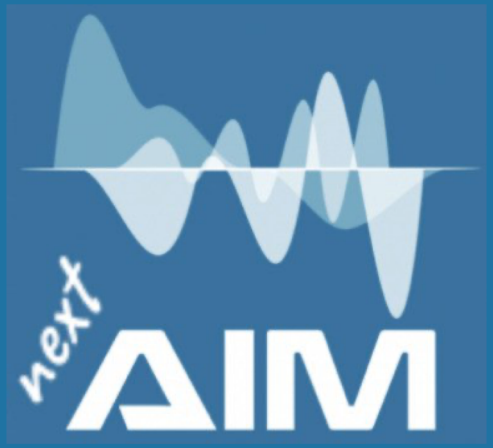
Artificial Intelligence in Medicine: next steps



Soft tissue sarcoma: transfer learning and fine-tuning for prediction of distant metastasis based on multimodal imaging

S. Piffer
INFN Sezione di Firenze

INFN

Outline

- Goal
- Clinical context
- Data
- Processing and training
- Results
- Conclusion

Goal

- Soft-tissue-sarcomas (STS) are uncommon, heterogeneous malignant tumors and their clinical management is particularly challenging. Accurate and precise STS patients stratification play an important role in clinical diagnosis and decision making for patient treatment

- Recent development on deep learning has shown great progress also in medical fields but the main limitation remains the small labeled dataset for training. To overcome this drawback transfer learning and fine-tuning have been investigated

- The goal of this study is to predict STS patients outcome to radiotherapy, in terms of distant metastasis development

Clinical context

- STS represent a rare and heterogeneous group of tumors, with more than 100 histological subtypes and account for 1% of solid cancers in adults
- Limbs are the most common primary site
- Traditional pathology approaches and molecular genetic assays play a crucial role in the classification of STS, since accurate histological diagnosis and an assessment of the risk of relapse are critical for delineating treatment strategies
- Classical pretreatment biopsy limits the classification of the entire tumor with few tissue samples
- A complete and in-depth histological analysis takes place only after surgery which is an advanced step in the therapeutic process and may differ from the preliminary grade provided on biopsy specimens
- Guidelines suggest a treatment protocol that includes pre-operative external beam radiotherapy surgery and chemotherapy

Available data

- All patients enrolled in this retrospective study received pre-operative radiation therapy for STS
- In this single-center analysis, data of 72 patients from January 2011 to August 2020 were initially analyzed for further inclusion
- Inclusion criteria:
  - availability of multi-parametric MRI, including axial T1-weighted, T2-weighted, contrast-enhanced (CE) T1-weighted maps
  - availability of radiotherapy CT, structures set, plan and 3D dose volume
  - availability of post-RT follow-up
- The study population included 61 patients with a complete data set

It can be considered a small database

Transfer learning

It is known that data-driven approaches, especially DL, have difficulty achieving high performance on limited data sets which consequently limits the stability and generalizability of the model

The use of pre-trained networks based on transfer learning and fine-tuning strategy is a valuable solution for the small dataset scenario

