2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Attention Based Deep Multiple Instance Learning Approach for Lung Cancer Prediction using Histopathological Images

João Moranguinho, Tania Pereira, Bernardo Ramos, Joana Morgado, José Luis Costa, and Hélder P. Oliveira *(Member, IEEE)*

***Abstract*—Deep Neural Networks using histopathological images as an input currently embody one of the gold standards in automated lung cancer diagnostic solutions, with Deep Convolutional Neural Networks achieving the state of the art values for tissue type classification. One of the main reasons for such results is the increasing availability of voluminous amounts of data, acquired through the efforts employed by extensive projects like The Cancer Genome Atlas. Nonetheless, whole slide images remain weakly annotated, as most common pathologist annotations refer to the entirety of the image and not to individual regions of interest in the patient's tissue sample. Recent works have demonstrated Multiple Instance Learning as a successful approach in classification tasks entangled with this lack of annotation, by representing images as a bag of instances where a single label is available for the whole bag. Thus, we propose a bag/embedding-level lung tissue type classifier using Multiple Instance Learning, where the automated inspection of lung biopsy whole slide images determines the presence of cancer in a given patient. Furthermore, we use a post-model interpretability algorithm to validate our model's predictions and highlight the regions of interest for such predictions.**

***Index Terms*—Deep Learning, Multiple Instance Learning, Lung Cancer, Histology Images**

I. Introduction

Recent numbers published by the American Cancer Society show that lung cancer will be the second leading cancer diagnosed to patients in the year 2020 [1]. Estimates also point out that lung cancer will be the leading cause for cancer-related fatalities registered this year. In order to limit lung cancer impact, patients must be correctly diagnosed in early stages, allowing a timely application of proper target therapies. One of the gold standards used in the classification and characterisation of a patients status is the microscopic inspection of histological slides [2]. These slides are obtained by thinly slicing a pre-processed and thoroughly prepared paraffin-embedded sample of tissue excised from a patient's suspicious region. Pathologists look for many insights in these slides to aid them in elaborating a diagnosis. These insights can arise from cell count, cell shape, nucleus size and shape, necrosis and many other features present in the tissue slide [2], [3]. However, the visual inspection of a significant and diverse amount of these slides can be quite tiresome and time-consuming, leading to potentially incorrect or non-consensual observations among colleagues [4]. Artificial Intelligence has proven its strengths and benefits in aiding professionals overcome issues present in situations like these by automating some tasks and processes, while demonstrating an above-average accuracy and precision, and sometimes revealing insights that escape the human eye.

The automation of these processes is currently done through histopathological image tiles as an input for the training process of Convolutional Neural Networks (CNN). These tiles refer to same-sized patches acquired from the original image as its dimensions are still not contemplated by the advances in neural networks architectures. However, an issue arises concerning image region segmentation and labelling, since most studies presented in the literature make use of labelled whole slide images that do not contain highlighted tumour regions of interest [5], [6]. This means that the tiles acquired from these images may be incorrectly labelled if the whole slide image label gets attributed to them. Such publications tackle this issue by employing Multiple Instance Learning techniques [5], [6].

Multiple Instance Learning (MIL) is considered a supervised/weakly supervised Machine Learning technique where data gets regarded as an amalgamation of smaller components [7], [8]. In the associated literature, a formalisation of this concept is achieved by denoting data samples as bags, and its components are denoted instances. In classification tasks using data sets comprised of images, a bag represents an entire image, and its instances represent some or all of its patches. This approach is particularly relevant in tasks where the classes of the data samples are known, while in contrast, the classes of its composing parts are not, *i.e.* a single label is available for the entirety of the bag, and no labels get provided for its instances. Hence, reducing the need for professional pathologists increased efforts in generating numerous annotations, while also increasing robustness concerning heterogeneity.

The focal point of this work is creating an automatic solution for the classification of lung biopsies expanding on the existing work employing CNN's by tacking a MIL approach. Thus, we propose a bag/embedding-level lung tissue type classifier using a CNN in a MIL approach, where the automated inspection of lung histopathological images

B. Ramos and J. Moura are with the INESC TEC - Institute for Systems and Computer Engineering, Technology and Science, Portugal and FEUP - Faculty of Engineering, University of Porto, Portugal.

T. Pereira is with the INESC TEC, Portugal.

J. L. Costa is with the i3S - Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Portugal and IPATIMUP - Institute of Molecular Pathology and Immunology of the University of Porto, Portugal.

