HIGH SCHOOL EDITION
Journal of Student Research
Volume 13 Issue 3 (2024)

Application of a Deep Learning Model for Effective Diagnosis of Osteosarcoma

Tanish Singh[1] and Kelly Ostrom[#]

[1]Godwin High School, USA
[#]Advisor

ABSTRACT

This experiment aimed to test the effects of different osteosarcoma tumor types on the accuracy of a trained neural network. Osteosarcoma is a subset of bone cancer seen greatly in pediatric cases and young adults, and its onset involves the rapid production of abnormal bone tissue, which leads to acute injuries. It is malignant by nature and life-threatening; therefore, it must be detected early by methods including X-ray, CT, MRI, and microscopy. The diversity of neoplastic formation makes classifying difficult, so using neural networks for detection can be beneficial. These machine-learning algorithms can identify patterns and analyze medical images for an accurate diagnosis. A dataset with three osteosarcoma classes was used: Viable Tumor, Non-Viable (Necrotic) Tumor, and Non-Tumor. All images were hematoxylin and eosin (H&E) stained cells used in training and testing. The structure of the network was a pre-trained model: VGG19, which uses the structure of a Convolutional Neural Network (CNN). The structure included the input, 18 convolution layers, max-pooling layers, and hidden layers for feature extraction. The multi-class classification accuracy was 96.27%, proving to be efficient. Large amounts of patient data can be quickly deciphered, and this could lead to advancements in medical imaging and disease prognosis.

Introduction

Early detection and treatment are pivotal for an expected prognosis of diseases, as untimely diagnosis could lead to transgressional health. As the current conventional diagnosis of osteosarcoma bone tumors consists of imaging and pathological examination of high histological variation, osteosarcoma detection has been difficult; reform is needed. Computer-aided technology, such as machine learning, has assisted in radiological and microscopy detection. A machine learning method, deep learning, utilizes multilayer convolutional neural networks (CNN) and can automatically extract features from a dataset, accurately classifying such tumors (Bansal et al., 2022). As less time is spent forming a diagnosis, the risk of a worse prognosis will be mitigated. The probability of implementing an incorrect diagnosis will also significantly decrease as deep learning algorithms have provided enhanced efficacy (Ahammad et al., 2022). This experiment will be conducted to display osteosarcoma tumor types on the neural network classification accuracy. Three types of tumors will be seen: Necrotic, Viable, and Non-Tumor. As osteosarcoma is a rare bone cancer, the research conducted can develop healthcare systems and classify images with comparable accuracy to a trained oncologist.

Real-world implications arise in the experiment due to AI advancement and deep neural networks' nurturing. The increasing adoption of deep learning in healthcare has accelerated the analytics of complex malignancies in tumor detection (Mohamed et al., 2022). Deep learning algorithms are directly applicable to the derivation of oncological issues. Applications help focus on omics data types and histological images such as microscopy and scan derived. By identifying patterns using mathematical models, clinical tasks that require a significant time consumption can be easily replaced to preface efficiency by automating these tasks with high

accuracy (Anisuzzaman et al., 2021). Computer-aided detection systems (CAD) focus on the detection of osteosarcoma using magnetic renaissance imaging (MRI) and computed tomography (CT). However, these systems have invoked limitations. Macroscopic information does not provide sufficient information for osteosarcoma tumor detection (Bansal et al., 2022). Therefore, whole slide imaging (WSI) can be utilized to differentiate viability in tumors. This can be seen on a larger scale as the direction of research towards microscopy can provide a more suitable detection method for other tumors that yield macroscopic analysis. Alongside this, depth in research will allow greater information on low-studied tumors such as osteosarcoma.

Osteogenic sarcoma (OS), or osteosarcoma, is a primary malignant tumor stemming from mesenchymal tissue, enabling osteoid proliferation. Osteosarcoma is the second contributor to child and adolescent death, and 5-10% of new pediatric cases consist of primary bone sarcomas (Prater & McKeon, 2022). Although 5 years old to teenage years are the general patients, the occurrence has been seen in older populations (>65) with a linkage to past radiation therapy or Paget's disease. These tumors generally stem in the lower limbs, and the onset of the site includes warmth, redness, and mild bone pain. OS is heterogeneous, creating divisions between subtypes based on factors such as infected sites and histological variation. These subtypes have a low correlation with demographics, biological behavior, and imaging appearance, as these factors all vary (Lindsey et al., 2017). The tumor originates in the metaphysis of long body bones in quick-growing bone locations. For this reason, there is a high incidence of OS in areas such as the proximal humerus, proximal tibia, and distal femur. The pelvis, sacrum, and spine contain a minimum percentage of all total cases (Bansal et al., 2022). Most patients contain only a single lesion upon diagnosis. Scientific research and treatment have improved for this disease; however, the etiology is imprecise. Around 80-90% of patients possess micro-metastases before onset, which cannot be detected accurately. Therefore, detecting sarcoma in subclinical situations with new diagnostic approaches is imperative.

