Early Experiences on using Triplet Networks for Histological Subtype Classification in Non-Small Cell Lung Cancer

Fatih Aksu[1] Fabrizia Gelardi[1,2] Arturo Chiti[1,2] Paolo Soda[3,4]

[1]Department of Biomedical Sciences, Humanitas University, Milan, Italy

[2]IRCCS Humanitas Research Hospital, Milan, Italy

[3]Research Unit of Computer Systems and Bioinformatics, Campus Bio-Medico University of Rome, Rome, Italy

[4]Department of Radiation Sciences, Radiation Physics, Biomedical Engineering, Umeå University, Umeå, Sweden

Outline

- Introduction
- Background
- Method
- Experiments
- Findings
- Conclusion

Introduction

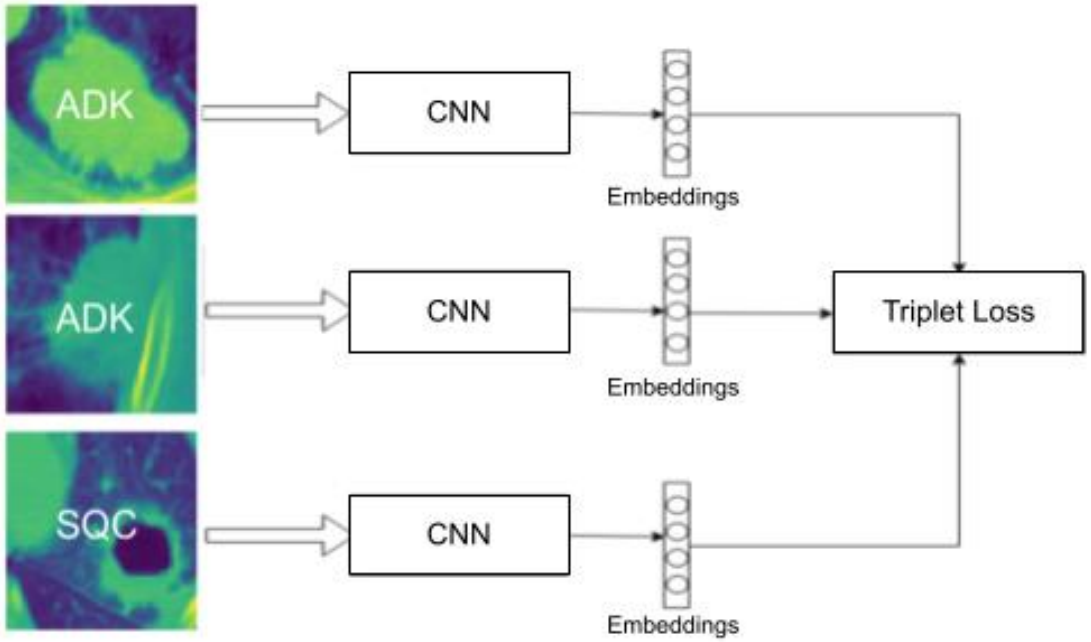
- Lung cancer is one of the leading cause of the cancer related deaths
- Accurate assessment of histotypes is essential for personalised treatment
- 85% of the lung cancer is Non-small cell lung cancer (NSCLC) with several histotypes ;
  - Adenocarcinoma (AC)
  - Squamous cell carcinoma (SQC)
  - Large cell carcinoma (LCC)
- Pathological confirmation requires invasive tissue sampling

Source: Medical News Today [1]

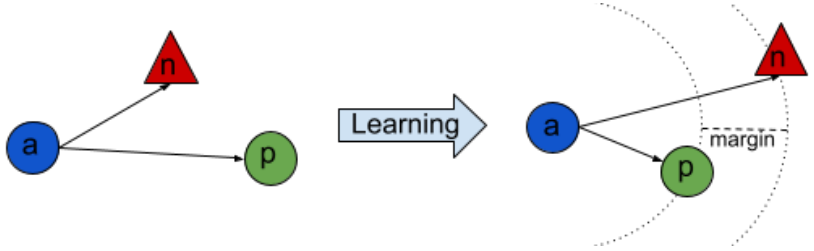
Background

- Based on hand-crafted radiomic features
  - Ferreria et al. (2018) [2]
    - 100 features are selected with ReliefF among ~2000 radiomic features
    - Classified with Naïve Bayes, k-Nearest Neighbours
  - Liu et al. (2019) [3]
    - 247 features are selected among 1029
    - Classified with an SVM
- Based on deep features
  - Han et al. (2021) [4]
    - Compared pretrained VGG-16 with 10 different ML models used with radiomic features
  - Chaunzwa et al. (2021) [5]
    - Pretrained VGG-16 used as feature extractor for kNN, SVM and Random Forest
    - Compared with end-to-end VGG-16

