Improving 2D CNN based feature extraction of 3D CT scan images for ovarian cancer

Sven Bierenbroodspot
s.a.k.bierenbroodspot@student.tue.nl, ID 1334859

***Abstract*—Ovarian cancer is a considerable cause of mortality for women in the Netherlands. Diagnosis relies on radiographic imaging such as ultrasound, magnetic resonance imaging (MRI), and computed tomography (CT). This paper contains a study to develop a computer-aided diagnostic (CAD) system to classify tumors into benign or malignant using 3D CT scans. pre-trained 2D convolutional neural networks (CNNs) are used for feature extraction of medical images. Due to the difference in the number of channels between the CNNs and the medical data, different input methods are explored to enhance the performance of 2D feature extraction on 3D medical images.**

The methods considered for input include a slice-based approach, an image projection based approach containing techniques such as Maximum Intensity Projection (MIP), Average Intensity Projection (AIP), Tumor MIP, and Minimum Intensity Projection (MINIP). Finally, the projection methods are combined into variable input configurations. The performance of these methods is evaluated using a Support Vector Machine (SVM) classifier, and the Area Under the Curve (AUC) is calculated to assess the quality of the classification results.

The results show that the 2D slice-based approach has the lowest AUC. The image projection methods, referred to as 2D+, considerably improve the AUC compared to the 2D slice input. Among the projection methods, Tumor MIP performs the best. Additionally, the study explores the combination of different projection methods, referred to as 2D+ VI, which leads to a slight increase in performance compared to 2D+ with a small rise in inference time.

In conclusion, this study demonstrates the effectiveness of using pre-trained CNNs for feature extraction in classifying ovarian masses as benign or malignant. The image projection methods, particularly Tumor MIP, enhance the performance compared to the 2D slice-based approach. The findings contribute to the development of a computer-aided diagnostic system for improved ovarian cancer diagnosis using CT scan images.

I. Introduction

In the Netherlands, over 1000 women pass away every year from the effects of epithelial ovarian cancer and fallopian tube cancer. [1]. In 2021 there were 1300 new cases discovered where 1090 cases are epithelial ovarian cancer. [2] this makes up 2.4% of all cancer cases diagnosed on women in the Netherlands [3].

Current diagnosis techniques involve ultrasound imaging, as well as computed tomography (CT) or magnetic resonance imaging (MRI) [4], [5]. One of the current state of the art methods for diagnosis is O-RAD MRI and it is able to provide doctors with the information to classify these tumors into benign, borderline, or malignant with a sensitivity of 93% and a specificity of 91% [6]. The disadvantage of current diagnostic techniques are the subjectivity of the diagnosis.

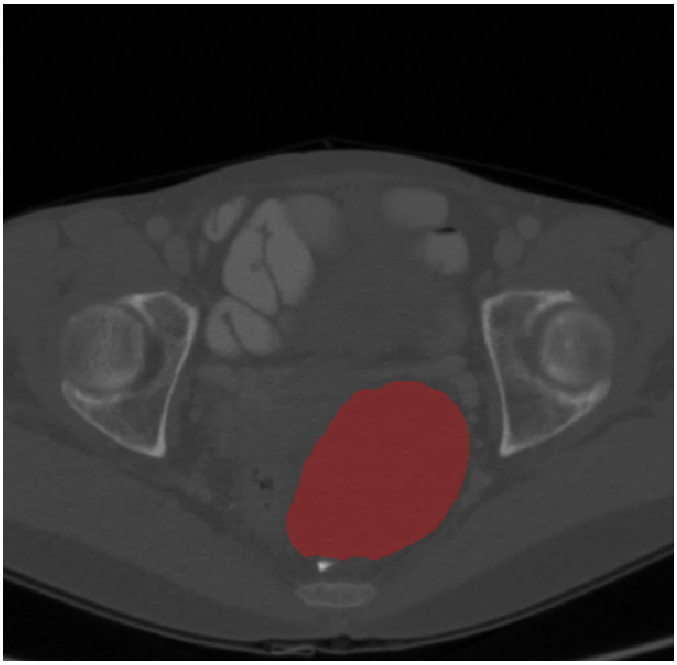

Figure 1. A 2D axial slice of the 3D-CT scan of a patient with ovarian cancer with the tumor marked in red.

Artificial intelligence has been increasingly utilized in healthcare applications over the past years. For lung nodule detection there have been great improvements in the performance, for example in the review article of Abdullah *et al.* they found that recent models have accuracies ranging from 84% to 99%. [7].