Whole Slide Images are being used to surpass limitations within MRI and CT scans. Cells are generally identified from Hematoxylin and Eosin-stained tissue (Bansal et al., 2022). Tissue taken from disease sites is placed under the microscope and mounted on a glass slide for pathological analysis. Experimentally, the dataset used has created digital WSIs and classified the given tumors into 3 subclasses (Viable Tumor, Non-Viable Tumor, and Non-Tumor). A viable tumor is defined as the presence of malignancy, a non-viable tumor shows signs of irreversible cell damage, and a non-tumor indicates a benign state. The cancerous bone cells are densely populated with an irregular shape, and darker stained shades are visible in image properties (Anisuzzaman et al., 2021). These features differentiate the three tumor subclasses and provide a more precise understanding than MRI or CT.

Machine learning algorithms in image classification have emerged because of the progress in medical imaging and the need for a method that can access large amounts of clinical data to aid physicians and scientists. Advancements in hardware allow for greater computational power, and the availability of datasets allows for research merging computer science and medicine. Deep convolutional neural networks (CNN) have been seen in many papers for cancer diagnosis and disease prevention, such as breast cancer, myocardial infarctions, pancreatic cancer, etc. These algorithms use mathematical functions to model and predict data points, introducing an external, non-biased source for classification in radiology or pathology. Early automatic detection can alleviate the workload for practitioners and serve as an assistance tool for surgeons pre-op. Problems in machine learning arise due to the data, and because the dataset did not include a balanced number of images, preprocessing techniques were used to increase the image count. The model can best converge when the data is high quality, which can be an issue in the medical field.

The conduction of this experiment directly tested the osteosarcoma tumor Whole Slide Images (WSI) on a multiclass classification of Viable Tumor, Non-Viable Tumor, and Non-Tumor. The levels were an experimental prerequisite as the dataset obtained had pathologist verification for sorted images of tumor types.

Methods and Materials

In this experiment, the dataset was derived from The Cancer Imaging Archive (TCIA) Public Access, consisting of 1144 Whole Slide Images (WSI) of 1024 X 1024 with 10X resolution of stained osteosarcoma histology images. Image preprocessing was crucial to creating a training dataset as the images were not uniform and had excess noise that the model could potentially pick up on. The standardization of images created consistent data and led to greater prediction accuracy. This technique assists in preventing overfitting and improving model generalization, which is crucial when working with small datasets. The datasets were split at a ratio of 70%, 20%, and 10%, respectively. The data was then resized to a 3-dimensional tensor with 224x224 pixels using the OpenCV library and iterating over the three image categories to place them in their respective folders. Augmentations were then performed on the data to create more images and implement adaptive changes within the testing set. One thousand one hundred image batches were created using Keras's ImageDataGenerator. Augmentations included rotations, width and height shifts, shear, zoom, vertical flip, and the nearest fill mode.

