

Self-supervised learning of a tailored Convolutional Auto Encoder for histopathological prostate grading.

Zahra Tabatabaei^{*§}, Adrián Colomer^{*†}, Kjersti Engan[‡], Javier Oliver[§], Valery Naranjo^{*}
Instituto Universitario de Investigación en Tecnología Centrada en el Ser Humano,
^{*} HUMAN-tech, Universitat Politècnica de València, Spain
[†] ValgrAI – Valencian Graduate School and Research Network for Artificial Intelligence
[‡] Department of Electrical Engineering and Computer Science, University of Stavanger, Norway
[§] Department of Artificial Intelligence, Tyris Tech S.L., Valencia, Spain

Abstract—According to GLOBOCAN 2020, prostate cancer is the second most common cancer in men worldwide and the fourth most prevalent cancer overall. For pathologists, grading prostate cancer is challenging, especially when discriminating between Grade 3 (G3) and Grade 4 (G4). This paper proposes a Self-Supervised Learning (SSL) framework to classify prostate histopathological images when labeled images are scarce. In particular, a tailored Convolutional Auto Encoder (CAE) is trained to reconstruct $128 \times 128 \times 3$ patches of prostate cancer Whole Slide Images (WSIs) as a pretext task. The downstream task of the proposed SSL paradigm is the automatic grading of histopathological patches of prostate cancer. The presented framework reports promising results on the validation set, obtaining an overall accuracy of 83% and on the test set, achieving an overall accuracy value of 76% with F1-score of 77% in G4.

Index Terms—Classification, Convolutional Auto Encoder (CAE), Histopathological images, Prostate cancer, Whole Slide Images (WSIs).

I. INTRODUCTION

Prostate cancer is the fifth most deadly disease, according to GLOBOCAN 2020, having an estimation of 375,304 cases. Prostate cancer tends to spread slowly, and in some cases, men who died of other diseases also had prostate cancer that had never impacted them in life [1]. In the traditional cancer diagnosis, if the therapist confirms prostate cancer based on blood tests, small portions of the cancerous tissue are extracted. These tissues need to pass some processing, and then they should be stained. Hematoxylin and Eosin (H&E) are one of the most widely used materials in staining tissues, and it provides pink and purple slides for pathologists [2]. Pathologists grade the tissues by examining them under a microscope to analyze how the glands, nuclei, and lumen are organized. Prostate cancer has four classes Non-Cancerous (NC), Grade 3 (G3), Grade 4 (G4), and Grade 5 (G5). The presence of one or more Gleason patterns in the tissue regions is an interesting element in diagnosing the Gleason score. This score defines the proper treatment to apply; however, in some cases detecting the abnormalities is complicated, and it is not

as easy as looking at the microscope [3]. To tackle these difficulties, the Computer-Aided Diagnosis system (CAD) provides a wide diversity of tools on scanned Whole Slide Images (WSIs) to increase the accuracy of a cancer diagnosis. These tools use Deep Learning (DL) and convolutional layers to extract features of the histopathological images, including color, texture, shape, size, etc. These features feed to fully connected layers to train the model that classifies the patches into different grades. Different cancer types have varying grades, and some of them only fall into the malignant and benign categories, necessitating a binary classification. To categorize the patches into more than two classes, multi-class classifiers are required.

The analysis of histopathological images has been done in previous literature using a variety of techniques. Spanhol, [5] proposed a method to allow using the high-resolution histopathological images from the BreakHis data set without computational costs in the architecture. Liu, [6] proposed an improved Auto Encoder (AE) network to have an automated binary classification on the BreakHis data set. Silva-Rodríguez, [3] proposed a patch-wise predictive model based on a Convolutional Neural Network (CNN) to grade the WSIs relating to the presence of cancerous patterns. Schmidt, [7] presented cancer classification by coupling semi-supervised and multiple instances learning on prostate and breast cancer. Kalapahar, [8] fed residual U-NET by SICAPv2 to grade the WSIs through semantic segmentation. Due to the similarities of some Gleason patterns in prostate histopathological tissues, multi-class classification is an open challenging task. Whereas both pathologists and DL techniques find it more challenging to grade prostate tissues into G3 and G4 than other grades, G3 and G4 need different medical treatments.