Computer aided diagnostics (CADs) can potentially aid in ovarian cancer diagnosis through their objective and potentially cost-effective character [8]. Several CADs have been developed for ovarian cancer prediction which are able to reach up to a sensitivity of 95.9% and specificity of 93.5%. [9]–[11]. However, these classifiers consist of pre-defined features and mainly use ultrasound imaging which means that these models depend on specific terminology and expertise of their users.

The aim of this paper is to develop a CAD which can classify ovarian masses as benign or malignant using CT scan images. This study uses pre-trained Convolutional Neural Networks (CNNs) to extract features from the medical images which will be used to classify the scans. As these CNNs are trained on the ImageNet dataset [12], they require an input of

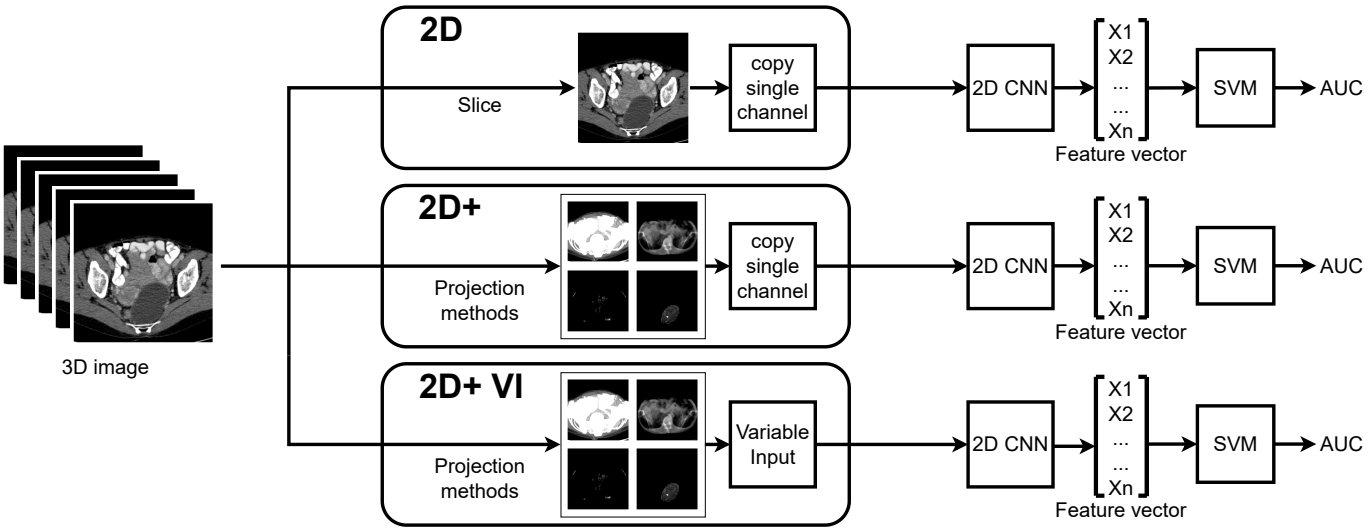

Figure 2. Illustration of pre-processing and feature extraction of different methods for 3D CT scan images. In the 2D method, the centre axial slice of the tumor is copied over the three input channels of the feature extractor. In the 2D+ method, different projection methods are used which extract information from the full 3D CT scan, these projection methods are included in Figure 3. The projections are copied over the three channels. In the 2D+ VI method, combinations of the different projection methods are used for the 3 channels.

3 channels. Since the medical data provided contains only 1 channel, it is required to expand on the other 2 channels before feature extraction. This leads to the opportunity to create inputs with variable images on different channels. Different input methods will be explored to find the best performance for 2D feature extraction on 3D medical images.

II. Methods

A. Dataset

This study uses data from the Catharina Hospital in Eindhoven, which included 146 patients. The data consists of 3D CT scans with masks of the ovarian tumors, which can be seen in Figure 1. The data has been labelled as benign, malignant and borderline using biopsy tests, where 51 cases have benign tumors, 68 have malignant tumors, and 27 have borderline tumors. For this study, only the benign and malignant cases are used due to the complexity of the classification of borderline tumors.

B. Feature extraction

CNN based feature extraction is a common technique in machine learning and computer vision used to extract meaningful features from raw data such as images. This study considers the following pre-trained 2D CNNs:

A) ResNet50 [13]
B) EfficientNet B4 [14]
C) Gluon ResNet [13]
D) ConvNext nano [15]
E) ConvNext Large [15]
F) ResNet v2 101 [16]
G) EfficientNet v2 [17]

All CNNs are trained on the ImageNet dataset [12], and have three input channels for Red, Green, and Blue (RGB) respectively. In this study, the CNNs will not be re-trained on the Ovarian Cancer dataset.