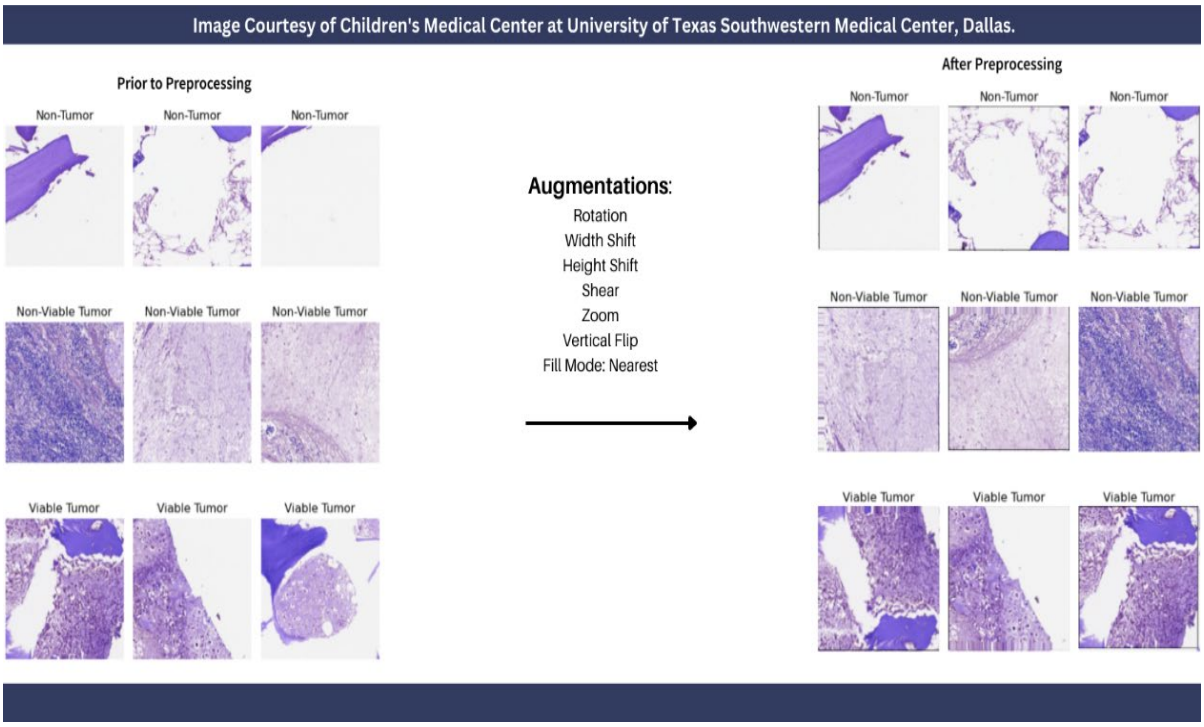


Figure 1. Image Preprocessing Method. This figure shows the augmentations done to the photos in each class to generate new images and increase variability in the model's predictions.

The specific architecture used to differentiate and make predictions for the classes -Non-Tumor, Necrotic Tumor, and Viable Tumor- was a CNN with 19 layers (not accounting for the pooling layers). Specifically, a pre-trained model (VGG19) was used and focused on feature extraction of specific colors and shapes in the osteosarcoma slide images. Multiple convolutional layers performed distinct operations on the data as the images were converted into tensors, in this case, three matrixes of RGB pixels, because of the images. After trying many optimizers, the Adam optimizer (Adagrad and RMSprop combined) with a learning rate of 0.01 was used to train the model, as it did not converge too quickly and had time to learn the complex patterns in the slides. After noticing a plateau or no significant growth in accuracy midway through testing, I employed a learning rate schedule to reduce the rate by a factor of 0.1 on the twentieth epoch. This technique helped the

model converge after no distinct growth. The model was trained for 35 epochs with a batch size of 32. The training and testing were performed on a computer containing an Nvidia GeForce GTX 1660 Super GPU. The osteosarcoma tumor type was assessed for an efficient addition to pathologist validation for patients. This measured the correctly predicted images among the data. All the data saved in the model prediction was statistically analyzed to create graphs of producible metrics and collect statistical data, including precision, recall, f1-score, and support. Some constants in this experiment were the dataset, WSI tile size, GPU used, train-test split, VGG19 base model layers, coding language, and deep learning API. To properly assess the risk and safety that could have been brought from this experiment, frequent breaks were taken within the experiment to prevent eye strain from the computer.