In this work, we propose a novel Self-Supervised Learning (SSL) strategy to classify prostate patches into four grades. Our proposed SSL classifier, in contrast with the methods cited above, can mitigate this problem with an accurate delimitation of the different grades. More precisely, we train an SSL classifier containing a tailored CAE in its pretext task and a stack of fully connected layers as a classifier module for

the downstream task in the place of the decoder in the CAE. Hence, the proposed classifier can detect very slight distinctions between grades, and it can provide a precise delineation of the complex patterns of G3 and G4. Compared to the prior patch-level techniques, the proposed SSL classifier performs comparably well at distinguishing G3 and G4.

II. MATERIAL

SICAPv2 is the biggest public data set of prostate cancer containing pixel-wise annotations of the tumor regions following the Gleason score scale. This contains 182 prostate biopsies from 96 patients. A group of pathologists from Hospital Clínico of Valencia analyzed the H&E stained images at $40\times$ magnification [4]. Table I describes the number of WSIs and patches in SICAPv2 per grade.

One of the challenges in working with WSIs is their big size. Thus, they were re-sampled into 10x magnification and divided into patches of $512\times 512\times 3$ pixels by a sliding window mechanism with 50% of overlapping [3]. In this paper, regarding constraints computation time, the images were resized to $128\times 128\times 3$.

TABLE I: SICAPv2 data set description.

Grades	NC	G3	G4	G5
WSIs	37	60	69	16
Patches	4417	1636	3622	655

III. METHOD

The proposed end-to-end patch-level framework based on a tailored CAE trained in an SSL strategy is depicted in Fig. 1 and Fig. 2.

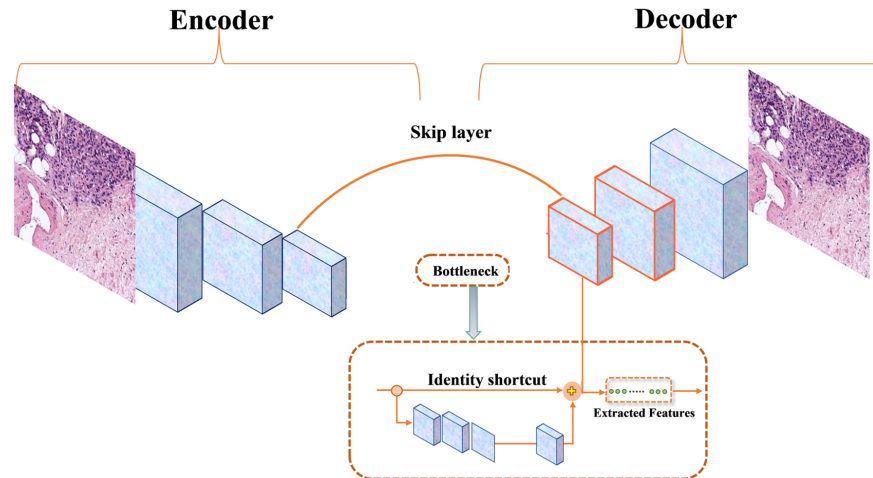


Fig. 1: The structure of the proposed CAE as the pretext task. The intermediate layers consist of a skip layer to transfer information between corresponding encoder-decoder blocks [4]. The bottleneck has four convolutional layers with stride = 1 and kernel size = (3, 3).