J. Morgado and H. P. Oliveira are with the INESC TEC, Portugal and FCUP - Faculty of Science, University of Porto, Portugal.

This work is financed by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia, within project UIDB/50014/2020.

determines the presence of cancer in a given patient, decreasing pathologists analytical burden and patient's diagnostic time. Furthermore, we employ a post-model interpretability algorithm to validate our model's predictions and highlight the regions of interest for such predictions while increasing the trust provided by pathologists by doing so.

II. Materials and Methods

A. Dataset

To fulfil the objectives set for this work, two datasets from The Cancer Genome Atlas (TCGA) research program were selected as they comprise two main features that are crucial: histopathological images and the label for lung cancer *vs* non-cancer. The data samples used in this work are lung tissue whole slide hematoxylin and eosin (H&E) stained images (see Figure 1), acquired from the TCGA-LUSC (Lung Squamous Cell Carcinoma) and TCGA-LUAD (Lung Adenocarcinoma) cohorts [9], [10]. The former encompasses 1612 samples where 1265 relate to a primary tumour and 347 relate to normal tissue. The latter gathers 1608 samples, with 1364 belonging to tumour tissue and only 244 belonging to normal tissue.

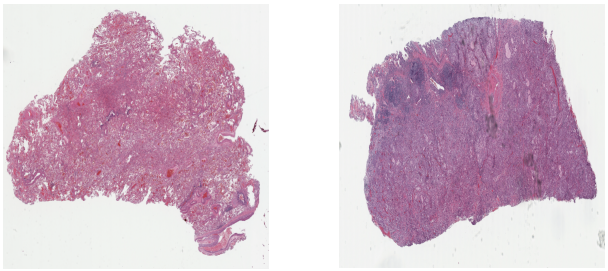

Fig. 1. Lung tissue H&E images from TCGA-LUAD dataset, classified as normal (left) and cancer (right).

B. Pre-processing

The original H&E whole slide images come in the .svs format and with infeasible dimensions to feed the classification network. Therefore, a set of pre-processing steps were required to turn these images manageable. The first step was to re-format them, and to this end, images were opened with *PIL* framework, transformed into *NumPy* arrays and saved into new PNG files.

Afterwards, the new re-formatted images get divided into several same-sized tiles acquired from a 40x magnification, whose dimensions were specified and set to 192x192. Only a few get selected from these tiles to serve as an input for our classification task. This selection was made with the aid of a scoring function introduced by Erikson *et al.* [11], where each tile gets scored according to their tissue percentage, tissue colour and saturation.

Inline with Erikson *et al.* [11], only the 50 highest scoring tiles of each image get selected and retrieved to represent instances composing the image bag. Every bag previously formed is then categorised into positive and negative classes. The positive class represents cancerous tissue, and the negative one represents normal tissue samples.

C. Attention Based Deep MIL

MIL algorithms bifurcate into patch/instance-level classification and bag/embedding-level classification tasks.

The former is concerned with the attribution of labels to the many individual instances present in each bag. It is the most prominent in literature exhibiting the current state of the art accuracy values in classification tasks analogous to the one herein discussed [5], [6]. This approach's theoretical formulation demonstrates an increased tendency to insufficiently train the model in a bag-level classification task and introduce additional error to the model as the incorrect instance-level attribution of labels will reproduce incorrect bag-level predictions [12].

The latter approach is only concerned with attributing a single label to each bag containing a set of instances, transforming our task into a fully-supervised classification one. This choice of classification level implies that *modus operandi* will always base itself in the analysis of multiple patches from each image, given that a bag-level classifier will most likely decrease its performance labelling single instances/patches, as stated by Cheplygina *et al.* [13].

In MIL, pooling layers are responsible for the attribution of a label to the whole bag. Standard pooling techniques commonly represent maximum and mean operators, where the maximum or mean value of a set of instance labels is attributed to the whole bag. However, these pooling techniques' static nature becomes a disadvantage, as stated by Ilse *et al.* [14] in a publication where two attention mechanisms are proposed to substitute the pooling layers. The disadvantage mentioned before is tackled by creating a trainable pooling layer in which the attention mechanism is responsible for the attribution of different weights to the many nodes of the last layers of the network. Figure 2 shows our work's full pipeline based on the implementation of these attention mechanisms.

