Hacettepe University

Artificial Intelligence Engineering Department

AIN491/492 End of Project Report

Project Details

Title	Soft Tissue Sarcoma Classification
Supervisor	Assoc. Prof. Aydın KAYA

Group Members

	Full Name	**Student ID**
1	Ahmet Emre USTA	2200765036

Abstract of the Project (/ 10 Points)

Explain the whole project shortly, including the introduction of the field, the problem statement, your proposed solution and the methods you applied, your results and their discussion, expected impact, and possible future directions. The abstract should be between 250-500 words.

Soft tissue sarcomas (STS) are rare and heterogeneous malignant tumors with over a hundred histological subtypes, posing significant diagnostic challenges due to their varied prognostic and clinical features. This project aims to develop an automated classification system for four STS subtypes: Rhabdomyosarcoma, Leiomyosarcoma, Synovial Sarcoma, and Malignant Peripheral Nerve Sheath Tumors, using deep learning models. High-resolution Hematoxylin & Eosin-stained images from the TCGA-SARC dataset and Istanbul University-Cerrahpaşa archives were used. The dataset was annotated and processed, involving masking and tiling operations to generate frames with at least 60% tumorous tissue. Pre-trained models like ResNet50V2, VGG19, InceptionV3, and EfficientNetV2S were tested for classification accuracy using cross-validation, with ResNet50V2 achieving the highest performance. Transfer learning was employed to maintain high-resolution details while optimizing inference times. The project demonstrated that deep learning models could significantly improve the diagnostic process for STS, potentially reducing diagnostic time and pathologists' workload. Future work will focus on integrating these models into clinical practice and expanding the dataset to enhance model robustness.

Keywords: Deep Learning, Artificial Intelligence in Health, Soft Tissue Sarcoma

Introduction, Problem Definition & Literature Review (/ 20 Points)

Soft tissue sarcomas (STS) are malignant tumors of mesenchymal origin, representing a highly heterogeneous group with many different subtypes. Sarcomas are rare, representing 6-15% of pediatric cancers (<15 years), 11% of adolescent and young adult cancers (15- 29 years), and 1-2% of adult cancers worldwide [(1)]. Although STSs are rare, their incidence has increased steadily over the last decade. They are also associated with significant morbidity and mortality [(2,3)]. The majority of AIDSs are sporadic. Various chemical mutagens and radiation exposure at an early age are known to underlie sporadic cases. The remaining cases are hereditary cases associated with multiple congenital syndromes, such as Li-Fraumeni syndrome.

According to the fifth edition of the World Health Organization (WHO) Classification of Soft Tissue and Bone Tumors, there are more than one hundred histological subtypes of soft tissue tumors. This diversity causes many difficulties in the differential diagnosis between histological types. In addition, the fact that each subtype has different prognostic and clinical features indicates that errors in the diagnostic process can lead to significant clinical problems. The fact that treatment approaches show remarkable differences between subtypes and that different treatment regimens are associated with different prognoses emphasizes another critical aspect of the problem [(4)]. In the current WHO guidelines, Soft Tissue Tumors are divided into 11 subcategories according to their cellular characteristics. In this project, we will study Rhabdomyosarcoma, Leiomyosarcoma, Synovial Sarcoma, and Malignant Peripheral Nerve Sheath Tumors, which are histopathologically differentiated from each other and may cause differential diagnosis difficulties for pathologists.

A biopsy is mandatory to evaluate the histologic type, subtype, and grade, and it is recommended for all deep-seated tumors >5 cm, according to clinical practice guidelines [(4)]. Hematoxylin & Eosin-stained preparations of the target tissue are first examined with biopsy. At this stage, histomorphologic features of the tissue are evaluated. In the ongoing process, immunohistochemical and molecular examinations specific to the subtypes included in the differential diagnosis are performed in most cases. Our project will focus on the

histopathologic evaluation process, the first stage of the diagnostic process. Hematoxylin & Eosin-stained preparation images will be used as a data set. Histologic type provides information about the surgical approach, neoadjuvant treatment decision, the chemotherapy agents to be applied after the surgical procedure, and the radiotherapy response so that misclassification will affect the treatment process. The subtypes included in our project are explained below respectively:

- Rhabdomyosarcoma (RMS) is children's most common soft tissue sarcoma and represents a high-grade neoplasm of myoblast-like cells in skeletal muscle [(5)]. According to the latest guidelines published by the World Health Organization, Rhabdomyosarcomas are divided into four subtypes: embryonic, Alveolar, Spindle/Sclerosing, and Pleomorphic. Each subtype contains different molecular patterns, but this classification is based more on morphology.
- Leiomyosarcoma (LMS) is one of the most common subtypes of soft tissue sarcomas in adults and can occur almost anywhere in the body. The most common anatomical locations are the abdomen, retroperitoneal region, and uterus. It develops from smooth muscle cells called leiomyocytes. The typical histologic pattern is intersecting fascicles formed by sharp-edged spindle cells with large eosinophilic cytoplasm and long, hyperchromatic nuclei [(6)].
- Synovial sarcomas are microscopically monophasic or biphasic. Monophasic synovial sarcoma consists of short, tightly packed bunches with narrow cytoplasm, dark chromatin, and uniform spindle cells. The biphasic type contains gland-like structures composed of cuboidal or columnar epithelioid cells and the spindle cell component mentioned above [(7)].
- Malignant peripheral nerve sheath tumors (MPNSTs) are malignant soft tissue sarcomas originating from Schwann cells with a high propensity for metastasis. MPNSTs constitute 5-10% of soft tissue sarcomas. It is among the most common soft tissue sarcomas other than rhabdomyosarcoma in pediatric patients. 20% of all MPNSTs are diagnosed in pediatric patients [(8)]. Histomorphologically, it consists of monotonous spindle cells with hyperchromatic nuclei in a fascicular pattern with

perivascular hypercellularity. High mitotic index and necrosis are striking features of MPNSTs [9].

The increasing use of artificial intelligence technologies in medicine shows that we can benefit from this technology in solving our problems. R&D studies using artificial intelligence technologies in medical image processing are progressing rapidly [10]. However, studies in the field of medical pathology are progressing more slowly due to the small number of samples, the large size of the digitized data, and the difficult access and high cost of the computer power required to work on the data.

Current studies are conducted on medically more easily solved tumors and are more common. [11] We are working on soft tissue sarcomas in our project but have not found enough place in such studies because they are less common than other tumor types. Our project will be a pioneering study on this subject, considering its clinical meaning and the difficulties it brings in data collection.

The great studies in computer vision contribute to solving our problems. Pre-trained models previously developed for image processing are the most effective tools for solving the current medical problem. In our project, we aim to use these pre-trained models to find the one with the optimum features for our problem.

The models to be developed with the knowledge that our study will provide insight will be pioneering in their application to routine clinical practice. This way, prolonged diagnostic processes will be shortened, and patients can reach the correct treatment regimens in shorter periods. In addition, the workload of expert pathologists working in this field will be reduced, and molecular diagnostic stages will be enabled faster.

Methodology (/ 25 Points)

Dataset

Phase One

In the initial phase, Dedifferentiated Liposarcoma, Leiomyosarcoma, Synovial Sarcoma, and Malignant Peripheral Nerve Sheath Tumor cases were extracted from the open-source TCGA-SARC dataset. High-resolution Hematoxylin and Eosin-stained scan images available on the GDC data portal were meticulously reviewed, and appropriate cases were selected for inclusion. A total of 20 resection slides were curated, comprising five instances from each of the four subtypes. Slides exhibiting less than 60% tumorous tissue in less than 60% of the total material were excluded from the study.

Phase Two

In the second phase, the same subtypes—Dedifferentiated Liposarcoma, Leiomyosarcoma, Synovial Sarcoma, and Malignant Peripheral Nerve Sheath Tumor—were sourced from the archival records of the Department of Medical Pathology, Department of Surgical Medical Sciences, at Istanbul University-Cerrahpaşa Cerrahpaşa Medical Faculty. Following ethical committee approval for the research, 30 slides meeting the criteria for pathological examination were selected under the guidance of expert pathologist Prof. Dr. Nil Çomunoğlu, with assistance from Expert Pathologist Research Assistant Ayse Mine Önenerk Men and Intern Dr. Hakan Tomac. The digitization of these slides, funded by Tübitak 2209-A student project support, was carried out by Medisa Dış Ticaret ve Pazarlama Limited Şirketi using a Leica Aperio GT 450 model whole slide imaging system at 40x magnification.

Data Labeling

The dataset was annotated to differentiate tumoral regions from non-tumoral areas. Digital labeling was conducted using the QuPath[17] platform, where tumor regions were marked. The marked tumor regions included artifacts such as staining errors, tissue debris, separations, and thick sections to replicate routine clinical conditions.