C. Slice based input for RGB channel models [2D]

To set a baseline to improve on, this method which will be called 2D contains a slice based input. It will only contain information from one axial slice hence its name. The most relevant slice is determined by the largest tumor portion. The slice is extracted from the 3D image which size is equal to $[512, 512, z]$ (where z is different among patients), thus creating a slice of size $[512, 512, 1]$. To match the input size of the 3-channel 2D CNN which is equal to $[512, 512, 3]$, the slice is copied 3 times. A visualization of an axial slice at with the largest tumor portion is included in Figure 1.

D. Image projection [2D+]

Since the 3D image contains more information than can be contained in only 1 slice the performance can possibly be improved when projecting the full 3D images onto 2D images. This study considers 4 projection methods;

1) Maximum Intensity Projection: The Maximum Intensity Projection (MIP) interprets a 3D image as a 2D image by selecting the maximum value along one axis [18]. For this study, axial MIP is used.

2) Average Intensity Projection: The Average Intensity Projection (AIP) is similar to the MIP but in stead of the maximum value, the average value along one of the axis is considered. The projection is also performed over the axial view.

3) Tumor MIP: The tumor MIP is the MIP of only the tumor. This is obtained by extracting only the tumor region of the 3D CT scan, using the 3D mask of the tumor, and then determining the MIP of this resulting image.

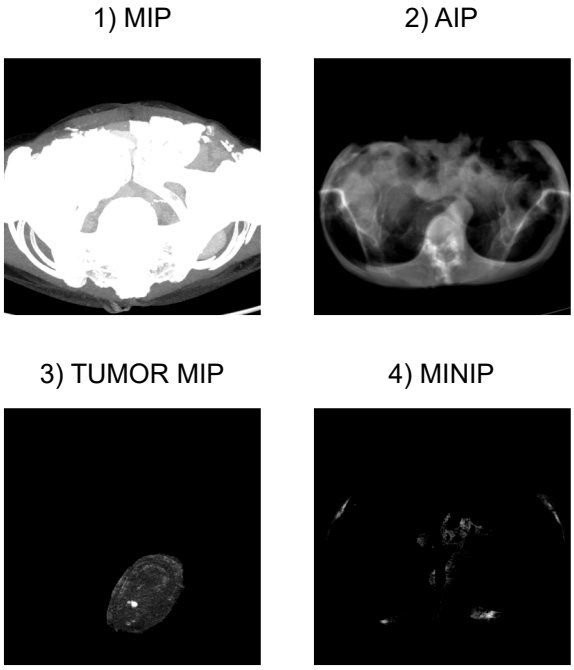

Figure 3. Illustration of the projection methods used in this study. In subfigure 1 the maximum intensity projection is displayed. In subfigure 2 the average intensity projection is displayed. In subfigure 3 the maximum intensity projection of the tumor is displayed. in subfigure 4 the minimum intensity projection is displayed.

4) Minimum Intensity projection: The Minimum Intensity Projection (MINIP) is identical to the MIP except that the maximum value is replaced with the minimal value along the relevant axis.

All image projections are included in Figure 3. From now on, this paper will refer to the projection methods by the abbreviation. To compare this method to the 2D method, the projections are copied to 3-channels and their features are extracted.

E. Variable Input [2D+ VI]

Since the 2D+ method only considers inputs of 3 identical projections this method will explore the effect of combining projection methods. Because the feature extractors will not be re-trained, the position of each projection in the input is a factor which influences the output. There are 4 different projection methods which leads to $4^3 = 64$ unique combinations. Since 4 methods are already considered in 2D+, the remaining amount of combinations is equal to 60.

F. Evaluation

In order to evaluate the quality of the feature extraction all results are evaluated identical as displayed in Figure 2. The extracted features are used to train a Support Vector Machine (SVM) [19]. This SVM is created with the scikit-learn python library. To reduce the impact of overfitting, a 5-fold cross-validation with a train/val split of 75%/25% is implemented. Since the result is not consistent for each time the SVM is trained, the SVM is trained 100 times and the best result is considered. This yields similar results to a well optimized SVM in MATLAB classificationLearner [20].

To determine whether the result is good the Area Under the Curve (AUC) is computed. To calculate the AUC, the ROC curve is constructed by varying the classification threshold, and for each threshold, the True Positive Rate (TPR) and False Positive Rate (FPR) are calculated. The TPR is the ratio of correctly predicted positive samples to the total number of positive samples, while the FPR is the ratio of incorrectly predicted negative samples to the total number of negative samples [21].

