END-TO-END DISCRIMINATIVE DEEP NETWORK FOR LIVER LESION CLASSIFICATION

Francisco Perdigón Romero*, Andre Diler*, Gabriel Bisson-Gregoire*, Simon Turcotte†, Real Lapointe †, Franck Vandenbroucke-Menu †, An Tang† and Samuel Kadoury*†

* MedICAL Laboratory, Polytechnique Montreal, Montréal, Canada
† Centre de recherche du CHUM (CRCHUM), Montréal, Canada

ABSTRACT

Colorectal liver metastasis is one of most aggressive liver malignancies. While the definition of lesion type based on CT images determines the diagnosis and therapeutic strategy, the discrimination between cancerous and non-cancerous lesions are critical and requires highly skilled expertise, experience and time. In the present work we introduce an end-to-end deep learning approach to assist in the discrimination between liver metastases from colorectal cancer and benign cysts in abdominal CT images of the liver. Our approach incorporates the efficient feature extraction of InceptionV3 combined with residual connections and pre-trained weights from ImageNet. The architecture also includes fully connected classification layers to generate a probabilistic output of lesion type. We use an in-house clinical biobank with 230 liver lesions originating from 63 patients. With an accuracy of 0.96 and a F1-score of 0.92, the results obtained with the proposed approach surpasses state of the art methods. Our work provides the basis for incorporating machine learning tools in specialized radiology software to assist physicians in the early detection and treatment of liver lesions.

Index Terms— liver lesion classification, metastases, CT imaging, deep learning, convolutional neural networks

1. INTRODUCTION

Colorectal liver metastasis (CLM) is one of most aggressive liver malignancies, with colorectal cancer (CC) being the second most common cancer in women and third in men with a estimate of 1.36 million cases per year based on the latest report by the WHO [1]. Due to the difficulty to detect CC in its early stages, there is a very high probability for CLM to develop in almost half of these patients. Computed tomography (CT) is the most commonly used medical imaging modality for the diagnosis and monitoring of patients with liver malignancies. The diagnosis of these malignancies through CT images is a time-consuming task for the specialist and also prone to undetected lesions. On CT images, it is possible to distinguish between liver lesions and healthy tissue because they have different contrasts [1]. On the other hand, it can be quite challenging for residents or inexperienced specialists to determine whether the lesion is benign, such as in the case of a cyst, or malignant, due to similar contrasts on CT images as illustrated in Figure 1.

In recent years there has been a growing interest in the development of computer-aided diagnostic tools that help radiologists to classify different types of lesions [2] [3] [4]. The use of deep learning techniques in medical imaging increased after the success of AlexNet [5] in the ImageNet competition [6]. Deep learning techniques are now not only used for classification, but also for segmentation of different anatomical structures using networks such as U-Net [7]. A recent survey presented by G. Litjens et al. [8] emphasizes the vast use of deep learning techniques applied to different fields in medical imaging, but also emphasized for the need to properly validate these models with respect to specific medical needs.

In this work, we propose an end-to-end trainable framework for liver lesion classification in CT images using a hierarchical deep network with fully connected classification layers to generate a probabilistic output discriminating between cysts and malignant tumors. The network was trained and validated on a clinical dataset of 63 patients screened for CLM.

Fig. 1. Sample axial slice of a CT image, with arrows indicating cysts and metastases with similar contrast levels.

2. RELATED WORK

In recent years, several methods were proposed to accurately classify liver lesions. A. Quatrehomme et al. used a trained classifier with support vector machines (SVM), in which several visual characteristics were extracted using Unser histograms, Law measures, Gaussian Markov random fields, and histogram statistics [2]. Based on the confusion matrix, the reported accuracy of cysts/malignant lesions classification yielded a rate 90.4%, which was trained on an in-house dataset that contained 95 CT lesion images. In [3], the authors performed a classification of lesions using multidimensional persistent homology, a bottleneck distance, and an SVM classifier. The dataset used for training and validation contained 132 CT lesion images and...
yielded a classification accuracy of 85%. Recently, K. Yasaka et al. [4] used a deep learning classifier of liver lesions in CT images that was trained with 1068 lesion images. The authors tested different models based on convolutional neural networks, yielding an overall accuracy of 84%. The main limitations with these methods is the lack of a discriminant framework when training the models in order to distinguish cysts from metastases.