D. Training

We train and evaluate the model three times for each experiment, assuring distinct training and testing samples. A general metric on the model's performance is generated by computing the mean bag-level prediction accuracy (Acc) and Area under the Curve (AUC), using the results from the three runs. The pre-processed data samples were split into the training, validation, test sets (70, 15, 15%). Furthermore, a 3-Fold Cross-Validation method was employed to do so. K-Fold Cross-Validation is a resampling technique in which the original dataset gets split into same-sized k folds/groups of samples. A new array of k datasets is generated, with each one randomly selecting different folds for training, validation and test phases. A manual grid-search was performed to fine-tune some of the hyper-parameters composing the model. Two hyperparameters were chosen, learning rate and weight decay, and their values transited between 1e-4, 5e-4 and 1e-3.

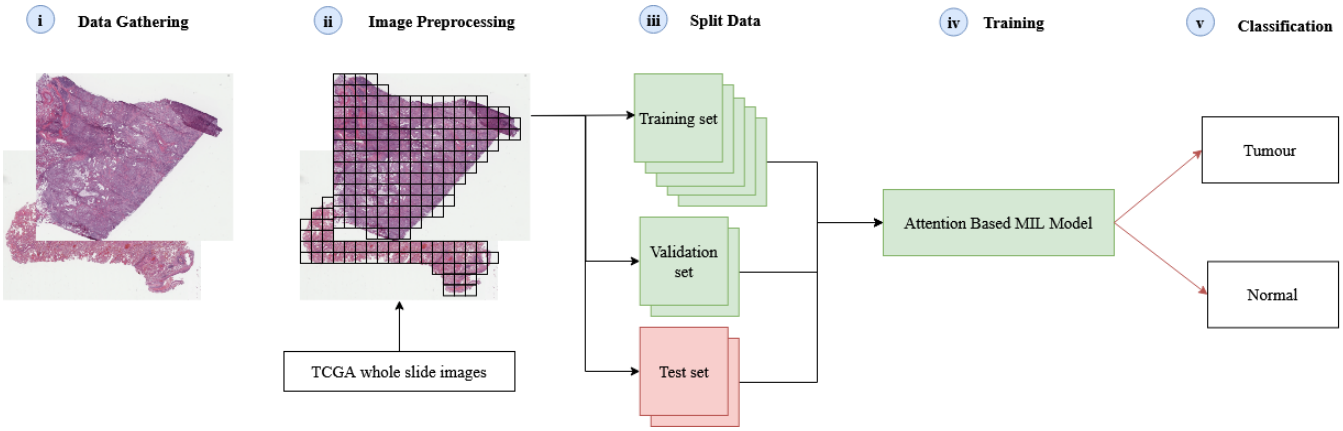

Fig. 2. Overview of the pipeline for the proposed MIL tissue type classification.

All of these experiments employ the Nadam optimizer, using the default values β_1, β_2 of 0.9 and 0.999.

III. Results

A. Performance Classification

The best hyperparameters and the correspondent performance results for the two attention methods used in this work are plotted on Table I.

TABLE I
BEST PERFORMANCE RESULTS FOR THE TWO ATTENTION MECHANISMS.

Attention Mechanism	LR	Weight Decay	Accuracy	AUC
Standard	1e-4	1e-4	0.900	0.939
Gated	1e-4	5e-4	0.912	0.945

The highest Acc and AUC values achieved belong to the experiments employing the Gated Attention mechanism and using the learning rate and weight decay values of 1e-4 and 5e-4. Moreover, experiments employing the Standard Attention mechanism also reached compelling Acc and AUC values, although using a slightly smaller weight decay factor. The slight improvement in results from the Gated Attention mechanism follows the original proposition by Ilse *et al.* [14]

B. Attention Maps

In order to discern the actual quality of our model's predictions and further aid pathologists in the visual inspection of our whole slide images, we adopted the Grad-Cam [15] algorithm. The Grad-Cam algorithm highlights regions of interest in the original images by inspecting gradient information generated through the network's layers. Moreover, the gradient values tie each neuron to its relative significance in the predictions generated and create a heatmap of "importance" superimposed over the original image.

Figure 3 displays the original image tiles fed into our fully-trained model and the superimposed images generated by the Grad-Cam algorithm highlighting the regions with a larger contribution to the model's prediction. In this particular experiment, the original image was classified as cancer tissue. This classification was correctly predicted by our model and the visual characteristics defining this classification are highlighted in the Grad-Cam algorithm output image, particularly, abnormal nuclei shape and size.