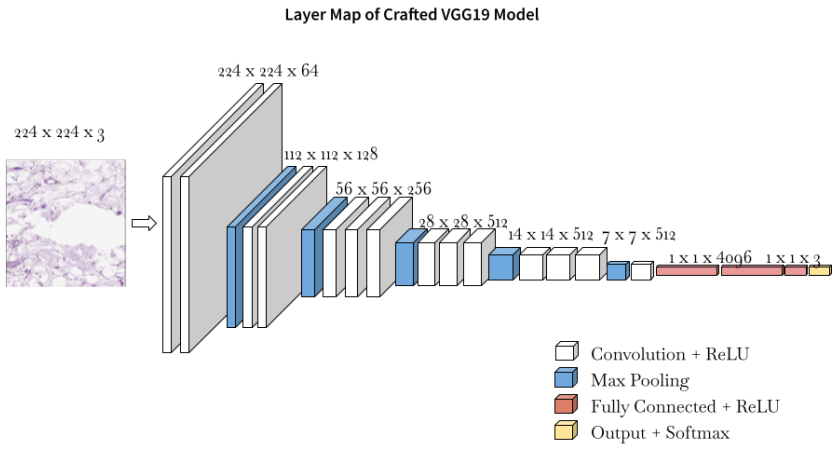
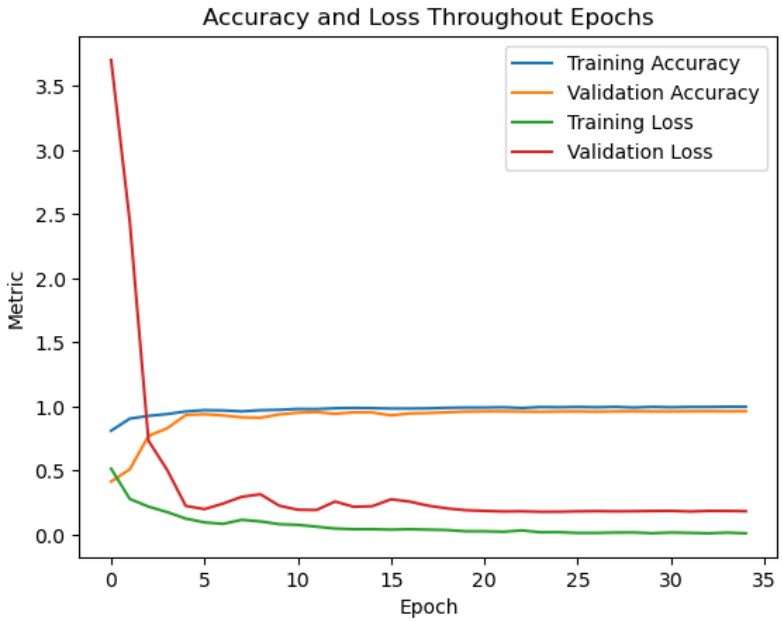


Figure 2. Layer Map of Crafted VGG19 Model. This map indicates the layer composition of a specific Convolutional Neural Network (CNN), showing the neural flow of the given image.

Results

The model focused on testing the effects of different osteosarcoma tumor types on the accuracy of a trained neural network. Multiple graphs, figures, and tables were generated from the Python libraries Matplotlib, Scikitlearn, and Seaborn. These were utilized to collect and summarize the results of an experiment. Table 1 displays the analysis of results, including the model metrics of validation accuracy, precision, recall, f1-score, and recall. Table 2 compares the proposed model to other state-of-the-art algorithms, comparing accuracy and model type. Graph 1 displays the graphical representation of the model's training with the metrics of training accuracy, validation accuracy, training loss, and validation loss throughout 35 epochs.



Graph 1. Accuracy and Loss Throughout Epochs. The graphed accuracies and loss metrics are present over 35 epochs, showing little to no overfitting within validation.

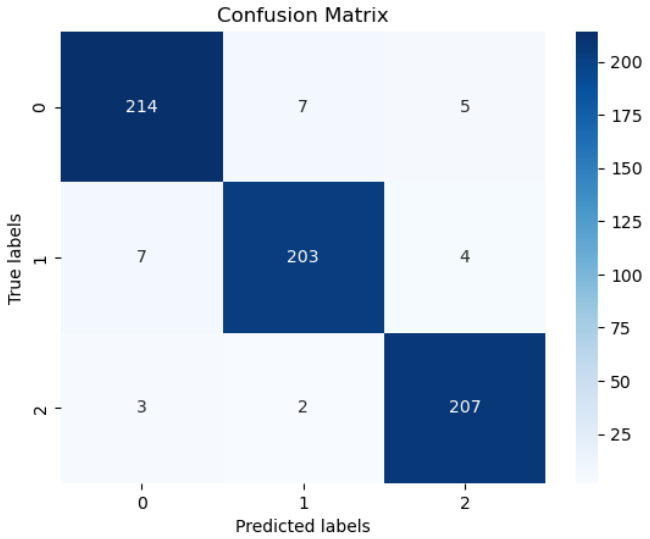


Figure 3. Confusion Matrix. A confusion matrix heatmap displays the correct predictions of the model in dark blue, with the labels 0, 1, and 2 being Non-Tumor, Viable Tumor, and Non-Viable Tumor, respectively.

Table 1. The statistical evaluation of the model's performance.