3. METHODS AND MATERIALS

3.1. Dataset

In this study, a dataset consisting of CT images from 63 consenting patients treated at the Centre hospitalier de l’Universit´e de Montr´eal was used, identified from a prospectively maintained hepatobiliary biobank. The dataset contains 230 lesions with an equal distribution of cysts and metastases. The largest cyst lesion was 13.8mL, while the largest metastasis was 534.6mL. On the other hand, the smallest volumes was 0.018mL and 0.272mL, respectively. The average volumes of metastases and cysts was 24.871 mL and 0.791 mL, respectively. The lesions were automatically segmented using a joint liver/liver-lesion segmentation model build from two fully convolutional networks, connected in tandem and trained together end-to-end [9], which yielded a Dice score over 90% from the 2017 MICCAI LITS challenge. The previous segmentation helps to minimize the size of the image, thus decreasing the computational cost, and also helps the convolutional network from isolating several lesions contained in the same axial slice. For metastases, the primary cancer are variants of malignant neoplasm of rectum or colon.

3.2. Proposed models

In this work, we implemented an end-to-end trainable deep network for liver lesion classification. At it’s core, we leverage the feature extraction capabilities of two well-known architectures: Inception-V3 [10] and InceptionResNet-V2 [11]. Two fully connected layers, each with a Dropout of 0.4 and ReLU activations, were used for the classification. Finally, a fully connected layer of size 2 was added to the architecture, due to the number of classes in the classification problem, followed by a Softmax layer (see Table 1).

| Table 1. Details of the framework architecture. |
|-----------------------------------------------|
| Block Type | Block Name | Feature Name | Number |
| Extraction Feature | Inception-V3 | InceptionResNet-V2 | 2048 |
| Extraction Global Average Pooling | - |
| Classification Fully connected | 512 |
| Classification Dropout(0.4) | - |
| Classification Fully connected | 512 |
| Classification Dropout(0.4) | - |
| Classification Fully connected | 2 |

The inputs to the network were composed of 3 orthogonal patches, where the first image belongs to the plane where the lesion is larger, while the other two images belong to the adjacent orthogonal slices. This ensures that the network is provided enough information for optimal classification accuracy (see Figure 2).

The proposed model is based on the Inception-V3 [10] architecture, following improvements from the original model [12]. The architecture addresses the common problem of high variations in the spatial location of targets. In medical images, the same object can be observed at different scales. Choosing the appropriate kernel size for convolutions is challenging as tumors and cysts can greatly vary in size (Figure 1). Moreover, the computational cost increases with the kernel size. To solve this problem, a concatenation of multiple feature extraction methods were used. Extraction with different kernel sizes and pooling steps are processed in parallel, and all extracted features are concatenated at the end of the Inception module. The best feature extraction filters are chosen through back-propagation.

The architecture implements the concept of factorization by asymmetric convolutions, which is a simple but efficient way to optimize matrix operations. This is done by reducing nxn convolutions into nx1 followed by 1xn convolutions, improving compared to the previous factorization of convolutions and reduced computational cost. For example, a pair of 3x3 convolutions can be replaced by one 5x5 convolution (reduction factor of 2.78). Although, a representational bottleneck is created, this does not affect performance given a properly balanced size, based on the hypothesis of a correlation presence between adjacent pixels in the image. This correlation makes the reduction of the image dimension (bottleneck) possible without losing relevant information. To counterbalance the effect of the representational bottlenecks, the filter banks were expanded. Hence, the module was not made deeper but wider, as the depth reduces the dimensions (and the information) and the width preserves it. Instead of putting 3x3 convolutions followed by 1x1 and 1x3 convolutions serially, the output of the 3x3 convolution is fed in parallel to the 1x3 and 3x1 convolutions.

Auxiliary classifiers (AC) were added to improve convergence at the end of the model. They reduce the vanishing gradient problem, by acting as regularizers. AC work well when combined with other regulators like batch normalization and dropout layers. Finally, additional techniques such as batch normalization and label smoothing were added to improve the computationally efficiency of the final architecture without affecting its performance.

