3D Supervised Contrastive-Learning Network for Classification of Ovarian Neoplasms

Tarun Roy TARUNKANTI-ROY@UIOWA.EDU
Jesus Gonzalez Bosquet JSUS-GONZALEZBOSQUET@UIOWA.EDU
Suely Oliveira SUELY-OLIVEIRA@UIOWA.EDU
Xiaodong Wu XIAODONG-WU@UIOWA.EDU
University of Iowa
Iowa City, IA 52242, USA

Editors: Under Review for MIDL 2023

Abstract

Ovarian cancer is the deadliest of all female reproductive system cancers and ranks the 5^{th} in cancer deaths among women. We propose a 3D contrastive learning based predictive model to discriminate benign from malignant masses in abdominal CT scans for ovarian cancer patients. We used fully supervised contrastive learning(SCL) approach which allowed us to effectively leverage the label information of our small dataset of 331 patients. All patients' data was collected at the University of Iowa. Three different architectures (VGG, ResNet and DenseNet) were implemented for feature extraction by contrastive learning. We showed that SCL consistently out-performed over the traditional cross-entropy based networks with VGG and two ResNet variants. With five fold cross validation, our best contrastive learning model achieves an accuracy of 92.8%, mean AUC of 92.4%, mean recall of 94.45% and mean specificity of 90.37%. This work shows that contrastive learning is a promising deep learning method to improve early detection of women at risk of harboring ovarian cancer.

Keywords: Supervised contrastive learning, ovarian cancer, classification, deep learning, feature encoder, efficentnet, resnet, cross validation

1. Introduction

American cancer society indicated, the pobability of a woman getting ovarian cancer is $\frac{1}{78}$. Moreover, the chance of dying from it is $\frac{1}{108}$. Diagnostic models for cancer patients may improve decision making to personalize management of cancer patients. In this study, we propose a deep learning-based predictive model for ovarian cancer patients to discriminate benign from malignant masses in abdominal CT scans. Our developed model uses 3D CT scan data obtained at the University of Iowa. A major challenge in the analysis of ovarian CT scans is that there are a large number of ovarian cysts existing in both malignant and benign patient data. Manually tracing all of them is cumbersome. Previous works also show that CNN based models out perform experienced field radiologists in terms of accuracy of prognosis (Saida et al., 2022). Most of the prior works done with ovarian cancer data only use 2D convolutional networks. In this study we trained 3D CNN models and got better performance compared to 2D models. We also implemented a new state-of-the-art contrastive-learning technique in 3D.

2. Methodology

In our proposed approach, we trained a 3D convolutional feature encoder using a supervised contrastive loss. The trained encoder was used on top of a multi-layer perceptron(MLP) netwrok to train the classifier. All the weights of the encoders were frozen during classifier training. The feature encoders we used had different convolutional architectures. The dataset contains CT scans of lower abdomens from 331 patients. Out of these samples, 196 scans contained malignant tumors and the rest of the 135 samples had benign tumors. Because of the small sample size, we trained models using five-fold stratified cross validation with a split of 264 for training and 67 for testing. For each volume image, the region of interest (ROI) with a dimension of $128 \times 128 \times 64$ was set around patients' lower abdomens where the ovaries were located, and the images were cropped to the ROI volume.

2.1. Representation Learning Framework

Our proposed predictive model consists of the following components, as in (Tian et al., 2019; Khosla et al., 2020)

- Data Augmentation module: 3D medical images are not suitable for any random augmentations. We experimented only with three different augmentations: translation, rotation and flipping (Solovyev et al., 2022). From each input sample n two random augmented images $\tilde{n} = Augment(n)$ were generated to train the encoder network with the objective of minimizing the contrastive loss for the same class and maximizing for the other classes.

- Encoder Network: In this work we used different 3D convolutional architectures as encoder networks that output the vector representation of the input CT volume. $x = Enc(\tilde{n}) \in \mathbf{R}^{D_E}$ In our experiments we empirically chose the representation vector size $D_E = 2048$.

- Projection Network: Maps the representation vector x to a projection vector $z = proj(x) \in \mathbf{R}^{D_p}$. In this paper we used MLP network as the projection head with output vector size of $D_p = 512$. The normalized output vector is used to measure the sample distances in the projection space. Even though we had different encoder networks, we used the same projection head in each case.