	Validation Accuracy (%)	**Precision (%)**	**Recall (%)**	**F1 Score (%)**	**Support**
Non-Tumor	96.45	96.00	95.00	95.00	226
Viable Tumor	97.02	96.00	95.00	95.00	214

Non-Viable Tumor	95.33	96.00	98.00	97.00	212
Average	96.27	96.00	96.00	95.67	217.33

Table 2. Comparison to algorithms of a similar task.

Authors	Year	Model	Accuracy (%)
Mishra, Rashika et al.	2017	CNN	84
Arunachalam, Harish Babu, et al.	2017	K-means flood fill	95.5
Mishra et al.	2018	AlexNet, LeNet, VGGNet	92
Arunachalam, Harish Babu, et al.	2019	VGG19	96
Anisuzzaman et al.	2021	CT, SVM, ENS	89.9
Liangrui Pan et.al.	2022	NRCA-FCFL	99.17
Nasir et al.	2022	AlexNet	99.30
Proposed Model	**2023**	**VGG19**	**96.27**

The Viable group had the greatest validation accuracy compared to the other comparable tumor types. The accuracy for this group was 97.2%, which was greater than Non-Tumor and Non-Viable Tumor at the values of 96.45% and 95.33%, respectively. Overall growth throughout each epoch was present, as the data was learning at a steady rate. Some fluctuations in the validation accuracy were seen throughout the testing, but there was a net increase in the final epoch. Results were very accurate on all accounts, as precision is achieved by mathematical computations such as Convolution and MaxPooling operations, and the model was compiled with a sufficient optimizer. There were minimal signs of overfitting as the validation and training accuracies stayed similar, as seen in Graph 1. These highly accurate findings are comparable to pathological human verification and are used in accordance, as seen in Figure 1. A batch size of 32, a learning rate of 0.01 from epochs 1-19, and 0.001 from 20-35, and the Adam optimizer yielded the results.

Conclusion

The proposed model tested different osteosarcoma tumor types on the accuracy of a neural network. Images from a dataset of three tumor types, Viable Tumor, Non-Viable (Necrotic), and Non-Viable Tumor, were derived to create a deep learning algorithm. The validation accuracy of the neural network was the main metric being measured. The data was categorically measured and showed the model's effectiveness in determining osteosarcoma tumor types. This research proved that high accuracies are possible for algorithm creation, which is better than standard diagnostic imaging techniques and human validation.

Countless scientists have studied and created machine-learning models for medical assessments of different ailments and tumors. Anisuzzaman et al. from the University of Wisconsin-Milwaukee conducted a deep-learning study for osteosarcoma detection using similar microscopy-derived images, achieving similar accuracy. Yogananda et al. conducted another study in 2019, which examined a deep-learning model for brain tumor segmentation. Both experiments proved that high accuracies are possible for algorithm creation, which is better than standard diagnostic imaging techniques and human validation.

Many conditions allowed for results in this overall experiment. The optimization techniques used for values of the layers significantly increased the accuracy, as did using a pre-trained model, VGG19. This displays the potential of such algorithms to achieve excellence in the diagnosis of patients and tumor detection. As large

amounts of data were processed, including over 3000 images, the model was set out to provide patient care by reducing the time of diagnosis and having accurate results to reduce the chance of misdiagnosis.

Using any computational technique, including a deep learning model for the classification of tumors, brings sources of errors that cannot be controlled. Datasets that are missing values or have incorrect labels can lead to errors in classification because the model will incorporate incorrect patterns in its schema, or there will not be enough information for the algorithm to make confident predictions. Limited computational power is also an obstacle in CNN use because complex models require enough graphical power and memory, which is difficult to implement in rural or low-income hospital settings. A fast, high-accuracy algorithm that requires minimal computational effort coupled with pathologist validation can assist in diagnosis and mitigate the troubles faced in medical society. This research was conducted to strengthen the identification of a highly misdiagnosed tumor, increase the efficiency and quality of patient care, and embark on future research in other cancers or diseases for classification.

Acknowledgments

I would like to thank my advisor for the valuable insight provided to me on this topic.