In our model, we leverage the use of residual connections present in the InceptionResNet [11] architecture, as they generally improve classification accuracy. The original InceptionResNet combined the Inception modules with residual connections. The so-called residual connections add features that were the input (officially called Identity connection or Identity mapping) of two or three previous convolutional layers. This approach introduced by He in al. [13] demonstrated that the residual connections reduce the time of convergence for a classification network (ILSVRC 2015)
as it minimizes the effect of vanishing gradients, especially in very deep networks.

Given the large size of the Inception-V3 architecture, it was expected that the natural evolution of the architecture would be the use of residual blocks. This resulted in reduced convergence time and slightly better results when compared to Inception-V3. Two architectures were proposed: InceptionResNet-V1, which has a computational cost similar to Inception-V3, and InceptionResNet-V2, a more computationally expensive version but with a much higher performance.

### 3.3. Training Protocol

In this section, we present the training protocol for the models, along with parameter selection strategies.

**Cropping and padding**: All volumes were cropped to the size of the bounding boxes generated by the segmentations during data preparation. To have the same size for all images, a padding was applied to fit the largest lesion, with a size of $252 \times 210$ pixels. The value of the padding was the mean value of the extracted bounding box. Using this technique, the mean value of the bounding box image is less affected compared to zero padding.

**Normalization**: Mean centering and standard deviation normalization was applied to each image patch.

**Train/Validation/Test split**: The main dataset containing 230 lesion images was divided into three sets: 138 lesions images was used as a training set, 46 images as a validation set, and 46 images as a test set. The images were randomly selected for each separate set. In our experiments, we tested the automatic initialization of the weights and using pre-trained weights in the ImageNet database.

**Data augmentation**: Due to the dataset’s limited size, data augmentation was used to avoid over-fitting. The data augmentation strategy included rotation (up to 30 degrees), horizontal and vertical shift (up to 25 pixels), and flipping. The previous parameter values were randomly chosen at every epoch, this was done automatically by the deep learning used framework (Keras/Tensorflow).

**Batch size**: A small batch size of 8 images was used to avoid CPU bottlenecks during training since image batches are generated in real-time on CPU due to the data augmentation strategies.

**Optimization and Learning Rate Scheduling**: The Adam optimization algorithm was used with an initial learning rate of $10^{-3}$. This learning rate was reduced by a factor of 2 after 10 epochs without improvements in the validation set accuracy, repeatedly. The allowed minimum learning rate was $10^{-10}$.

**Iterations**: The initial number of epochs was $10^3$, but due to the early stopping criterion that was used, the final number of epochs was 55. For early stopping, the validation set accuracy was monitored, after 50 epoch without improvements the train is ended.

Training time on an NVIDIA Titan Xp GPU was 170 minutes, while the inference time per single lesion was 3 ms (89 ms on an I7 7th generation generation CPU). The above timings were calculated for the proposed model, including InceptionNetV3.

### 4. RESULTS AND DISCUSSION

Table 2 presents the classification results in comparison to a texture-based approach trained with SVMs, InceptionResNetV2 and Inception-V3. It shows the proposed model with pre-trained weights from the ImageNet performed better in 4 out of the 7 metrics. Despite the fact that the pre-trained weights are obtained from natural images from the ImageNet dataset, using these weights as initialization for the network produced better results due to the important variability between the existing classes in ImageNet ($>1000$). Furthermore, due the large number of images provided to the first layers, this allows the model to train the primary filters with highly discriminant features, allowing to distinguish cysts from metastases. This feature makes a model that uses these weights to achieve better accuracy than a model trained from scratch.

Figure 3 shows the efficiency of the classifiers implemented through the receiver operating characteristic (ROC) curves. Figure 4 shows the performance of the model with the proposed model based on InceptionV3 using pre-trained ImageNet weights through the confusion matrix.