A. Self Supervised Learning (SSL)

Supervised DL has achieved great success in the last decade, but it depends on manual labeling and vulnerability. Learning with scarce labeled data is a fundamental problem in DL. Also, it is critical for histopathological image analysis because annotating medical images is time-consuming and expensive. The emergence of SSL tackled this problem. SSL has soared in performance on representation learning in the last several years. SSL is appealing because it allows for the pre-training usage of unlabeled domain-specific images in order to learn more appropriate representations [9]. It derives its labels from a co-occurring modality for the given data. To be able to understand the geometry of the input, SSL learns the representation of the data by observing different parts of it [10]. Generally, SSL includes a pretext and a downstream (real) task. In the pretext task, the visual representations are learned to have the obtained model weights for the downstream task. SSL leverages input data as supervision in pretext and benefits almost all types of downstream tasks. Then, the representations in the downstream task can be fine-tuned depends on the goal of the system with a few labeled images [11], [12]. In this work, the downstream task is classification with insufficient annotated patches.

B. Pretext and downstream tasks for prostate cancer gradation

The intuition behind the proposed method is image reconstruction by a tailored Convolutional Auto Encoder (CAE) and classification by a multi-layer perceptron from a latent space as its pretext and downstream task, respectively. Pretext task has three benefits in our work including the classifier's initial weights are not random, the classifier is aware of the general characteristics of the WSIs before the first epoch, and it takes

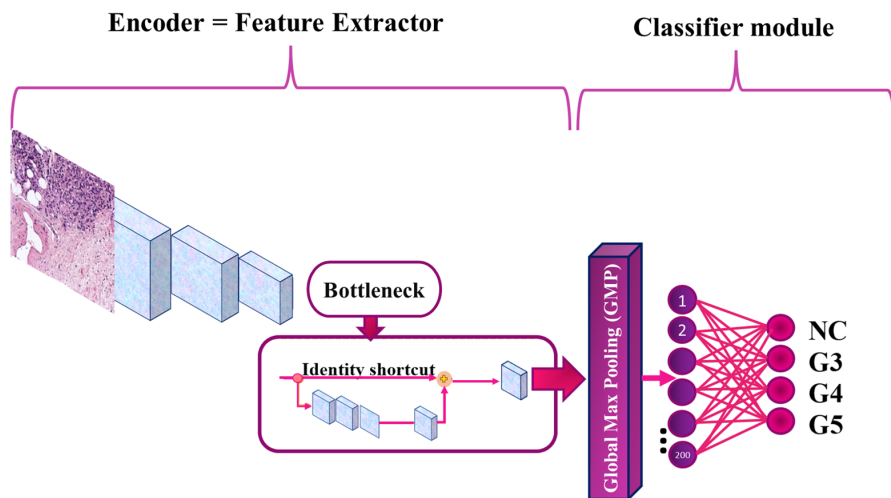


Fig. 2: The structure of the proposed classifier on SICAPv2 as the downstream task (real task). The extracted features from the pretext task are fed into the downstream task, and the output layer presents one neuron per target grade with soft-max activation.

less time to categorize the images. These benefits are attained because, in contrast to other transfer learning techniques, our model is fed by histopathological images in pretext.

We proposed a tailored CAE in the pretext task to deal with the complexity of the histopathological images. As can be seen in Fig. 1 an encoder’s middle block is connected to the corresponding decoder’s block through a skip connection. In order to extract more representative features, a residual block is added to the bottleneck. Notably, the tailored CAE compresses the input and reconstructs it using convolutional layers without a pooling layer. In the pretext stage, the decoder has thus been dropped after pre-training, and the encoder and bottleneck are set as the Feature Extractor (FE) for the classifier in the downstream task. In the downstream task, FE and the classifier module have been trained together to minimize the amount of categorical cross-entropy in the loss function. The architecture of the proposed classifier in Fig. 2 demonstrates how each image in the input is categorized into the corresponding label.

IV. EXPERIMENTS AND RESULTS

We perform a series of experiments to train and validate the proposed classifier by splitting the data set into 80:20 for

training and validating, respectively. In order to perform a comparison of the different classifiers described previously, the test set is completely isolated.

In particular, the tailored CAE was trained using Adam optimizer with a learning rate of 5×10^{-4} and a batch size of 16 images. Then, in the downstream task, weights of the first 29 levels of the encoder to the first 29 layers of the classifier were loaded. The bottleneck’s output was fed into the classifier component, which has a Global Max Pooling 2D (GMP) layer and two dense layers [200, 4] to obtain the probability distribution of four categories. In downstream, the classifier has with a batch size of 8 images, a learning rate of 0.5, and an SGD optimizer.