IV. Discussion

Our model displays a compelling potential in the classification of lung histology slides. Additionally, a significant contribution originates from our implementation of the Grad-Cam algorithm, increasing our model's trust by validating its predictions and highlighting specific cells exhibiting abnormal features. Despite holding lower performance metrics values than the state-of-the-art competitors, our work indicates that further improvements may arise with the increase of data samples, more meticulous image pre-processing, and an increase in the model's architecture complexity. Furthermore, a few limitations are imposed in this work line, such as the imbalance between normal and cancerous data sample count, as lung biopsies are not excised in non-suspicious patients. An additional limitation comes from the computational cost of the employed techniques. Using a voluminous amount of input images can be exhausting to the computer RAM and only with further advances in hardware improvement can this issue be minimised. A final effort to further improve our work envisions the validation of the regional highlights generated with the Grad-Cam algorithm by a professional pathologist.

V. Conclusions

This study showed that an approach based on attention-based deep MIL for lung cancer prediction in histopathological images could be a relevant tool to help the pathologists analysis. The proposed approach based on the attention method allows the identification of the most relevant regions on the images that contribute to the lung cancer prediction. When allied with the Grad-Cam algorithm's implementation, this information adds extra knowledge that helps clinicians trust the models and visually validate the highlighted regions.

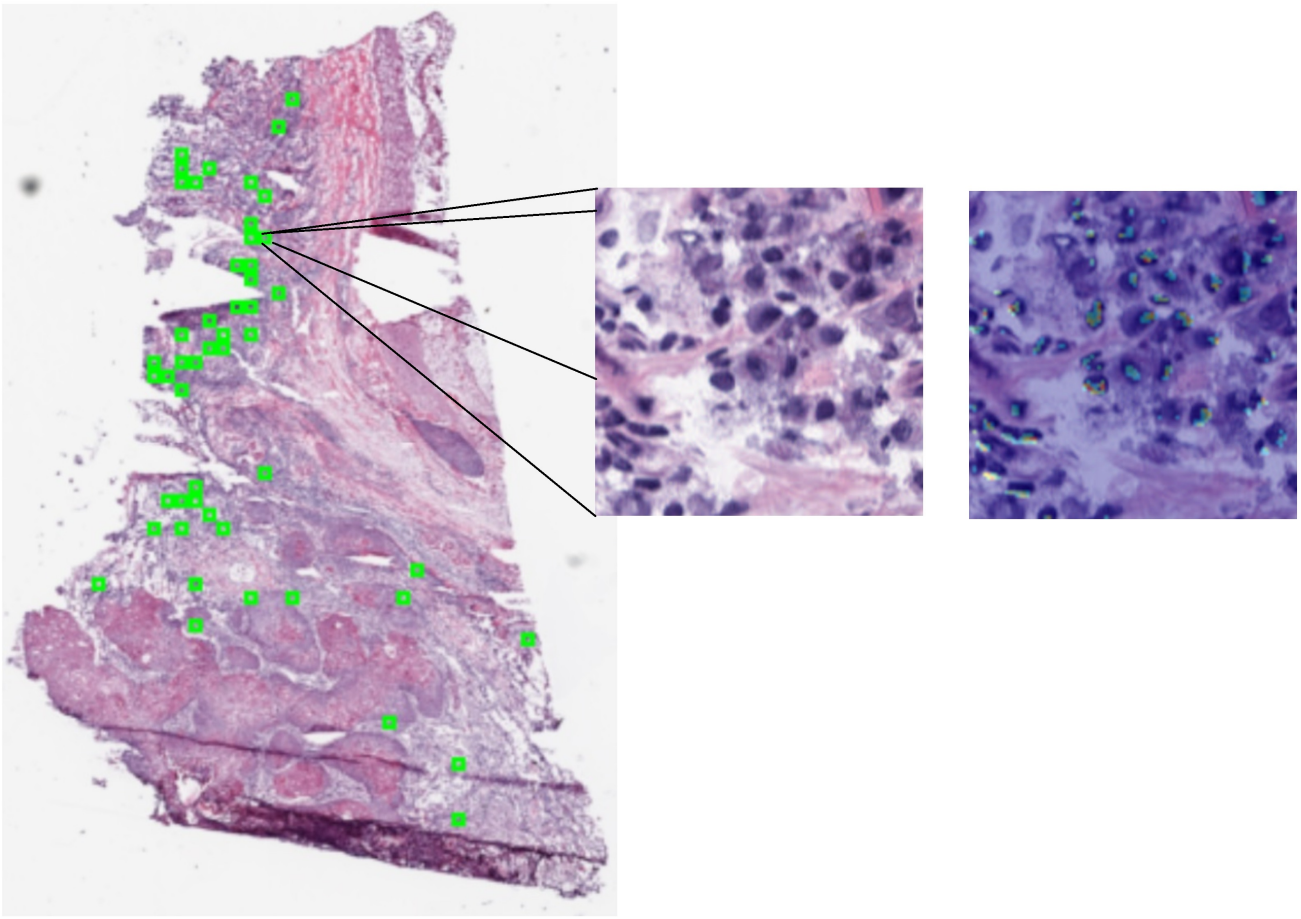

Fig. 3. Original whole slide image, tiles in green, with a tile highlighted by the GradCam algortihm.

ACKNOWLEDGMENT

We acknowledged the National Cancer Institute and the Foundation for the National Institutes of Health for the free publicly available "The Cancer Genome Atlas (TCGA)" Database used in this work. The database used in the experiments of this work ensures, on the correspondent cited description papers, that the necessary ethical approvals regarding data access were obtained.

REFERENCES

[1] N. N. C. Institute, "Cancer Facts & Figures 2020," *CA: A Cancer Journal for Clinicians*, pp. 1–76, 2020.