Triplet Loss



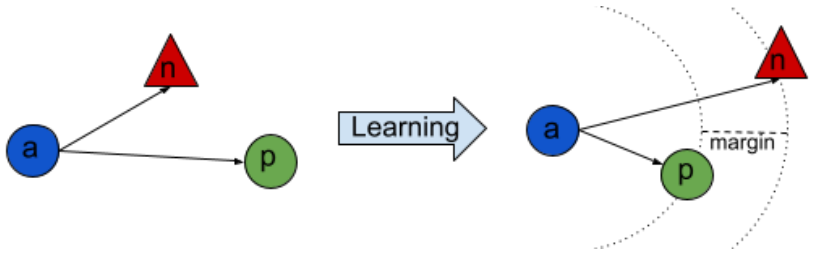
$$Loss = \sum_{i=1}^{N} \left[\| f_i^a - f_i^p \|_2^2 - \| f_i^a - f_i^n \|_2^2 + \alpha \right]_+$$



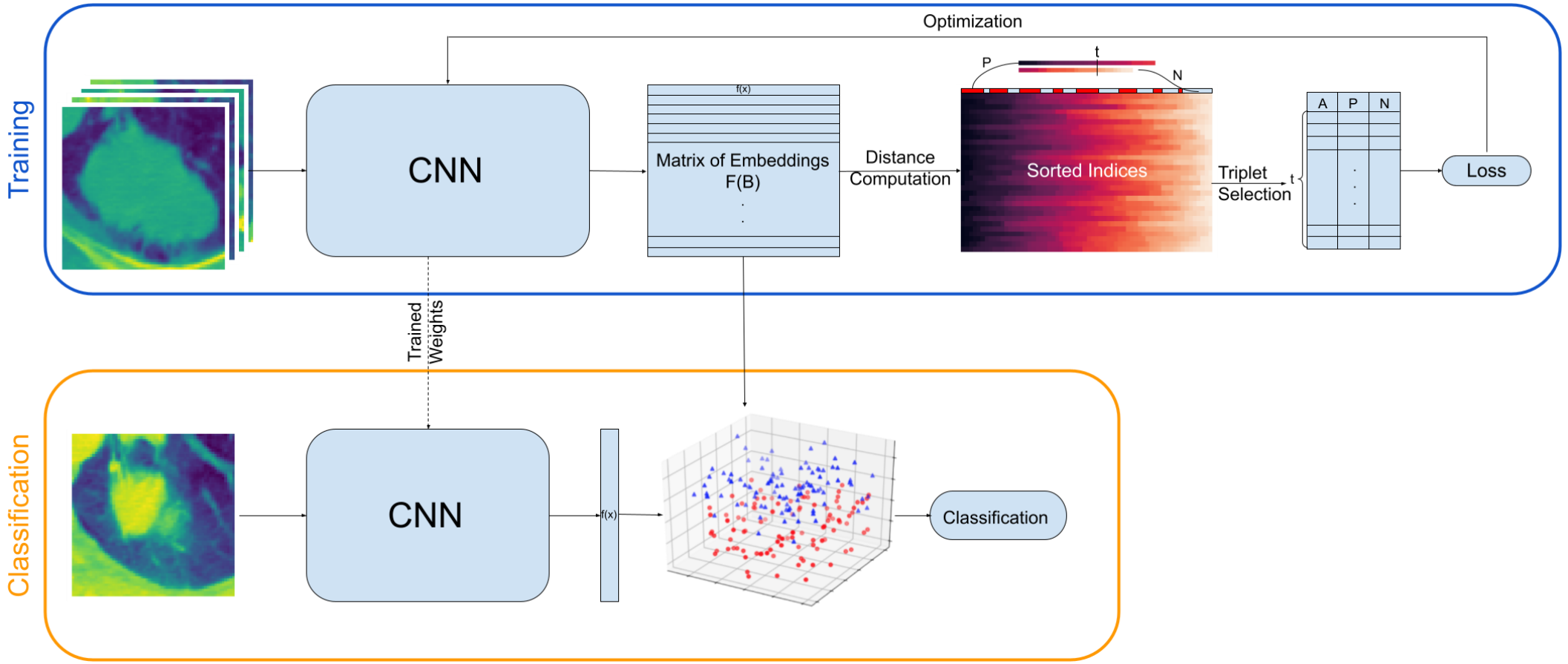
Triplet Selection Methods

- TSM-1
  - Easy positives, hard negatives
  - Static difficulty
- TSM-2
  - Easy positives, hard negatives
  - Dynamic difficulty
- TSM-3
  - Easy positives, easy negatives
  - Dynamic difficulty