To establish a proper comparison with previous work, five indicators were selected, including accuracy (ACC) [13], F1-score (F1) [14], F1-avg, Kappa (k), and Confusion Matrix (CM). Table III and Table II summarize that our model reaches a promising performance on both validation and test set while it does not need any pre-trained model with a huge number of images. Comparing the k values in the test set, which is reported in Table III demonstrates how our suggested method is superior to recent works.

In Table II, the performance of different models such as

TABLE II: The obtained results of patch-level classification by our proposed classifier in the validation set of SICAPv2. The presented metrics are accuracy (ACC), F1-Score (F1S) computed per class and its average (F1-avg), and Cohen’s quadratic kappa (k).

Method	ACC		F1S		F1-avg		k
Ours	0.83	0.92	0.73	0.82	0.67	0.78	0.76
VGG19 + FC [3]	0.72	0.88	0.66	0.60	0.52	0.66	0.73
VGG19 + GMP [3]	0.72	0.87	0.64	0.60	0.54	0.66	0.71
ResNet + FC [3]	0.69	0.83	0.66	0.57	0.48	0.64	0.68
ResNet + GMP [3]	0.68	0.83	0.64	0.55	0.50	0.63	0.67

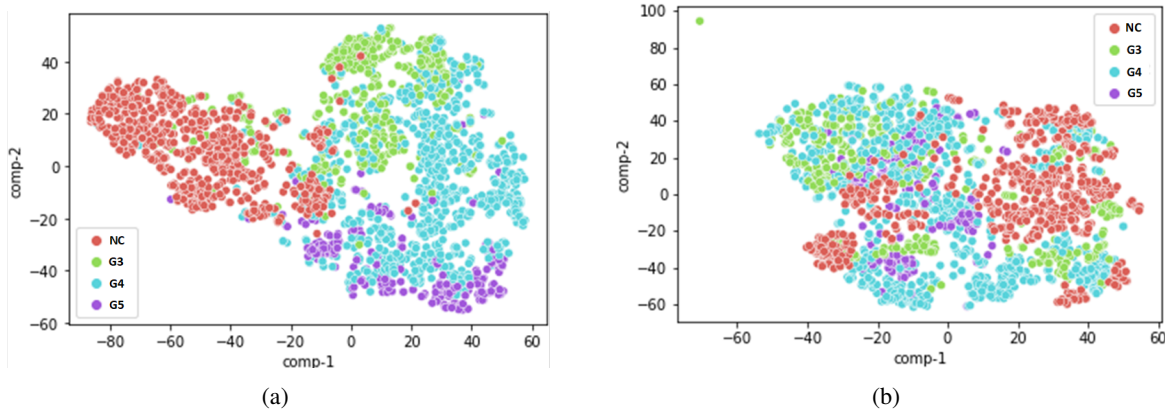


Fig. 3: TSNE projection of the extracted features of the test set in a 2D representation. Fig. 3(a) and Fig. 3(b) respectively represent the feature space as a result of the proposed method and Schmidt, Arne (2022) [7].

TABLE III: The obtained results of patch-level classification by our proposed classifier in the test set of SICAPv2. The presented metric is Cohen’s quadratic kappa (k).

Method	k
Ours	0.80
FSconv+GMP (2020) [3]	0.77
Nir et al. (2018) [15]	0.61
Otalora et al. (2020) [16]	0.55/0.59
Inter-Pathologists (2018) [17]	0.65
Arviniti et al. (2018) [17]	0.55/0.49

ResNet, VGG19, and FSConv are presented with different configurations of top models. We observe that our model can achieve competitive results with respect to other methods, such as VGG19, that had been pre-trained on a large number of non-histopathological images. Our SSL classifier measures k better than more contemporary approaches, as shown in Table II, which supports the efficacy of the suggested SSL algorithm on a small training data set. The F1S for each grade in the validation set demonstrates that the proposed method is more effective in distinguishing between the four grades, particularly in the challenging task of differentiating between G3 and G4. This suggests that the model is more robust in its ability to classify the grades accurately.