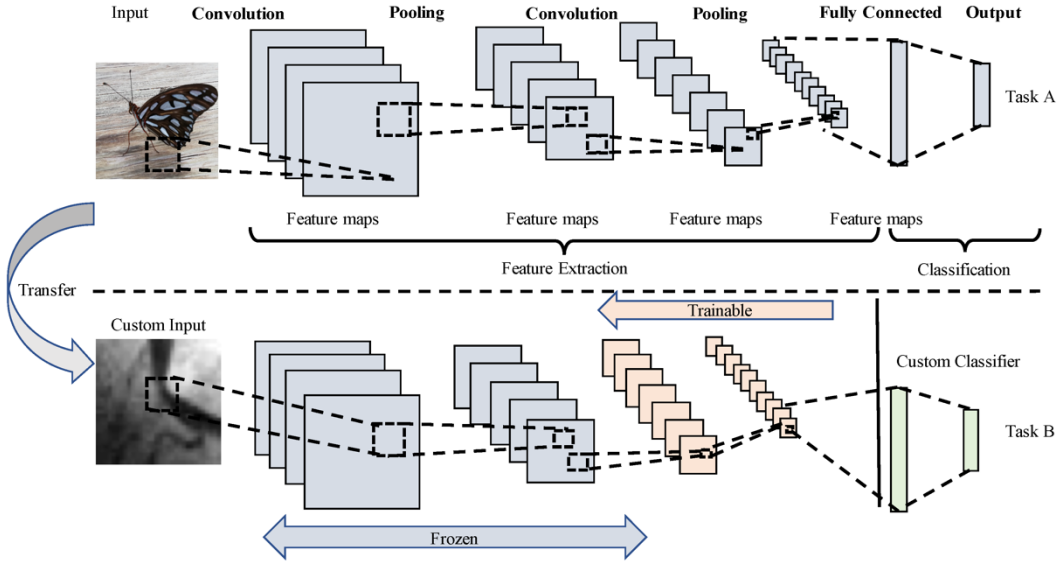


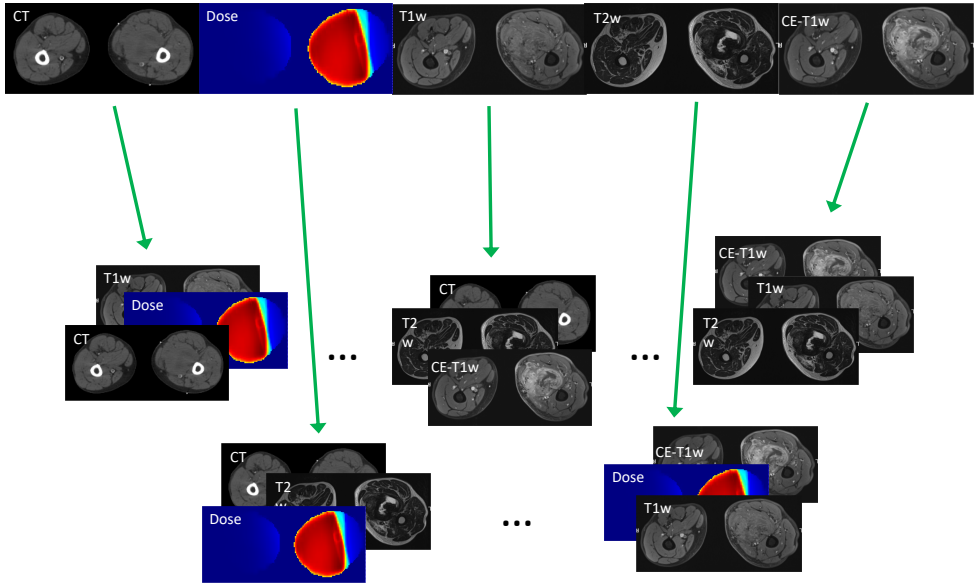
Image processing

- The proposed patient classification is based on 2D images for practical reasons and because obtaining transversal 2D slices from 3D volume for each patient allow to increase the number of training samples for the deep neural networks
- From every patient in the training set we could generate as many training samples as transversal slices are available from the patient tumor, going from several tens in the original dataset to thousands after slicing the patients (2395 images in total)
- MR images co-registered to CTs and applied deformable registration for 3D tumor contours propagation
- Since the acquisition volume of clinical images is much larger than the tumor region, we opted to extract only the tumor volume with a margin of 1 cm both superiorly and inferiorly in the cranio-caudal direction
- For MRI it is usual to have a reduced FOV compared to CT, so we resized all the multimodal images in 512 x 512 matrices, where pixel value outside the perimeter of the smaller images was set to 0
- The images were normalized employing min-max normalization to scale the image intensity values between 0 and 1
- The pre-trained network requires input images of 224 x 224 pixels with RGB 3-channels, so the images were further scaled to fit the required input size
- This double step was done to maintain the aspect ratio within the images

3-channels

- 5 different types of images available (CT, dose distribution, T1w, T2W and CE-T1W)
- With 3 channels in the input, it was possible to create 10 different combinations

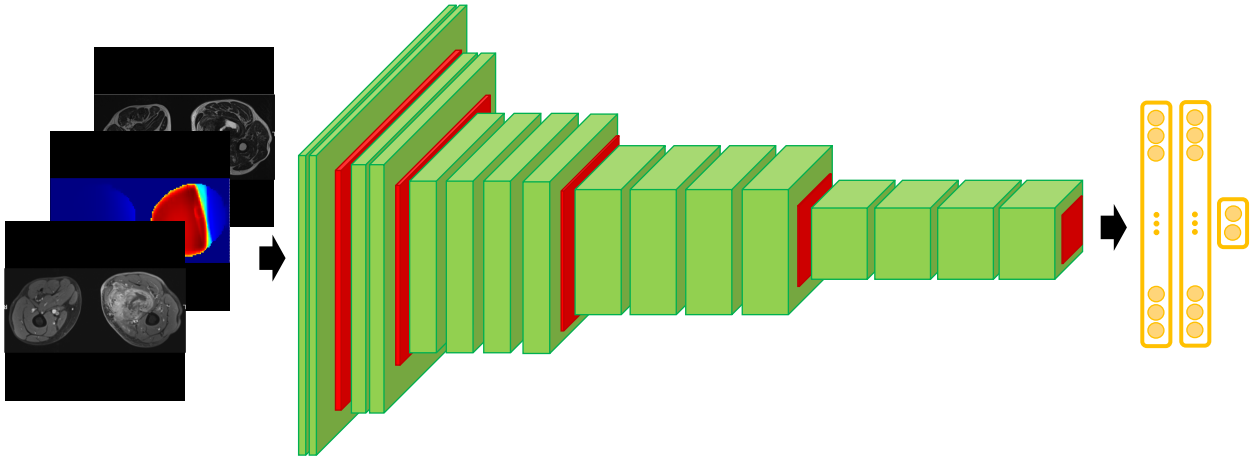
1) CT – Dose – T1w
2) CT – Dose – CE-T1w
3) CT – Dose – T2w
4) CT – T1w – CE-T1w
5) CT – T1w – T2w
6) CT – T2w – CE-T1w
7) Dose – T1w – CE-T1w
8) Dose – T1w – T2w
9) Dose – T2w – CE-T1w
10) T1w – T2w – CE-T1w