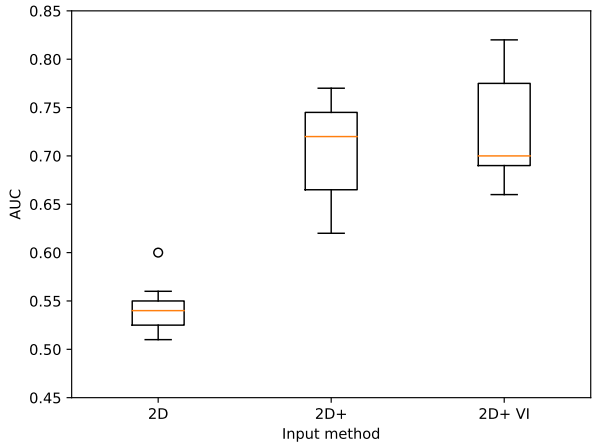

Figure 4. Classification AUC score over 5-fold cross validation for the different input methods. The plot is made based on AUC of each individual feature extractor discussed in subsection II-B.

Table I
10-RUN AVERAGE INFERENCE TIME OF THE FEATURE EXTRACTOR CONVNEXT LARGE.

	Inference time [s]
2D	0.4635
2D+	0.5016
2D+ VI	0.5298

G. Hardware

The hardware used for testing consists of an AMD Ryzen 5 5600X CPU, 32GB 3600mHz RAM, and a NVIDIA Geforce GTX 1660 GPU.

III. RESULTS AND DISCUSSION

In this section the methods are compared by considering their resulting (AUC) as well as the inference time.

A. 2D

In Figure 4 an AUC comparison over the different methods is shown. It can be seen that the 2D approach to the problem yields the lowest classification AUC compared to the other approaches for all feature extractors. The AUC values range from 0.51 to 0.60, which means that the model barely has

better performance than a coin flip. The inference time of this method is shown in Table I. Since this method does not require as much pre-processing compared to the other methods, it is the fastest.

B. 2D+

The classification AUC of the image projection methods is considerably higher than that of the 2D slice input. Figure 4 shows an increase in AUC for all feature extractors compared to the 2D input method. The median AUC saw an increase of 33% and the maximum AUC saw an increase of 28%. As shown in Table I, the inference time is increased by only 8%. The performance for each projection method is included in Figure 5. The TUMOR MIP projection method is the best performing method by a considerable margin. This means this projection shows more useful information compared to other methods. It is also note worthy that the MINIP performs better than the MIP, while from visual inspection the MIP is more clear than the MINIP. This means it is more important that the information is relevant rather than the amount of information in the projection.

C. 2D+ VI

Figure 4 shows a maximum performance increase of 6.5% for this method compared to the 2D+ method. The median performance sees a decrease of 2.8% compared to the 2D+ method. The increase in this method is not as considerable as the increase from 2D slice based to 2D+. However, small performance increases will make a difference when combined with other methods to improve the performance. For all of the feature extractions the highest AUC was found with two out of the three input images containing the TUMOR MIP, this further confirms the quality of this projection method. The inference time of this method increased by 5.6% compared to the 2D+ method.

There is no direct comparison to the inference time of 3D feature extraction on the Ovarian Cancer dataset, Wang *et al.* compared 3D feature extraction to 2D extraction and found the inference time to double for a 3D input size of $[64, 64, 64]$ and a 2D input size of $[64, 64]$ [22]. Since the image size of this study is much larger ($[512, 512, \sim 175]$), it can be expected that the inference time for 3D feature extraction on the ovarian cancer dataset will also be much larger.

D. Future research

It would be interesting to compare this study to a 3D CNN based feature extraction. The inference time is expected to be much larger than its 2D counterpart. The main question would be if the performance of a 3D based feature extraction will be better or if the projection methods already contain the most important information.

IV. Conclusion

This study considered 3 different input methods for 2D CNN feature extraction on 7 different feature extractors. The 2D slice based method yields the lowest maximum classification AUC of 0.60. The 2D+ input method leads to a considerable increase in performance with a small increase in inference time. The maximum classification AUC of this method is equal to 0.77. Finally, the 2D+ VI method also saw a small increase in inference time. The maximum classification AUC of this method is equal to 0.82. The TUMOR MIP projection method is the best performing projection method. Since the improved input methods lead to better performance, this methods should be considered in further research on this topic.