In order to analyze the correlation of the extracted features for each class, T-distributed Stochastic Neighbor Embedding (TSNE) technique was applied to them, Fig. 3. Fig. 3(a) shows a 2D TSNE plot, representing features of the four classes under this study. The plot shows a discriminative behavior of the proposed framework, allowing a promising gradation of prostate histopathological patches with a F1-avg value of 72% on the test set. As can be seen in Fig. 3(a) compared to Fig3(b)¹, our proposed model can reduce the inter-distance and can increase intra-distance between features of each class

¹Fig. 3(b) was plotted based on the features shared by Schmidt, Arne [7]

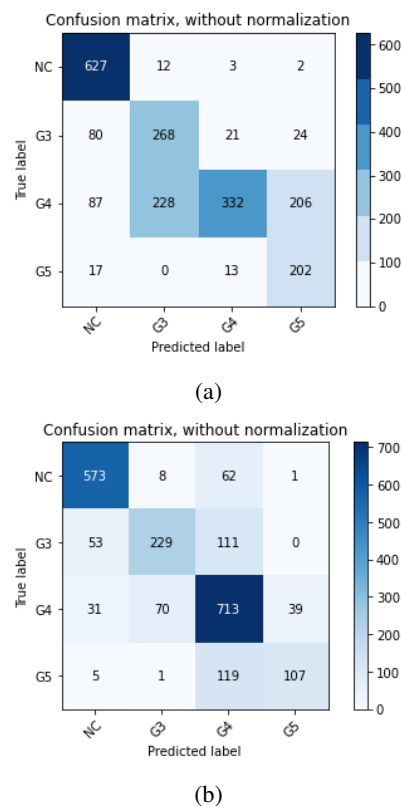


Fig. 4: CMs of the grading by (a) proposed method in [3] and (b) our proposed classifier.

much better than the recent work [7]. In Fig. 3(a), the extracted features of G4 are more spread in the feature space, while it brings F1S equal to 77% on the test set, which is higher than compared methods.

Most notably, from a pathologist’s point of view, G3 and G4 are slightly different. However, G3 must take medication, and G4 must have surgery. This emphasizes the significance of accurately grading G3 and G4. In Fig. 4 CMs show that the

model classifies the images into four classes more accurately and it can surpass and outperform the reported CM in [3], especially in grading G3 and G4.

V. CONCLUSIONS

In this paper, we have presented a novel end-to-end DL structure that is able to classify a multi-class data set without the need for a pre-trained model on a pool of images. The proposed SSL method enables learning the first meaningful features of WSIs by training a tailored CAE to reconstruct the images in the pretext task. In our downstream task, we can get a classification accuracy of 76% in the test set, surpassing competing methods in the same and other data sets and indicating the value of our method. The highlighted power point in our method is an accurate delimitation of the complex patterns of G3 and G4 in prostate tissues resulting in an FIS value of 77%, which outperforms the previous works.

VI. ACKNOWLEDGMENTS

In this study, Zahra Tabatabaei is funded by European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 860627 (CLARIFY Project).

The work of Adrián Colomer has been supported by the ValgrAI – Valencian Graduate School and Research Network for Artificial Intelligence & Generalitat Valenciana and Universitat Politècnica de València (PAID-PD-22).

There are no material financial or non-financial interests to disclose for the authors.