These combinations can be interpreted as 10 different sets on which it is possible to train the neural network

Transfer learning and fine-tuning

The underlying skeletal deep convolutional neural network is a VGG-19 pre-trained on ImageNet

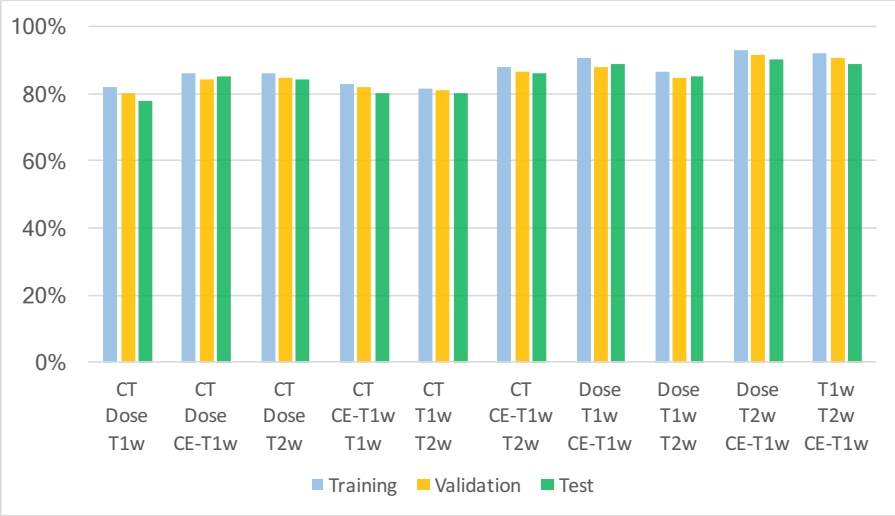


- The last fully connected layer was adapted to our binary classification task
- We applied transfer learning and fine-tuned the final block, made up of the last three fully connected layers, keeping the weights of all other layers frozen