- Supervised Contrastive Losses used in this work can leverage the label information more effectively compared to the cross-entropy loss. The idea here is to cluster the points belonging to the same class that are pulled together in embedding space while simultaneously pushing apart cluster of samples from different classes (Khosla et al., 2020)

$$\mathcal{L}^{sup} = \sum_{i=1}^{2N} \mathcal{L}_i^{sup}$$

$$\mathcal{L}_i^{sup} = \frac{-1}{2N_{\tilde{\boldsymbol{y}}_i} - 1} \sum_{j=1}^{2N} \mathbf{1}_{i \neq j} \cdot \mathbf{1}_{\tilde{\boldsymbol{y}}_i = \tilde{\boldsymbol{y}}_j} \cdot \log \frac{\exp\left(\boldsymbol{z}_i \cdot \boldsymbol{z}_j / \tau\right)}{\sum_{k=1}^{2N} \mathbf{1}_{i \neq k} \cdot \exp\left(\boldsymbol{z}_i \cdot \boldsymbol{z}_k / \tau\right)}$$

For a minibatch of $X_{1..b}$ samples, here $N_{\tilde{\boldsymbol{y}}_i}$ is the total number of images in the minibatch that have the same label, y, as the anchor image, i. Augmented images are indicated by $\tilde{y}$. This loss has important properties well suited for supervised learning: (a)generalization to an arbitrary number of positives, (b) contrastive power increases with more negatives.

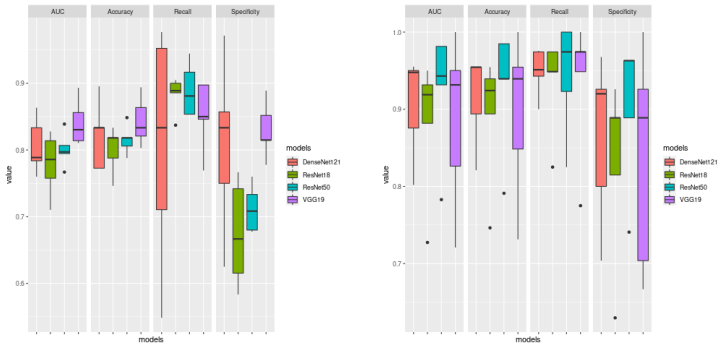

Figure 1: Performance overview of the five fold cross validation (a) Networks trained with Cross-entropy loss (b) Networks trained with Contrastive loss

3. Result and Discussion

All the models shown in Table 1 are cross validated with leave-one-fold-out fashion. This demonstrates the robustness of the models to new data. Fig. 1 depicts the performance boxplot of 5-fold cross validation in terms of accuracy, AUC, recall and specificity scores. Supervised contrastive learning models outperformed the baseline models trained with binary cross-entropy loss.

Table 1: Performance Comparison of models on CT volume size of ($64 \times 128 \times 128$)

Panel A: BaseLine 3D models				
	Acc. (%)	**AUC(%)**	**Recall(%)**	**Spec.(%)**
VGG19	84.3	84.1	85.2	82.96
ResNet18	80.1	77.9	88.33	67.5
ResNet50	81.6	80.1	88.99	71.18
DenseNet121	82.15	80.58	80.42	80.73
Panel B: SCL 3D models				
VGG19	89.48	88.58	93.45	83.7
ResNet18	89.17	88.2	93.42	82.96
ResNet50	**92.8**	**92.4**	94.45	**90.37**
DenseNet121	91.16	90.61	**94.89**	86.35

This work leverages the state-of-the-art contrastive learning method to develop an automated diagnosis model for the classification of ovarian tumors. We studied fully supervised

contrastive learning for tackling this problem and investigated the predictive powers with respect to four common CNN baselines. We expect that with a large training dataset (even without annotations), higher accuracy will be achievable using semi-supervised contrastive learning as well.

References

Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. *CoRR*, abs/2004.11362, 2020. URL https://arxiv.org/abs/2004.11362.

Tsukasa Saida, Kensaku Mori, Sodai Hoshiai, Masafumi Sakai, Aiko Urushibara, Toshitaka Ishiguro, Manabu Minami, Toyomi Satoh, and Takahito Nakajima. Diagnosing ovarian cancer on mri: A preliminary study comparing deep learning and radiologist assessments. *Cancers*, 14(4), 2022. ISSN 2072-6694. doi: 10.3390/cancers14040987. URL https://www.mdpi.com/2072-6694/14/4/987.

Roman Solovyev, Alexandr A Kalinin, and Tatiana Gabruseva. 3d convolutional neural networks for stalled brain capillary detection. *Computers in Biology and Medicine*, 141: 105089, 2022. doi: 10.1016/j.compbiomed.2021.105089.

Yonglong Tian, Dilip Krishnan, and Phillip Isola. Contrastive multiview coding. *CoRR*, abs/1906.05849, 2019. URL http://arxiv.org/abs/1906.05849.