PAPER 7

The Use of Deep Learning for Necrosis Assessment in Patients with Osteosarcoma

Christa L. LiBrizzi M.D, Jeremias Sulam PhD, Zhenzhen Wang, Aaron James, M.D, Adam Levin, M.D, Carol D. Morris, M.D

Background: Percent necrosis after neoadjuvant chemotherapy is considered a powerful prognostic factor for survival in osteosarcoma. Historically, less than 90% necrosis is associated with inferior survival whereas greater than 90% necrosis is associated with superior survival at 5 years. Currently, the assessment of percent necrosis is calculated by musculoskeletal (MSK) pathologists estimating tumor viability over an average of whole slide images (WSI). This process is labor intensive, requires subspecialized training, has high inter-observer variability, and is not standardized across institutions.

Questions/Purpose: We aim to develop, train, and validate a machine learning system capable of quantifying percent necrosis from OS WSI at an equivalent accuracy with improved precision compared to standard practice.

Patients/Methods: We performed a pilot proof of concept study. We retrospectively obtained 6 WSI of patients with high-grade, intramedullary osteosarcomas at our institution. The WSI were created from the definitive surgical procedure following preoperative chemotherapy. A weakly supervised machine learning system was developed to automatically segment viable tumor, necrotic tumor, and non-tumor cells and determine percent necrosis from WSI. We defined viable tumor as areas of >75% viable tumor, necrotic tumor as areas of >75% necrotic tumor, and areas of no tumor whatsoever. The deep learning system was taught by utilizing coarse segmentations by an MSK pathologist, which was then processed into individual patches. Patches were processed into groups of like patches called "bags," which were used to train the machine. All original labels were forgotten once the machine was trained. To compare the accuracy and precision of our machine learning system, an MSK pathologist was asked to annotate and calculate percent necrosis of each slide. The MSK pathologist was blinded to all patient HPI, type of osteosarcoma, and documented percent necrosis in the pathology report. We compared the machine learning system and the MSK pathologist percent necrosis by Pearson's correlation.

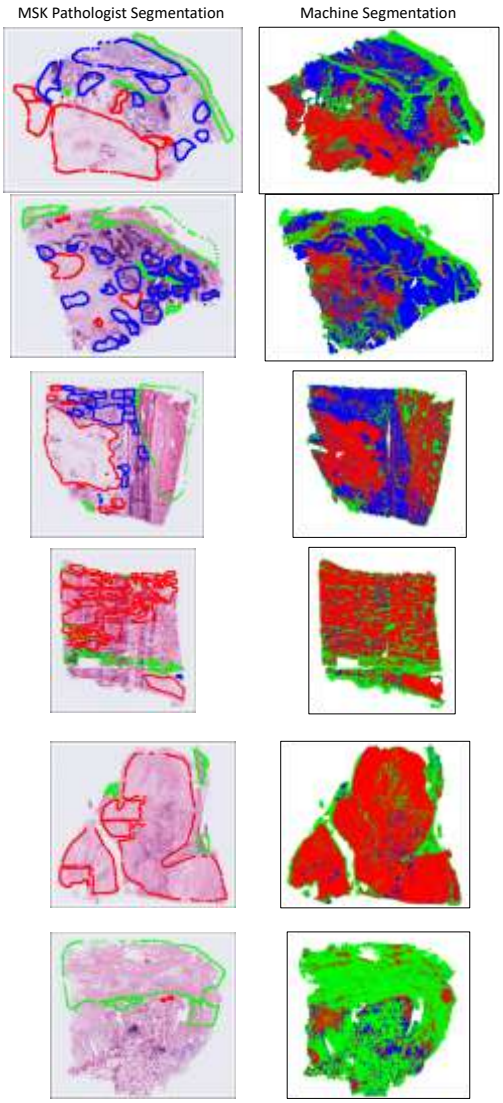
Results: Six WSI were segmented into areas of viable tumor, necrotic tumor, and non-tumor cells by both the machine learning system and the MSK pathologist. For slide one, percent necrosis estimations by the MSK pathologist versus machine system was 45% versus 58.28%, respectively; slide two was 20% versus 33.95%; slide 3 was 65% versus 60.32%; slide 4 was 5% versus 61.65%; slide 5 was 97% versus 94.45%; and slide 6 was 99% versus 96.89%. There was a strong, positive correlation between the machine system and MSK pathologist percent necrosis estimations by way of Pearson's correlation (r=0.84). Chondroblastic tissue proved to be an area for improvement in our algorithm.

Conclusion: Through this preliminary analysis, we were able to create and teach a machine learning system that was able to determine malignant versus benign mesenchymal cells and automatically calculate percent necrosis. This study also demonstrated a strong correlation compared to the standard of care of MSK pathologist percent necrosis estimation. Further segmentation of chondroblastic and other mesenchymal lineage cells are necessary to teach the machine learning system to allow for more accurate percent necrosis estimations.

Table 1: Percent necrosis estimates by musculoskeletal pathologist and machine learning system in patients with osteosarcoma

Slide ID	Pathologists	System
16714507	45%	58.28%
16714505	20%	33.95%
16714503	65%	60.32%
16714499	5%	61.65%
16714498	97%	94.45%
16714495	99%	96.89%

Figure 1: Segmentation of whole slide imaging by musculoskeletal pathologist and machine learning system



Red: necrotic tumor, Green: non-tumor, Blue: viable tumor