REFERENCES

- [1] Mallika Marar, Qi Long, Ronac Mamtani, Vivek Narayan, Neha Vapiwala, and Ravi B Parikh. Outcomes among african american and non-hispanic white men with metastatic castration-resistant prostate cancer with first-line abiraterone. *JAMA network open*, 5(1):e2142093–e2142093, 2022.
- [2] Neel Kanwal, Saul Fuster, Farbod Khoraminia, Tahlita CM Zuiverloon, Chunming Rong, and Kjersti Engan. Quantifying the effect of color processing on blood and damaged tissue detection in whole slide images. In *2022 IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*, pages 1–5. IEEE, 2022.
- [3] Julio Silva-Rodríguez, Adrián Colomer, María A Sales, Rafael Molina, and Valery Naranjo. Going deeper through the gleason scoring scale: An automatic end-to-end system for histology prostate grading and cribriform pattern detection. *Computer Methods and Programs in Biomedicine*, 195:105637, 2020.
- [4] Zahra Tabatabaei, Adrián Colomer, Kjersti Engan, Javier Oliver, and Valery Naranjo. Residual block convolutional auto encoder in content-based medical image retrieval. In *2022 IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*, pages 1–5, 2022.
- [5] Fabio Alexandre Spanhol, Luiz S. Oliveira, Caroline Petitjean, and Laurent Heutte. Breast cancer histopathological image classification using convolutional neural networks. In *2016 International Joint Conference on Neural Networks (IJCNN)*, pages 2560–2567, 2016.
- [6] Min Liu, Yu He, Minghu Wu, and Chunyan Zeng. Breast histopathological image classification method based on autoencoder and siamese framework. *Information*, 13(3), 2022.
- [7] Arne Schmidt, Julio Silva-Rodríguez, Rafael Molina, and Valery Naranjo. Efficient cancer classification by coupling semi supervised and multiple instance learning. *IEEE Access*, 10:9763–9773, 2022.
- [8] Amartya Kalapahar, Julio Silva-Rodríguez, Adrián Colomer, Fernando López-Mir, and Valery Naranjo. Gleason grading of histology prostate images through semantic segmentation via residual u-net. In *2020 IEEE International Conference on Image Processing (ICIP)*, pages 2501–2505. IEEE, 2020.
- [9] Shekoofeh Azizi, Basil Mustafa, Fiona Ryan, Zachary Beaver, Jan Freyberg, Jonathan Deaton, Aaron Loh, Alan Karthikesalingam, Simon Kornblith, Ting Chen, et al. Big self-supervised models advance medical image classification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3478–3488, 2021.
- [10] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020.
- [11] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
- [12] Xiao Liu, Fanjin Zhang, Zhenyu Hou, Li Mian, Zhaoyu Wang, Jing Zhang, and Jie Tang. Self-supervised learning: Generative or contrastive. *IEEE Transactions on Knowledge and Data Engineering*, 2021.
- [13] Amalia Luque, Alejandro Carrasco, Alejandro Martín, and Ana de Las Heras. The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition*, 91:216–231, 2019.
- [14] Luis Cuadros-Rodríguez, Estefanía Pérez-Castaño, and Cristina Ruiz-Samblás. Quality performance metrics in multivariate classification methods for qualitative analysis. *TrAC Trends in Analytical Chemistry*, 80:612–624, 2016.
- [15] Guy Nir, Soheil Hor, Davood Karimi, Ladan Fazli, Brian F Skinnider, Peyman Tavassoli, Dmitry Turbin, Carlos F Villamil, Gang Wang, R Storey Wilson, et al. Automatic grading of prostate cancer in digitized histopathology images: Learning from multiple experts. *Medical image analysis*, 50:167–180, 2018.
- [16] Sebastian Otálora, Niccolò Marini, Henning Müller, and Manfred Atzori. Semi-weakly supervised learning for prostate cancer image classification with teacher-student deep convolutional networks. In *Interpretable and Annotation-Efficient Learning for Medical Image Computing*, pages 193–203. Springer, 2020.
- [17] Eirini Arvaniti, Kim S Fricker, Michael Moret, Niels Rupp, Thomas Hermanns, Christian Fankhauser, Norbert Wey, Peter J Wild, Jan H Rueschoff, and Manfred Claassen. Automated gleason grading of prostate cancer tissue microarrays via deep learning. *Scientific reports*, 8(1):1–11, 2018.