

Student Name : Onur Şero
Student ID: 2020700276

Student Name : Yunus Emre Aydın
Student ID: 2018203099

Student Name : Yunus Emre Türkoğlu
Student ID: 2019203174

Student Name : Ful Belin Korukoğlu
Student ID: 2018206162

Multimodal deep learning model for the prediction of lymph node metastasis from PET-CT images

The project's purpose was to implement a binary classification model to predict lymph node metastasis from PET-CT images of prostate cancer patients. This is important because of several reasons:

1. Treatment Planning: If lymph node metastasis is detected early, it may influence the choice of treatment options, such as surgery, radiation therapy, or systemic therapies.
2. Risk Stratification: Identifying patients who are at a higher risk of lymph node metastasis can aid in determining the aggressiveness of treatment required. Patients with lymph node involvement may require more intensive therapies to ensure effective disease management.
3. Prognosis: Lymph node metastasis is a significant prognostic factor in prostate cancer. It affects the overall survival rate and disease progression.
4. Personalized Medicine: Predictive models can contribute to personalized medicine by guiding individualized treatment decisions. By accurately identifying patients at risk of lymph node metastasis, clinicians can tailor treatment plans, avoiding overtreatment in low-risk cases and ensuring optimal therapy for high-risk cases..

We used the pet_ct dataset which included three image types (PET, CT, Seg), and tabular data that consists of clinical and histopathological information. For data preprocessing, the entropy of the categorical variables and the variance of the numerical variables were used to assess the need for feature selection in the preprocessing stage. Except for node and number columns in tabular data, all of the data had similar entropy and variance values, therefore rest was used for the model.

In numerical data, missing values were imputed using kNNs. Cropping was done on the sections of the images to speed up processing. The tabular data was one-hot encoded, and the PET and CT images were normalized. According to CT scans, pet and segmentation images were registered¹.

Model Architecture

The model takes two inputs: `x_image` (pet_ct image data) and `x_clinical` (clinical data). The image data goes through the ResNet architecture for feature extraction, while the clinical data goes through FC layers. The features from both parts are concatenated and further processed through FC layers. Finally, the output is obtained through the final FC layer and sigmoid activation function, representing the predicted probability of lymph node metastasis.

The ResNet-50² architecture consists of 50 layers, including convolutional layers, residual blocks, and pooling layers. Here's a breakdown of the ResNet-50 architecture:

Part1: ResNet-50 for feature extraction from pet_ct images

- Input: `x_image` (image data)
- Convolutional layer with 64 filters, kernel size of 7x7, stride of 2
- Batch normalization and ReLU activation
- Max pooling with kernel size of 3x3, stride of 2
- Residual blocks:
 - Stage 1: Three residual blocks with 64 filters
 - Stage 2: Four residual blocks with 128 filters
 - Stage 3: Six residual blocks with 256 filters
 - Stage 4: Three residual blocks with 512 filters
 - Each residual block consists of two convolutional layers, batch normalization, and shortcut connections
- Global average pooling to obtain a fixed-size feature vector
- Fully Connected Layer:
 - A linear layer that maps the features to the desired output size (e.g., 2048-dimensional feature vector).

Part 2: FC layers for clinical information

- Input: `x_clinical` (clinical data)
- Linear layer with 20 input features and 64 output features
- ReLU activation and dropout regularization

Part 3: FC layers for feature concatenation

- Concatenation of features from Part 1 and Part 2
- Linear layer with 2048 (from ResNet) + 64 (from Part 2) input features and 128 output features
- ReLU activation and dropout regularization

Part 4: Final FC layer for lymph node metastasis prediction

- Linear layer with 128 input features and 1 output feature
- Output is passed through a sigmoid activation function to produce the predicted probability

Results

For the evaluation metrics³, the mean and the standard deviation of 5-fold cross-validation (CV) loss, accuracy, f1-score, area under the ROC curve, sensitivity and specificity were used. Importance of each metric is as follows:

1. **Cross-Validation Loss:** Cross-validation loss provides an indication of how well the model is performing during training. It measures the discrepancy between the predicted outputs and the ground truth values across multiple folds. A lower cross-validation loss indicates better model performance in terms of minimizing prediction errors.
2. **Accuracy:** Accuracy measures the overall correctness of the model's predictions. It represents the proportion of correctly classified instances out of the total number of instances. Accuracy is a widely used metric and gives a general overview of the model's performance. However, it may not be suitable for imbalanced datasets as it can be influenced by the majority class.
3. **F1-Score:** F1-score is a measure of the model's accuracy, but it considers both precision and recall. It takes into account the trade-off between precision (ability to correctly identify positive cases) and recall (ability to identify all positive cases). F1-score provides a balanced assessment of the model's performance, especially in scenarios where class imbalance is present.
4. **Area Under the ROC Curve (AUC-ROC):** AUC-ROC is a commonly used metric for evaluating binary classification models. It represents the model's ability to distinguish between positive and negative instances by plotting the true positive rate (sensitivity) against the false positive rate (1 - specificity) at different classification thresholds. A higher AUC-ROC indicates a better discriminatory power of the model.
5. **Sensitivity:** Sensitivity, also known as the true positive rate or recall, measures the model's ability to correctly identify positive instances. It represents the proportion of true positive predictions out of all actual positive instances. Sensitivity is particularly important when the cost of false negatives (misclassifying positive cases) is high, as it ensures the identification of as many positive cases as possible.
6. **Specificity:** Specificity measures the model's ability to correctly identify negative instances. It represents the proportion of true negative predictions out of all actual negative instances. Specificity is crucial when the cost of false positives (misclassifying negative cases as positive) is high, as it ensures the avoidance of unnecessary interventions or treatments for individuals who do not have the condition.