Data Preparation and Processing

Masking and Tiling Operations

The dataset was subjected to masking and tiling operations at 40x magnification. Frames with at least 60% tumorous tissue were generated during tiling. For each case, 2,000 random frames were selected, resulting in a total of 40,000 samples from each tumor group and case for the initial phase of the study. For the subsequent phase, an additional 40,000 samples per group were selected for training purposes. Five thousand frames were chosen from the remaining samples for the test set to enhance diversity and optimize sample utilization. Data augmentation was not applied during the training phase, as existing literature suggests it has a negligible impact on model performance.

Segmentation Problem

High-resolution images were registered at a 4x magnification level for the segmentation task, and data augmentation techniques were applied to enhance the dataset.

Accelerating Classification through Inheritance

Given the resolution of the images, which reaches hundreds of thousands of pixels, the initial approach to expedite the inference process was to reduce the resolution. However, since lowering resolution leads to information loss, as noted in the literature, transfer learning was employed to mitigate this issue. The data from the archive of Cerrahpaşa Medical Faculty were re-tiled with a zoom value of 16x, utilizing transfer learning to minimize the loss of information and preserve the integrity of the data. This approach ensured that high-resolution details were maintained while optimizing the efficiency of the inference process.

Phase One: Pretrained Model Comparison

Model Training and Testing

Model training and testing were conducted using Nvidia A100 40 GB graphics cards on the Google Colab Pro+ platform. This phase utilized the dataset created from TCGA-SARC data, with system limitations considered when selecting appropriate pre-trained model versions and making the dataset. The pre-trained models chosen for this study included ResNet50V2 with 25.6M parameters, VGG19 with 143M parameters, InceptionV3 with 24M parameters, and EfficientNetV2S with 21.6M parameters. These models were chosen based on their widespread use and development.

Model Selection and Training

The dataset was employed to train four different model types using the cross-validation method, incorporating the minimum necessary layers to evaluate the base performance of the models without increasing complexity. During model construction, both classification accuracy and simplicity of the model structure were prioritized, with an effort to avoid increasing the number of trainable parameters. The pre-trained models' existing parameters were frozen, and training was performed solely on the added layers. The features extracted from the pre-trained models were adapted to the problem using a classification segment consisting of one sampling layer and two fully connected layers.

Evaluation Metrics

The performance of the trained models was monitored using cross-validation, with classification accuracy, sharpness, and recall as the key metrics. For validation, the dataset was split into 80% training and 20% validation sets, ensuring no data leakage between cases. One case from each subgroup was entirely included in the validation set during cross-validation. Five different partitions were performed, with four cases in the validation set and sixteen in the training set out of the twenty cases. This methodology ensured robust evaluation and accurate assessment of model performance.

Phase Two: Training the Best Model

The model that demonstrated the best performance during the comparison process was subsequently trained using data obtained from the archive of Cerrahpaşa Medical Faculty. This training was conducted using the cross-validation method with an ImageNet-like approach. Additional layers were incorporated to achieve four-class outputs, ensuring no bottleneck at the network's end. All model parameters were utilized during the training process.

Monitoring and Validation

The success of the trained models was monitored using the cross-validation method, with key metrics including classification accuracy, sharpness, and recall. For validation, the dataset was divided into an 80% training set and a 20% validation set. Additionally, model performance was re-evaluated using the generated test set.

Careful measures were taken during cross-validation to ensure robust evaluation and prevent data leakage. One case from each subgroup was included in the validation set, ensuring no overlap between training and validation data. Out of 20 cases, five distinct partitions were created, with four cases in the validation set and sixteen in the training set. This methodology ensured a comprehensive and accurate assessment of the model's performance across different subsets of the data.

Phase Three: Training the Segmentation Model

Model Architecture and Training

Based on the initial benchmarking phase results, the most suitable architecture for the problem was selected and trained to perform binary segmentation. All 50 masked images that had undergone the annotation process were utilized, with the data being cross-validated using an 80% training and 20% validation split. No distinction was made between individual cases during this process.

Monitoring and Validation

The success of the trained models was assessed using the dice loss and accuracy metrics implemented through the cross-validation method. To enhance the model's robustness, various data augmentation techniques, including mirroring, freezing, and clipping, were applied during the training process. These augmentation methods aimed to increase the diversity of the training data and improve the model's generalizability.

