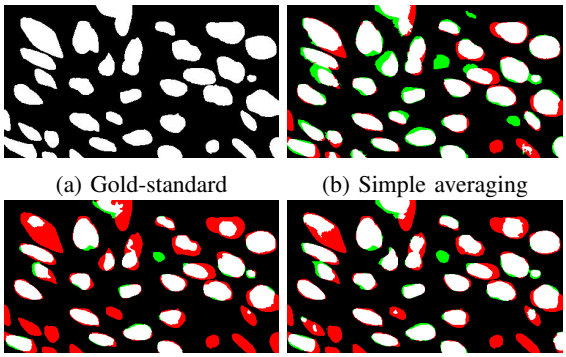
Table I presents a quantitative analysis of the method's performance on the gold-standard images, showing the average values obtained with the metrics for the OED classes. It is noteworthy that *Model* 4 obtained the highest values, with $A_{CC} = 91.95\%$ and $D_C = 0.87$. *Models* 1, 5, and 6 presented A_{CC} values close to each other, but with a drop in the D_C measure. It is important to highlight the relationship between *Models* 4 and 5, as both models are based on FPN, and between *Models* 4 and 1, as Mask R-CNN is an extension of FPN. Therefore, these close results may be a result of this relationship. The values yelded by *Models* 2, 3 and 7 were lower, with A_{CC} of 84.72%, 80.83%, and 85.48%, respectively. The results suggest that models using ResNet50 backbone achieved higher values than models employing MobileNetV3. Furthermore, it is observed that models based on FPN obtained better results according to the evaluation metrics.

Fig. 7 shows the segmentation masks obtained by each ensemble approach. Visually, the results obtained by the simple averaging present outcomes similar to the gold-standard, with few occurences FP (indicated by green regions) and FN (red regions). However, some deformations can be observed at some nuclei borders. In contrast, the weighted averaging and majority voting present nuclei with deformed contours, unsegmented objects, and the presence of small artifacts in

TABLE I: Evaluation of CNN models for segmentation of nuclei in histological tissues.

Model	A_{CC} (%)	S_E (%)	S_P (%)	D_C
Model 1	89.49	84.55	91.83	0.84
Model 2	84.72	81.47	86.64	0.77
Model 3	80.83	72.84	84.76	0.71
Model 4	91.95	87.99	92.76	0.87
Model 5	89.32	82.79	92.55	0.83
Model 6	89.31	78.50	94.67	0.83
Model 7	85.48	87.88	82.97	0.80

the images.



(a) Gold-standard (b) Simple averaging

(c) Weighted averaging (d) Majority voting

Fig. 7: Visual comparison between the gold-standard and nuclei masks generated by the ensemble methods.

Table II, depicts the evaluation values obtained for the ensemble strategies in comparison to the gold-standard. The results show that the simple averaging approach yielded higher values than the other methods, achieving A_{CC} of 92.14% and D_C of 0.86. The weighted averaging resulted in A_{CC} and D_C values of 91.21% and 0.85, respectively. The majority voting obtained 90.57% and 0.85 for these metrics. It can be observed that the ensemble approaches outperformed many of the individual CNN models in terms of A_{CC}. The weighted averaging and majority voting achieved higher values compared to all CNN models, except for *Model 4*. The evaluated ensemble strategies provided S_E values higher than most of the models, indicating that the detection of nuclei region improved. These ensemble methods shown D_C values higher than most of the individual models.

TABLE II: Evaluation of ensemble strategies.

Ensemble Strategy	A_{CC} **(%)**	S_E **(%)**	S_P **(%)**	D_C
Simple averaging	92.14	88.12	91.94	0.86
Weighted averaging	91.21	86.82	91.06	0.85
Majority Voting	90.67	84.62	89.94	0.85

The simple averaging method was evaluated on three distinct datasets, namely, CryoNuSeg [12], MoNuSeg 2018 [13], and GlaS [14]. This evaluation is presented in Table III, which shows results obtained by algorithms present in the literature. For CryoNuSeg, the proposed method achieved a D_C value of 0.86, surpassing Mahbod et al.'s method [12] and showing comparable results with Hassan et al. [36] and Sitnik et al. [37]. Similarly, for the MoNuSeg 2018 dataset, the ensemble achieved a value higher than the methods proposed by Liao et al. [38] and Li et al. [39], but lower compared to the method by Hassan et al. [36]. For GlaS, the proposed method presented D_C value of 0.86, being comparable to the algorithms of Liao et al. [38] and Sirinyukunwattana et al [14]. Overall, the proposed ensemble model demonstrated similar performance to state of the art algorithms.

TABLE III: Evaluation between proposed ensemble strategy and segmentation methods in the literature (D_C).