By considering these evaluation metrics collectively, we gain a comprehensive understanding of the model's performance. Accuracy provides an overall assessment, while F1-score accounts for the balance between precision and recall. AUC-ROC provides insights into the model's discriminatory power, and sensitivity and specificity evaluate its ability to correctly identify positive and negative instances. The combination of these metrics helps in selecting the most suitable model and assessing its performance for the specific task at hand.

CNN MODEL		FUSION MODEL(clinical)	FUSION MODEL(clinical and histopathological)
Pet_seg		Pet_seg	Pet_seg
Mean of best accuracy: 0.74850640113		Mean of best accuracy: 0.79160739687	Mean of best accuracy: 0.80739687055
AUC Mean: 0.7263970737654947		AUC Mean: 0.7809591546433653	AUC Mean: 0.7943507417191628
AUC Std: 0.04521200809322972		AUC Std: 0.053605526818856816	AUC Std: 0.07710158932688323
		F1 Score Mean: 0.5857773261410849	
Ct_Seg		Ct_Seg	Ct_seg
Mean of best accuracy: 0.74822190611		F1 Score Std: 0.12857474122464774	Mean of best accuracy: 0.82332859175
AUC Mean: 0.7321479374110953		Precision Mean: 0.8039332057748522	AUC Mean: 0.8046027230237759
AUC Std: 0.0841499875761339		Precision Std: 0.13107747108226195	AUC Std: 0.03740922568728924
		Recall Mean: 0.4924009425299508	
Ct_mult_seg		Ct_seg	Ct_mult_seg
Mean of best accuracy: 0.72788051209		Recall Std: 0.17951603857808424	Mean of best accuracy: 0.80213371266
AUC Mean: 0.7156878683194472		Mean of best accuracy: 0.79715504978	AUC Mean: 0.7823816297500509
AUC Std: 0.08321667695476623		AUC Mean: 0.7826864458443404	AUC Std: 0.06340325864841052
		AUC Std: 0.05446064232762959	
		F1 Score Mean: 0.6002953640052251	
		F1 Score Std: 0.13956283545645917	
Mixed		Precision Mean: 0.7500216484069824	Mixed
Mean of best accuracy: 0.71678520625		Precision Std: 0.15550189237777887	Mean of best accuracy: 0.80199146514
AUC Mean: 0.7023166023166024		Recall Mean: 0.5116474100521633	AUC Mean: 0.7704938020727494
AUC Std: 0.057146976130813425		Recall Std: 0.1377273773340029	AUC Std: 0.08041070393101744

FigA. Summary of accuracy results of models

We created two models for analyzing the data. First, we created a 3-D CNN model which is Res-Net 50 to analyze images and secondly we created a custom model which did joint-fusion on image and tabular data. We tested the fusion model first on only clinical data and then both with clinical and histopathological data. As seen on FigA. above, after including both types of tabular data, the accuracy of the custom model increases up to 80%. It must be stated here that the mean of accuracy refers to the mean of each fold in 5 fold cross-validation. Likewise, Pet_Seg refers to the case when segmentation masks are applied to Pet images by pointwise addition. Similarly, Ct_Seg refers to the case when segmentation masks are applied to Ct images by pointwise addition. For the Ct_mult_seg case, segmentation masks are applied to Ct images with some weighted pointwise multiplication (min 0.1 and max 1.0). For the final test case called Mixed, both three images are added point-wise then fed to the network.

Lastly, except for learning rate, no other network parameters were tuned and early fusion was used to concatenate images, but joint fusion is the proper method. Our method was inspired by a multimodal learning article⁴ but in our model there is only one Res-Net50 network because Pet and Ct images are tested separately and their results are reported separately for Pet_Seg, Ct_Seg and Ct_mult_seg test cases. For the Mixed test case, all three images are merged to a single image then fed to the system. As a result there is one image input for the whole system.

REFERENCES

- 1- Z. Yaniv, B. C. Lowekamp, H. J. Johnson, R. Beare, "SimpleITK Image-Analysis Notebooks: a Collaborative Environment for Education and Reproducible Research", J Digit Imaging., [doi: 10.1007/s10278-017-0037-8](https://doi.org/10.1007/s10278-017-0037-8), 31(3): 290-303, 2018.
- 2- K. Hara, H. Kataoka, Y. Satoh; Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 6546-6555
- 3- Dsouza, J. (2021) How to improve the accuracy of your image recognition models, freeCodeCamp.org. Available at: <https://www.freecodecamp.org/news/improve-image-recognition-model-accuracy-with-these-hacks/> (Accessed: 02 June 2023).
- 4- Joo, S., Ko, E.S., Kwon, S. et al. Multimodal deep learning models for the prediction of pathologic response to neoadjuvant chemotherapy in breast cancer. Sci Rep 11, 18800 (2021). <https://doi.org/10.1038/s41598-021-98408-8>