Phase Four: Reducing Model Inference Times with Transfer Learning

Model Re-Training

In the second stage, the previously trained model was re-trained using the newly generated 16x dataset without altering the model structure. The dataset was divided into an 80% training set and a 20% validation set to ensure rigorous evaluation.

Monitoring and Validation

The performance of the re-trained models was monitored using classification accuracy, understanding, and recall metrics. The dataset was divided into five distinct partitions, with four cases in the validation and sixteen in the training sets. This method ensured that 20% of the data was consistently reserved for validation across all partitions, preventing data leakage and ensuring robust evaluation.

The model's inference times were shortened by employing transfer learning and re-training with a higher zoom level while maintaining high performance, as evidenced by the classification accuracy, understanding, and recall metrics. This approach effectively optimized the model for practical application without compromising the integrity of the results.

Results & Discussion (/ 30 Points)

Phase One Results: Pre-Trained Model Performance

Cross-validation and Model Performance

In the initial stage, five cross-validation tests were conducted to evaluate the performance of various pre-trained models on the validation set. The average classification accuracy rates achieved by the trained models were as follows: ResNet50V2: 0.49, VGG19: 0.41, EfficientNetV2S: 0.27, and InceptionV3: 0.39. These averages were derived from five different tests for each model.

Findings and Implications

The results of our study indicate that, in the realm of deep learning applications in digital pathology, the ResNet50V2 model—with its multi-layer architecture—demonstrates superior performance for soft tissue sarcoma classification. This finding suggests that ResNet50V2 is more suitable for developing models tailored to this classification task.

Dissemination of Results

The findings from the first phase of our study were documented and presented in a paper titled "Ön eğitimli görüntü işleme modellerinin yumuşak doku sarkomlarına yönelik sınıflandırma başarımlarının TCGA-SARC veri seti kullanılarak karşılaştırılması" This paper was presented at the 32nd National Pathology Congress in October 2023, contributing to the ongoing research and development in digital pathology and deep learning model optimization.

Phase Two Results: Enhancing Model Performance

Improvements and Cross-Validation

In the second phase, augmenting the size of the fully connected layers at the end of the network and conducting comprehensive training led to a 20-point improvement in the ResNet50V2

architecture's performance, which was identified as the best model in the first phase. The five-fold cross-validation yielded an average accuracy of 70.16%. In experiments conducted on the test set, an accuracy of 60% was achieved.

Phase Three: Segmentation Mask Removal and Model Pipeline Installation

Planned Improvements and Challenges

The third phase aimed to remove segmentation masks and implement a model pipeline for automatic glass-based evaluation. This approach was intended to reduce the areas needing evaluation on high-resolution images, thereby shortening evaluation times. Despite achieving good numerical results with the ResNet architecture, the models failed to generate masks effectively during the test phase. The high similarity of tissue samples and the substantial error rate among professionals indicated that the model struggled to learn with limited samples.

Phase Four: Faster Evaluation with Transfer Learning

Method and Results

In the fourth phase, another method proposed for faster evaluation involved compensating for the model's loss of information at low resolution through transfer learning. Despite this approach, the model's accuracy at the end of the training period was 68%, which was lower than in the first phase. Results from the test set, with an accuracy of 48%, confirmed that the information loss was significant at 16x magnification. This outcome demonstrated that high-resolution details are crucial for maintaining model accuracy and that important information loss occurs at lower resolutions.

The Impact and Future Directions (/ 15 Points)

Current and Upcoming Publications

We have published a paper detailing the results of this study and are currently preparing another manuscript based on our ongoing research. A literature review on image processing studies indicates that the primary reason for suboptimal results is the limited number of available images. To address this issue, we have prepared a new dataset comprising 250 STS slides with spindle morphology at the Department of Medical Pathology, Cerrahpaşa Medical Faculty. Upon completing the analysis of this new dataset, we intend to publish our findings and submit the entire process to a peer-reviewed journal.

Technical Considerations and Future Directions

Another strategy that may enhance model performance is resolving technical limitations to enable the use of entire frames rather than selecting samples from the dataset. Future studies are expected to focus on the further development and integration of auxiliary diagnostic systems into routine clinical practice. This advancement will reduce patients' reliance on reference centers, allowing diagnostic procedures to be performed at local centers. Implementing fast and efficient systems will mitigate the economic and medical issues caused by unnecessary diagnostic procedures and treatment delays, ultimately improving patient outcomes and healthcare efficiency.