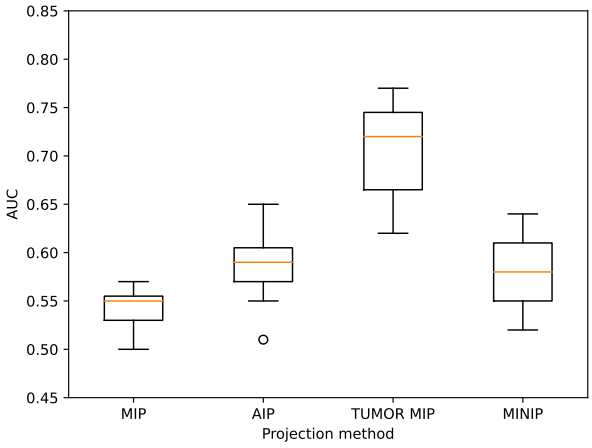

Figure 5. Classification AUC score over 5-fold cross validation for the different projection methods. he plot is made based on AUC of each individual feature extractor discussed in subsection II-B.

References

[1] IKNL. Sterfte eierstokkanker;. Available from: https://iknl.nl/kankersoorten/eierstokkanker/registratie/sterfte.

[2] IKNL. Incidentie eierstokkanker;. Available from: https://iknl.nl/kankersoorten/eierstokkanker/registratie/incidentie.

[3] IKNL. Figuren en uitleg per kankersoort;. Available from: https://iknl.nl/nkr-cijfers/figuren-en-uitleg-per-kankersoort.

[4] Orr B, Edwards RP. Diagnosis and Treatment of Ovarian Cancer. Hematology/Oncology Clinics. 2018 12;32:943-64.

[5] Jelovac D, Armstrong DK. Recent progress in the diagnosis and treatment of ovarian cancer. CA: A Cancer Journal for Clinicians. 2011 5;61:183-203.

[6] Thomassin-Naggara I, Poncelet E, Jalaguier-Coudray A, Guerra A, Fournier LS, Stojanovic S, et al. Ovarian-Adnexal Reporting Data System Magnetic Resonance Imaging (O-RADS MRI) Score for Risk Stratification of Sonographically Indeterminate Adnexal Masses. JAMA Network Open. 2020 01;3(1):e1919896-6.

[7] Abdullah DM, Ahmed NS. A Review of most Recent Lung Cancer Detection Techniques using Machine Learning IJSB Literature Review. 2021;5:159-73.

[8] Koch AH, Jeelof LS, Muntinga CLP, Gootzen TA, van de Kruis NMA, Nederend J, et al.. Analysis of computer-aided diagnostics in the preoperative diagnosis of ovarian cancer: a systematic review. Springer Science and Business Media Deutschland GmbH; 2023.

[9] Timmerman D, Verrelst H, Bourne TH, Moor BD, Collins WP, Vergote I, et al. Artificial neural network models for the preoperative discrimination between malignant and benign adnexal masses. Ultrasound in Obstetrics and Gynecology. 1999 1;13:17-25.

[10] Zimmer Y, Tepper R, Akselrod S. Computerized quantification of structures within ovarian cysts using ultrasound images. Ultrasound in Medicine Biology. 1999 2;25:189-200.

[11] Biagiotti R, Desii C, Vanzi E, Gacci G. Predicting Ovarian Malignancy: Application of Artificial Neural Networks to Transvaginal and Color Doppler Flow US. Radiology. 1999 2;210:399-403.

[12] ImageNet. About ImageNet;. Available from: https://www.image-net.org/about.php.

[13] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016. .

[14] Tan M, Le QV. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. 2019 5.

[15] Liu Z, Mao H, Wu CY, Feichtenhofer C, Darrell T, Xie S. A ConvNet for the 2020s; 2022.

[16] Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning;.

[17] Tan M, Le QV. EfficientNetV2: Smaller Models and Faster Training. 2021 4.

[18] Cody DD. AAPM/RSNA physics tutorial for residents: topics in CT. Image processing in CT. Radiographics. 2002;5:22.

[19] Cortes C. Support-Vector Networks. Machine Learning. 1995;20:273-97.

[20] MathWorks. Classification Learner;. Available from: https://nl.mathworks.com/help/stats/classificationlearner-app.html.

[21] Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology. 1983;148(3):839-43.

[22] Wang X, Su R, Xie W, Wang W, Xu Y, Mann R, et al. 2.75D: Boosting learning by representing 3D Medical imaging to 2D features for small data. Biomedical Signal Processing and Control. 2023;84:104858.