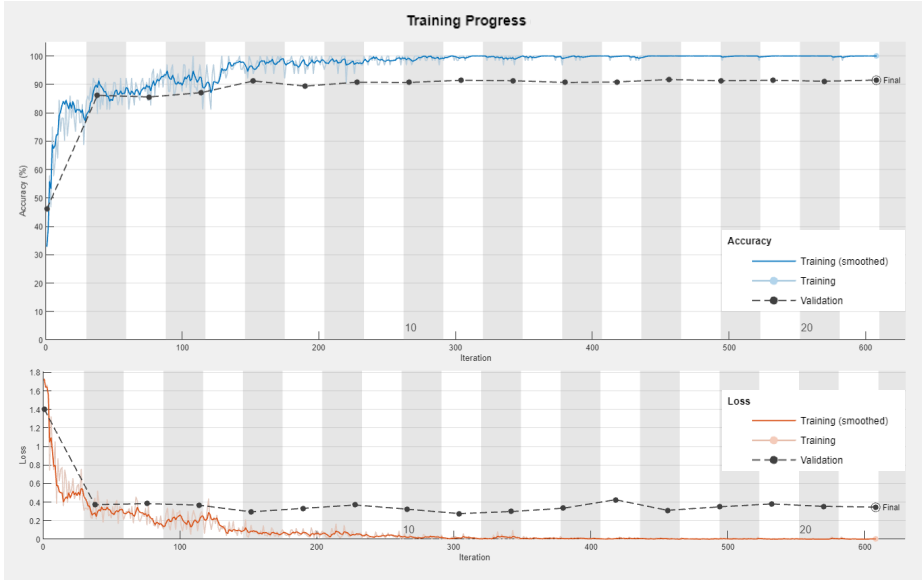
Training Deep Learning Model

- SGDM optimizer was used with momentum at 0.9, mini-batch size of 64, an initial learning rate of 0.01 for 50 epochs
- Training process was validated after every epoch and early stopping of the training was considered if the validation loss did not improve for 8 sequential epochs
- No data augmentation was applied
- For each combination set, we randomly divided the 2395 image samples into ten subsets of equal size such that each patient had been tested at least once during the independent testing stage
- We ensured no overlap and balanced classes across the ten subgroups. We exploited ten-fold cross-validation to evaluate the response prediction process. The final result is the average classification performance of the ten-fold test dataset

Results 1



The slice-based average prediction accuracy on training, validation, and test sets over the ten combination sets



The benefit of transfer learning and fine-tuning is to reduce overfitting and speed the convergence, evident from the accuracy and the loss history of our proposed model

Results 2

The first combination set had the worst performance on the test set while combination set 9 had the best prediction

Combination set	Accuracy	Sensitivity	Specificity	Precision	F1-score
1. CT – Dose – T1w	0.78 ± 0.02	0.67 ± 0.02	0.75 ± 0.02	0.73 ± 0.05	0.70 ± 0.03
2. CT – Dose – CE-T1w	0.85 ± 0.03	0.81 ± 0.03	0.82 ± 0.03	0.77 ± 0.06	0.79 ± 0.04
3. CT – Dose – T2w	0.84 ± 0.03	0.75 ± 0.04	0.83 ± 0.03	0.79 ± 0.06	0.77 ± 0.05
4. CT – T1w – CE-T1w	0.80 ± 0.02	0.68 ± 0.02	0.72 ± 0.02	0.69 ± 0.05	0.68 ± 0.04
5. CT – T1w – T2w	0.80 ± 0.02	0.71 ± 0.03	0.70 ± 0.02	0.68 ± 0.04	0.69 ± 0.03
6. CT – T2w – CE-T1w	0.86 ± 0.04	0.81 ± 0.06	0.73 ± 0.04	0.71 ± 0.06	0.76 ± 0.06
7. Dose – T1w – CE-T1w	0.89 ± 0.04	0.75 ± 0.05	0.84 ± 0.04	0.81 ± 0.07	0.78 ± 0.06
8. Dose – T1w – T2w	0.85 ± 0.02	0.76 ± 0.03	0.80 ± 0.02	0.77 ± 0.05	0.76 ± 0.04
9. Dose – T2w – CE-T1w	**0.90 ± 0.03**	**0.85 ± 0.06**	0.84 ± 0.04	0.78 ± 0.05	**0.81 ± 0.04**
10. T1w – T2w – CE-T1w	0.89 ± 0.02	0.77 ± 0.02	**0.86 ± 0.02**	**0.83 ± 0.03**	0.80 ± 0.03

Combinations in which the CT and T1w images are present give the least performing results. While for the combinations in which the other three types of multimodal images are present (Dose – T2w – CE-T1w) the results are better. A possible explanation could be based on the greater information content present in these last images

Dicussion & Limitations

Dicussion

- Despite the result is still being preliminary, it demonstrated the feasibility and the efficacy of the proposed workflow on predicting post-RT patient outcome

- The great power of the DNN to capture the different heterogeneities and microstructures present within the lesson, as well as the ability to obtain a high level of abstraction thanks to the depth of the network itself, have allowed to obtain satisfactory results

Limitations

- We have implicitly assumed that all tumor imaging slices had the same score, so the data size is sufficient to allow the use of deep learning-based prediction

- Repetitive cross-validation was applied to estimate the robustness and the results are potentially biased because we lack an external independent test set

- Although patients enrolled in this study each received the same treatment scheme, their histology subtypes are different

- The limited database and the hardware at our disposal allowed us to perform a shallow fine-tuning, in fact only the final VGG-19 block was fine-tuned; it will soon be superseded thanks to the use of INFN facilities

Future work & Conclusion

Future work

- We are enrolling more patients on an expansion cohort
- We would like to engage a deep fine-tuning (in a layer-wise manner)
- We plan to exploit Gradient Weighted Class Activation Map (Grad-CAM) as a starting point to delve into the world of explainability
- We want to highlight how the combination of multimodal images is more statistically informative than examining only one type of image at a time

Conclusion

- Pre-trained networks based on transfer learning and fine-tuning strategy possess important characteristics that make them natural candidates when applying deep learning to small medical databases
- The model demonstrates good prognostic accuracy and provides a non-invasive opportunity for personalized treatment adaptation for improved treatment efficacy

Thanks for the attention!