References

1) van der Graaf, W. T. A., Orbach, D., Judson, I. R., & Ferrari, A. (2017). Soft tissue sarcomas in adolescents and young adults: a comparison with their pediatric and adult counterparts. The Lancet. Oncology, 18(3), e166– e175. https://doi.org/10.1016/S1470-2045(17)30099-2

2) Foersch, S., Eckstein, M., Wagner, D. C., Gach, F., Woerl, A. C., Geiger, J., Glasner, C., Schelbert, S., Schulz, S., Porubsky, S., Kreft, A., Hartmann, A., Agaimy, A., & Roth, W. (2021). Deep learning for diagnosis and survival prediction in soft tissue sarcoma. Annals of oncology: official journal of the European Society for Medical Oncology

3) Toro J.R., Travis L.B., Wu H.J., Zhu K., Fletcher C.D.M., Devesa S.S. Incidence patterns of soft tissue sarcomas, regardless of primary site, in the surveillance, epidemiology, and results program, 1978–2001: An analysis of 26,758 cases. Int. J. Cancer. 2006; 119:2922–2930. doi: 10.1002/ijc.22239.

4) WHO Classification of Tumours Editorial Board. Soft tissue and bone tumors [Internet]. Lyon (France): International Agency for Research on Cancer; 2020 [cited YYYY Mmm D]. (WHO Classification of Tumours series, 5th ed.; vol. 3). Available from: https://tumourclassification.iarc.who.int/chapters/33.

5) Rudzinski, E. R., Teot, L. A., Anderson, J. R., Moore, J., Bridge, J. A., Barr, F. G., Gastier-Foster, J. M., Skapek, S. X., Hawkins, D. S., & Parham, D. M. (2013). Dense pattern of embryonal rhabdomyosarcoma, a lesion easily confused with alveolar rhabdomyosarcoma: a Soft Tissue Sarcoma Committee of the Children's Oncology Group report. American journal of clinical pathology, 140(1), 82–90

6) George, S., Serrano, C., Hensley, M. L., & Ray-Coquard, I. (2018). Soft Tissue and Uterine Leiomyosarcoma. Journal of clinical oncology: official journal of the American Society of Clinical Oncology, 36(2), 144–150.

7)Kumar, V., Abbas, A. K., & Aster, J. C. (Eds.). (2018). Robbins primary pathology (10th ed.). Elsevier.

8) Knight, S., Knight, T. E., Santiago, T., Murphy, A. J., & Abdelhafeez, A. H. (2022). Malignant Peripheral Nerve Sheath Tumors Comprehensive Review of Pathophysiology, Diagnosis, and Multidisciplinary Management. Children (Basel, Switzerland), 9(1), 38.

9) Somatilaka, B. N., Sadek, A., McKay, R. M., & Le, L. Q. (2022). Malignant peripheral nerve sheath tumor: models, biology, and translation. Oncogene, 41(17), 2405–2421. https://doi.org/10.1038/s41388-022-02290-1

10) Pesapane, F., Codari, M., & Sardanelli, F. (2018). Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine. European radiology experimental, 2(1), 35.

11) Cai, L., Gao, J., & Zhao, D. (2020). A review of the application of deep learning in medical image classification and segmentation. Annals of translational medicine, 8(11), 713.

12) arXiv:1603.08029

13) Srikanth Tammina (2019); Transfer learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images; International Journal of Scientific and Research Publications (IJSRP) 9(10) (ISSN: 2250- 3153), DOI: http://dx.doi.org/10.29322/IJSRP.9.10.2019.p9420

14) Koonce, B. (2021). EfficientNet. In: Convolutional Neural Networks with Swift for Tensorflow. Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-6168-2_10

15) Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2017). Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. Proceedings of the AAAI Conference on Artificial Intelligence, 31(1). https://doi.org/10.1609/aaai.v31i1.11231

16) arXiv:2010.11929

17) Bankhead, P., Loughrey, M.B., Fernández, J.A. et al. QuPath: Open source software for digital pathology image analysis. Sci Rep 7, 16878 (2017). https://doi.org/10.1038/s41598-017-17204-5

18) Hicks, S. A., Strümke, I., Thambawita, V., Hammou, M., Riegler, M. A., Halvorsen, P., & Parasa, S. (2022). On evaluation metrics for medical applications of artificial intelligence. Scientific reports, 12(1), 5979.