References

Ahammad, M., Abedin, M. J., Khan, Md. A. R., Alim, Md. A., Rony, M. A. T., Alam, K. M. R., Reza, D. S. A. A., & Uddin, I. (2022). A Proficient Approach to Detect Osteosarcoma Through Deep Learning. *2022 10th International Conference on Emerging Trends in Engineering and Technology - Signal and Information Processing (ICETET-SIP-22)*. https://doi.org/10.1109/icetet-sip-2254415.2022.9791502

Anisuzzaman, D. M., Barzekar, H., Tong, L., Luo, J., & Yu, Z. (2021). A deep learning study on osteosarcoma detection from histological images. *Biomedical Signal Processing and Control*, *69*, 102931. https://doi.org/10.1016/j.bspc.2021.102931

Arooj, S., Atta-ur-Rahman, Zubair, M., Khan, M. F., Alissa, K., Khan, M. A., & Mosavi, A. (2022). Breast Cancer Detection and Classification Empowered With Transfer Learning. *Frontiers in Public Health*, p. *10*. https://doi.org/10.3389/fpubh.2022.924432

Arunachalam, H. B., Mishra, R., Daescu, O., Cederberg, K., Rakheja, D., Sengupta, A., Leonard, D., Hallac, R., & Leavey, P. (2019). Viable and necrotic tumor assessment from whole slide images of osteosarcoma using machine-learning and deep-learning models. *PLOS ONE*, *14*(4), e0210706. https://doi.org/10.1371/journal.pone.0210706

Bansal, P., Gehlot, K., Singhal, A., & Gupta, A. (2022). Automatic detection of osteosarcoma based on integrated features and feature selection using binary arithmetic optimization algorithm. *Multimedia Tools and Applications*, *81*(6), 8807–8834. https://doi.org/10.1007/s11042-022-11949-6

Lindsey, B. A., Markel, J. E., & Kleinerman, E. S. (2016). Osteosarcoma Overview. Rheumatology and Therapy, 4(1), 25–43. https://doi.org/10.1007/s40744-016-0050-2

Mishra, R., Daescu, O., Leavey, P., Rakheja, D., & Sengupta, A. (2017). Histopathological Diagnosis for Viable and Non-viable Tumor Prediction for Osteosarcoma Using Convolutional Neural Network. *Bioinformatics Research and Applications*, pp. 12–23. https://doi.org/10.1007/978-3-319-59575-7_2

Mishra, R., Daescu, O., Leavey, P., Rakheja, D., & Sengupta, A. (2018). Convolutional Neural Network for Histopathological Analysis of Osteosarcoma. *Journal of Computational Biology*, *25*(3), 313–325. https://doi.org/10.1089/cmb.2017.0153

Mohamed, E. A., Rashed, E. A., Gaber, T., & Karam, O. (2022). Deep learning model for fully automated breast cancer detection system from thermograms. *PLOS ONE*, *17*(1), e0262349. https://doi.org/10.1371/journal.pone.0262349

Nasir, M. U., Khan, S., Mehmood, S., Khan, M. A., Rahman, A., & Hwang, S. O. (2022). IoMT-Based Osteosarcoma Cancer Detection in Histopathology Images Using Transfer Learning Empowered with Blockchain, Fog Computing, and Edge Computing. *Sensors*, *22*(14), 5444. https://doi.org/10.3390/s22145444

Pan, L., Wang, H., Wang, L., Ji, B., Liu, M., Chongcheawchamnan, M., Yuan, J., & Peng, S. (2022). Noise-reducing attention cross fusion learning transformer for histological image classification of osteosarcoma. *Biomedical Signal Processing and Control*, p. *77*, 103824. https://doi.org/10.1016/j.bspc.2022.103824

Prater, S., & McKeon, B. (2023, May 23). Osteosarcoma. Nih.gov; StatPearls Publishing. https://www.ncbi.nlm.nih.gov/books/NBK549868/

Sarker, I. H. (2021). Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. *SN Computer Science*, *2*(6). https://doi.org/10.1007/s42979-021-00815-1

Yang, S., Zhu, F., Ling, X., Liu, Q., & Zhao, P. (2021). Intelligent Health Care: Applications of Deep Learning in Computational Medicine. *Frontiers in Genetics*, p. *12*. https://doi.org/10.3389/fgene.2021.607471

Zhao, X., Wu, Q., Gong, X., Liu, J., & Ma, Y. (2021). Osteosarcoma: a review of current and future therapeutic approaches. *BioMedical Engineering OnLine*, *20*(1). https://doi.org/10.1186/s12938-021-00860-0

Zhou, X., Wang, H., Feng, C., Xu, R., He, Y., Li, L., & Tu, C. (2022). Emerging Applications of Deep Learning in Bone Tumors: Current Advances and Challenges. *Frontiers in Oncology*, p. *12*. https://doi.org/10.3389/fonc.2022.908873