$$\|f_i^a - f_i^p\|_2^2 + \alpha < \|f_i^a - f_i^n\|_2^2$$



Overall Framework



Experiments

- On a private dataset with;
  - 60 AC and 27 SQC patients
- 4 different architectures
  - ResNet-50, VGG-16, MobileNetV2, GoogleNet
- 3 different triplet selection methods
  - TSM-1, TSM-2, TSM-3
- 5 different numbers of triplets
  - 1, 2, 4, 8, 16
- 5 different numbers of neighbours
  - 1, 3, 5, 7, 9

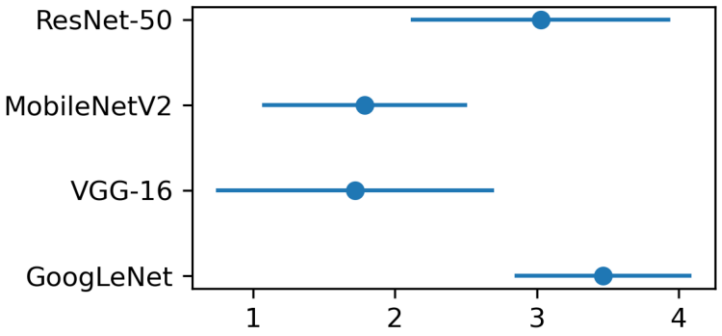
Findings

Triplet vs Softmax

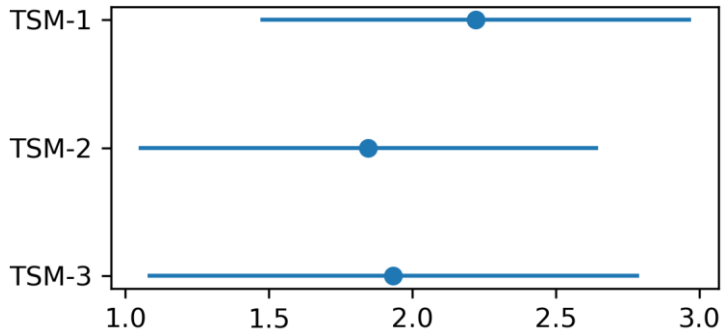
Model	Wins	Loss
ResNet-50	74 (30)	1 (0)
MobileNetV2	47 (25)	28 (0)
VGG-16	43 (12)	32 (1)
GoogLeNet	64 (18)	11 (0)

Numbers inside parantheses indicate the significant ($p<0.1$) wins/losses according to Wilcoxon signed rank test.

Model scores



Triplet selection method scores



Conclusion

- Triplet loss is a viable option to be used instead of softmax loss
- GoogLeNet and ResNet-50 are more suitable backbone networks than MobileNetV2 and VGG-16
- Choosing easiest positive and hardest negative during the whole training is a more preferred method

Future Work

- Extend our dataset
- Investigate the class unbalance problem
- Explore more architectures and triplet selection methods
- Have a multimodal approach
  - PET scans and genomic data

References

- [1] A. Sandoiu, “New protein may help to catch lung cancer early,” Medical News Today, https://www.medicalnewstoday.com/articles/321749 (accessed Jun. 17, 2023).
- [2] J. R. F. Junior et al., “Radiomics-based features for pattern recognition of lung cancer histopathology and metastases,” Computer methods and programs in biomedicine, vol. 159, pp. 23–30, 2018.
- [3] J. Liu, et al., Multi-subtype classification model for non-small cell lung cancer based on radiomics: SLS model, Medical Physics 46 (7) (2019) 3091–3100.
- [4] Y. Han, et al., Histologic subtype classification of non-small cell lung cancer using PET/CT images, European journal of nuclear medicine and molecular imaging 48 (2) (2021) 350–360.
- [5] T. L. Chaunzwa, et al., Deep learning classification of lung cancer histology using CT images, Scientific reports 11 (2021).