[2] W. D. Travis, E. Brambilla, H. K. Muller-Hermelink, and C. C. Harris, "World health organization classification of tumours," *Pathology and genetics of tumours of the lung, pleura, thymus and heart*, vol. 10, pp. 179–84, 2004.

[3] W. W. LaMorte, "Characteristics of Cancer Cells," 2016.

[4] A. Stang, H. Pohlabeln, K. M. Müller, I. Jahn, K. Giersiepen, and K.-H. Jöckel, "Diagnostic agreement in the histopathological evaluation of lung cancer tissue in a population-based case-control study," *Lung Cancer*, vol. 52, no. 1, pp. 29 – 36, 2006.

[5] L. Hou, D. Samaras, T. M. Kurc, Y. Gao, J. E. Davis, and J. H. Saltz, "Patch-based convolutional neural network for whole slide tissue image classification," in *Proceedings of the ieee conference on computer vision and pattern recognition*, 2016, pp. 2424–2433.

[6] N. Coudray, P. Ocampo, T. Sakellaropoulos, N. Narula, M. Snuderl, D. Fenyö, A. Moreira, N. Razavian, and A. Tsirigos, "Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning," *Nature Medicine*, vol. 24, 10 2018.

[7] T. G. Dietterich, R. H. Lathrop, and T. Lozano-Pérez, "Solving the multiple instance problem with axis-parallel rectangles," *Artificial Intelligence*, vol. 89, no. 1, pp. 31 – 71, 1997.

[8] M.-A. Carbonneau, V. Cheplygina, E. Granger, and G. Gagnon, "Multiple instance learning: A survey of problem characteristics and applications," *Pattern Recognition*, vol. 77, pp. 329 – 353, 2018.

[9] C. G. A. R. Network *et al.*, "Comprehensive genomic characterization of squamous cell lung cancers," *Nature*, vol. 489, no. 7417, p. 519, 2012.

[10] C. G. A. R. Network and Others, "Comprehensive molecular profiling of lung adenocarcinoma," *Nature*, vol. 511, no. 7511, pp. 543–550, 2014.

[11] D. Eriksson and F. Hu, "Whole-slide image preprocessing in Python," 2018.

[12] X. Wang, Y. Yan, P. Tang, X. Bai, and W. Liu, "Revisiting multiple instance neural networks," *Pattern Recognition*, vol. 74, pp. 15 – 24, 2018.

[13] V. Cheplygina, L. Sørensen, D. M. J. Tax, M. de Bruijne, and M. Loog, "Label stability in multiple instance learning," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 539–546.

[14] M. Ilse, J. Tomczak, and M. Welling, "Attention-based deep multiple instance learning," 02 2018.

[15] R. Rs, M. Cogswell, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," 10 2017, pp. 618–626.