Model	**CryoNuSeg**	**MoNuSeg 2018**	**GlaS**
Mahbod et al. [12]	0.78	-	-
Sitnik et al. [37]	0.85	-	-
Hassan et al. [36]	0.86	0.89	-
Li et al. [39]	-	0.79	-
Liao et al. [38]	-	0.69	0.85
Sirinyukunwattana et al. [14]	-	-	0.87
Proposed - Simple averaging	0.86	0.86	0.86

IV. Conclusions

This study proposed a method for cell nuclei segmentation in histological images of OED using CNN ensembles. Seven CNN models were trained to segment nuclei regions and morphological operations were used to remove artifacts. Three ensemble strategies were employed to improve the models results. The ensemble systems improved the CNNs segmentation, achieving A_{CC} of 92.14%, 91.21%, and 90.57% for simple averaging, weighted averaging, and majority voting, respectively, demonstrating the potential of the proposed methodology to minimize segmentation errors. When the method was applied to datasets in other domains, it was able to provide relevant results compared to values present in the state of the art.

Furthermore, a comparison of the results with those of public datasets revealed that the proposed segmentation ensembles achieved comparable results with existing methods from the literature. However, there is still room for improvement in the proposed method, and future studies will explore the impact of data augmentation on the CNNs segmentation. Additionally, other ensemble approaches and CNN models will be investigated for further improvement of the segmentation results.

The results of this study can be used to assist pathologists during the histological analysis of OED lesions. The proposed method accuracy on cell nuclei segmentation on OED histological images shows that the development CAD systems can improve the accuracy of nuclei segmentation on OED lesion images.

Acknowledgment

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. The authors gratefully acknowledge the financial support of National Council for Scientific and Technological Development CNPq (Grants #313643/2021-0, #311404/2021-9 and #307318/2022-2), the State of Minas

Gerais Research Foundation - FAPEMIG (Grant #APQ-00578-18 and Grant #APQ-01129-21) and São Paulo Research Foundation - FAPESP (Grant #2022/03020-1).

REFERENCES

[1] K. Wan and D. Kademani, "Oral epithelial dysplasia," in *Improving Outcomes in Oral Cancer*. Springer, 2020, pp. 1–12.

[2] H. Irshad, A. Veillard, L. Roux, and D. Racoceanu, "Methods for nuclei detection, segmentation, and classification in digital histopathology: a review—current status and future potential," *IEEE reviews in biomedical engineering*, vol. 7, pp. 97–114, 2014.

[3] A. Belsare, M. Mushrif, M. Pangarkar, and N. Meshram, "Breast histopathology image segmentation using spatio-colour-texture based graph partition method," *Journal of microscopy*, vol. 262, no. 3, pp. 260–273, 2016.

[4] R. C. Gonzalez and R. Woods, "Digital image processing," 2018.

[5] V. Kumar, J. C. Aster, and A. Abbas, *Robbins and Cotran Patologia-Bases Patológicas das Doenças*. Elsevier Brasil, 2010.

[6] R. B. Tokime, H. Elassady, and M. A. Akhloufi, "Identifying the cells' nuclei using deep learning," in *2018 IEEE Life Sciences Conference (LSC)*. IEEE, 2018, pp. 61–64.

[7] D. F. dos Santos, T. A. Tosta, A. B. Silva, P. R. de Faria, B. A. Travençolo, and M. Z. do Nascimento, "Automated nuclei segmentation on dysplastic oral tissues using cnn," in *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*. IEEE, 2020, pp. 45–50.

[8] A. J. Shephard, S. Graham, S. Bashir, M. Jahanifar, H. Mahmood, A. Khurram, and N. M. Rajpoot, "Simultaneous nuclear instance and layer segmentation in oral epithelial dysplasia," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 552–561.

[9] A. B. Silva, A. S. Martins, T. A. A. Tosta, L. A. Neves, J. P. S. Servato, M. S. de Araújo, P. R. de Faria, and M. Z. do Nascimento, "Computational analysis of histological images from hematoxylin and eosin-stained oral epithelial dysplasia tissue sections," *Expert Systems with Applications*, p. 116456, 2022.

[10] X. Ying, "An overview of overfitting and its solutions," in *Journal of Physics: Conference Series*, vol. 1168. IOP Publishing, 2019, p. 022022.

[11] H. Lumerman, P. Freedman, and S. Kerpel, "Oral epithelial dysplasia and the development of invasive squamous cell carcinoma," *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology*, vol. 79, no. 3, pp. 321–329, 1995.

[12] A. Mahbod, G. Schaefer, B. Bancher, C. Löw, G. Dorffner, R. Ecker, and I. Ellinger, "Cryonuseg: A dataset for nuclei instance segmentation of cryosectioned h&e-stained histological images," *Computers in biology and medicine*, vol. 132, p. 104349, 2021.

[13] N. Kumar, R. Verma, D. Anand, Y. Zhou, O. F. Onder, E. Tsougenis, H. Chen, P.-A. Heng, J. Li, Z. Hu *et al.*, "A multi-organ nucleus segmentation challenge," *IEEE transactions on medical imaging*, vol. 39, no. 5, pp. 1380–1391, 2019.

[14] K. Sirinukunwattana, J. P. Pluim, H. Chen, X. Qi, P.-A. Heng, Y. B. Guo, L. Y. Wang, B. J. Matuszewski, E. Bruni, U. Sanchez *et al.*, "Gland segmentation in colon histology images: The glas challenge contest," *Medical image analysis*, vol. 35, pp. 489–502, 2017.