![ROC curve](image1)

![Confusion matrix](image2)

Once lesions are classified with the proposed framework, the different segmented lesions in the CT volume are color-coded, thus allowing specialists to quickly validate the lesion type (see Figure 5). Due to the limited number of openly available liver lesion datasets with confirmed presence of malignant tumors, direct comparison with other published results is difficult. However the comparative results reported in this study show the improvement with regards to
Table 2. Classification performance of the trained models. IPW: with ImageNet pretrained weights NPW: no pretrained weights.

| Model                         | Accuracy | Balanced Accuracy | F1 score | AUC     | Precision | Recall | Specificity |
|-------------------------------|----------|-------------------|----------|---------|-----------|--------|-------------|
| SVM with texture features     | 0.73     | 0.71              | 0.63     | 0.78    | 0.70      | 0.81   | 0.65        |
| InceptionResNet-V2 (NPW)      | 0.87     | 0.85              | 0.78     | 0.92    | 0.90      | 0.9    | 0.78        |
| InceptionResNet-V2 (IPW)      | 0.89     | 0.90              | 0.83     | 0.97    | 0.96      | 0.87   | 0.92        |
| Inception-V3 (NPW)            | 0.86     | 0.91              | 0.82     | 0.97    | 1.00      | 0.81   | 1.00        |
| Proposed model (IPW)          | 0.96     | 0.93              | 0.92     | 0.97    | 1.00      | 0.94   | 0.85        |

state of the art methods, including texture-based classifiers.

![Fig. 5. Sample classification result; green: cyst, red: metastasis.](image)

5. CONCLUSION

In this work, we proposed a framework to classify cysts and metastases in CT images using deep neural networks from automated segmentations. A convolutional neural network derived from Inception-V3 was developed, which was trained from scratch using pre-trained weights in the ImageNet dataset and includes fully connected classification layers generating probabilistic outputs. The proposed model yielded the highest accuracy and AUC compared to previous state-of-the-art approaches. The framework can therefore provide a screening tool for early detection of malignant lesions which can be integrated in a CAD software.

6. REFERENCES

[1] W.B. Stewart and C.P. Wild, “World cancer report 2014. lyon, france: International agency for research on cancer,” World Health Organization, p. 630, 2014.

[2] A. Quatrehomme, I. Millet, D. Hoa, G. Subsol, and W. Puech, “Assessing the classification of liver focal lesions by using multi-phase computer tomography scans,” in MICCAI International Workshop on Medical Content-Based Retrieval for Clinical Decision Support. Springer, 2012, pp. 80–91.

[3] A. Adcock, D. Rubin, and G. Carlsson, “Classification of hepatic lesions using the matching metric,” Computer vision and image understanding, vol. 121, pp. 36–42, 2014.

[4] K. Yasaka, H. Akai, O. Abe, and S. Kiryu, “Deep learning with convolutional neural network for differentiation of liver masses at dynamic contrast-enhanced ct: a preliminary study,” Radiology, vol. 286, no. 3, pp. 887–896, 2017.

[5] A. Krizhevsky, I. Sutskever, and E.G. Hinton, “Imagenet classification with deep convolutional neural networks,” in Advances in neural information processing systems, 2012, pp. 1097–1105.

[6] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. Ieee, 2009, pp. 248–255.

[7] 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.

[8] G. Litjens, T. Kooi, A.B. Bejnordi, A.A. Setio, F. Ciompi, M. Ghafoorian, A.J. van der Laak, B. Van Ginneken, and I.C. Sanchez, “A survey on deep learning in medical image analysis,” Medical image analysis, vol. 42, pp. 60–88, 2017.

[9] E. Vorontsov, A. Tang, C. Pal, and S. Kadoury, “Liver lesion segmentation informed by joint liver segmentation,” in Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on. IEEE, 2018, pp. 1332–1335.

[10] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2818–2826.

[11] C. Szegedy, S. Ioffe, V. Vanhoucke, and A.A. Alemi, “Inception-v4, inception-resnet and the impact of residual connections on learning,” in AAAI, 2017, vol. 4, p. 12.

[12] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1–9.

[13] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[14] P.D. Kingma and J. Ba, “Adam: A method for stochastic optimization,” arXiv preprint arXiv:1412.6980, 2014.