[15] G. Jader, J. Fontineli, M. Ruiz, K. Abdalla, M. Pithon, and L. Oliveira, "Deep instance segmentation of teeth in panoramic x-ray images," in *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. IEEE, 2018, pp. 400–407.

[16] E. K. Wang, C.-M. Chen, M. M. Hassan, and A. Almogren, "A deep learning based medical image segmentation technique in internet-of-medical-things domain," *Future Generation Computer Systems*, vol. 108, pp. 135–144, 2020.

[17] M. Ji and Z. Wu, "Automatic detection and severity analysis of grape black measles disease based on deep learning and fuzzy logic," *Computers and Electronics in Agriculture*, vol. 193, p. 106718, 2022.

[18] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2117–2125.

[19] M. Kisantal, Z. Wojna, J. Murawski, J. Naruniec, and K. Cho, "Augmentation for small object detection," *arXiv preprint arXiv:1902.07296*, 2019.

[20] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.

[21] X. Wang, S. Han, Y. Chen, D. Gao, and N. Vasconcelos, "Volumetric attention for 3d medical image segmentation and detection," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 175–184.

[22] K. Long, L. Tang, X. Pu, Y. Ren, M. Zheng, L. Gao, C. Song, S. Han, M. Zhou, and F. Deng, "Probability-based mask r-cnn for pulmonary embolism detection," *Neurocomputing*, vol. 422, pp. 345–353, 2021.

[23] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[24] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2881–2890.

[25] A. Arlova, C. Jin, A. Wong-Rolle, E. S. Chen, C. Lisle, G. T. Brown, N. Lay, P. L. Choyke, B. Turkbey, S. Harmon *et al.*, "Artificial intelligence-based tumor segmentation in mouse models of lung adenocarcinoma," *Journal of Pathology Informatics*, vol. 13, p. 100007, 2022.

[26] S. H. Kassani, P. H. Kassani, M. J. Wesolowski, K. A. Schneider, and R. Deters, "Deep transfer learning based model for colorectal cancer histopathology segmentation: A comparative study of deep pre-trained models," *International Journal of Medical Informatics*, vol. 159, p. 104669, 2022.

[27] A. O. Vuola, S. U. Akram, and J. Kannala, "Mask-rcnn and u-net ensembled for nuclei segmentation," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE, 2019, pp. 208–212.

[28] J. Dolz, C. Desrosiers, L. Wang, J. Yuan, D. Shen, and I. B. Ayed, "Deep cnn ensembles and suggestive annotations for infant brain mri segmentation," *Computerized Medical Imaging and Graphics*, vol. 79, p. 101660, 2020.

[29] E. B. Kablan, H. Dogan, M. E. Ercin, S. Ersoz, and M. Ekinci, "An ensemble of fine-tuned fully convolutional neural networks for pleural effusion cell nuclei segmentation," *Computers & Electrical Engineering*, vol. 81, p. 106533, 2020.

[30] Z.-H. Zhou, *Ensemble methods: foundations and algorithms*. CRC press, 2012.

[31] L. Nanni, A. Lumini, and S. Ghidoni, "Ensemble of deep learned features for melanoma classification," *arXiv preprint arXiv:1807.08008*, 2018.

[32] B. Krawczyk, L. L. Minku, J. Gama, J. Stefanowski, and M. Woźniak, "Ensemble learning for data stream analysis: A survey," *Information Fusion*, vol. 37, pp. 132–156, 2017.

[33] F. J. Estrada and A. D. Jepson, "Benchmarking image segmentation algorithms," *International Journal of Computer Vision*, vol. 85, no. 2, pp. 167–181, 2009.

[34] A. Baratloo, M. Hosseini, A. Negida, and G. El Ashal, "Part 1: simple definition and calculation of accuracy, sensitivity and specificity," *Emergency*, vol. 3, no. 2, pp. 48–49, 2015.

[35] Z. Ma, X. Wu, Q. Song, Y. Luo, Y. Wang, and J. Zhou, "Automated nasopharyngeal carcinoma segmentation in magnetic resonance images by combination of convolutional neural networks and graph cut," *Experimental and therapeutic medicine*, vol. 16, no. 3, pp. 2511–2521, 2018.

[36] L. Hassan, M. Abdel-Nasser, A. Saleh, O. A. Omer, and D. Puig, "Efficient stain-aware nuclei segmentation deep learning framework for multi-center histopathological images," *Electronics*, vol. 10, no. 8, p. 954, 2021.

[37] D. Sitnik and I. Kopriva, "Lefm-nets: Learnable explicit feature map deep networks for segmentation of histopathological images of frozen sections," *arXiv preprint arXiv:2204.06955*, 2022.

[38] Z. Liao, K. Xu, and N. Fan, "Swin transformer assisted prior attention network for medical image segmentation," in *Proceedings of the 8th International Conference on Computing and Artificial Intelligence*, 2022, pp. 491–497.

[39] X. Li, Y. Wang, Q. Tang, Z. Fan, and J. Yu, "Dual u-net for the segmentation of overlapping glioma nuclei," *Ieee Access*, vol. 7, pp. 84 040–84 052, 2019.