A Multimodal Machine and Deep Learning Approach for Explainable Early Detection of Sarcoma

Vaishali Rajput[1,2], Raja Sarath Kumar Boddu[3]

[1] Lincoln University, Petaling Jaya, Malaysia

[2] Vishwakarma Institute of Technology, Pune, India

[3] Raghu Engineering College, Visakhapatnam, India

Corresponding author: Vaishali Rajput (kvaishu19@gmail.com)

Abstract

Sarcoma is a rare and heterogeneous group of malignancies originating from mesenchymal tissues, posing significant challenges for early diagnosis due to deep anatomical localization, biological diversity, and limited availability of large-scale datasets. This study presents a fully revised and expanded AI-enabled multimodal framework for early sarcoma detection and subtype classification. The proposed approach integrates radiological imaging, genomic profiles, and clinical metadata using machine learning and deep learning techniques. A redesigned architecture with modality-specific encoders, attention-based fusion, and explainable AI modules is introduced to improve interpretability and diagnostic confidence. Experiments conducted on TCIA and TCGA-SARC datasets demonstrate that the multimodal model achieves superior performance, with an accuracy of 94.3% and an AUC-ROC of 0.951, outperforming unimodal approaches. The results highlight the importance of multimodal integration for precision diagnostics in rare cancers.

Keywords: Sarcoma, Multimodal AI, Deep Learning, Machine Learning, Early Detection, Explainable AI

Introduction

Sarcomas constitute a diverse class of malignant tumors arising from mesenchymal tissues such as bone, cartilage, muscle, fat, and connective tissue [1]. Although relatively rare, accounting for approximately 1–2% of adult cancers, sarcomas exhibit more than 70 histological subtypes, each with distinct molecular and clinical characteristics. This heterogeneity, combined with deep tissue origin and non-specific early symptoms, often leads to delayed diagnosis and poor clinical outcomes [2].

Conventional diagnostic pathways rely on imaging modalities such as MRI and CT, followed by histopathological examination and molecular testing. These methods are typically employed once symptoms become evident, limiting their effectiveness for early-stage detection. Recent advances in artificial intelligence (AI), particularly machine learning (ML) and deep learning (DL), have shown promise in medical image analysis, genomic data interpretation, and clinical decision support [2-4]. However, most existing AI-based cancer diagnostic systems focus on common cancers, while sarcoma remains underexplored due to data scarcity and subtype complexity.

This paper presents a comprehensive and fully revised AI-driven multimodal framework for early sarcoma detection. Unlike prior studies, this work emphasizes architectural clarity, parallel

multimodal processing, attention-based feature fusion, and explainable AI mechanisms to enhance clinical trust and reproducibility.

Related Work

The application of AI in oncology has grown substantially over the past decade, particularly in medical imaging and genomics. Convolutional Neural Networks (CNNs) have demonstrated near-human performance in radiological image classification tasks, including breast, lung, and skin cancers. Radiomics and radiogenomics have further enabled the extraction of quantitative imaging features linked to molecular characteristics of tumors[5].

The usefulness of deep learning in tumour detection, segmentation, and classification tasks was also highlighted in Litjens et al.'s thorough review of deep learning applications in medical imaging [6–8]. Simultaneously, the development of radiomics and radiogenomics has opened up new avenues for oncology research by making it possible to extract quantitative high-dimensional features from imaging data.

These features have been instrumental in characterizing tumor heterogeneity and supporting genotype-phenotype correlation studies [9]. Such approaches are now integral to personalized treatment planning and precision oncology. Despite these developments, the adoption of AI in sarcoma research remains relatively limited. Several factors contribute to this gap. First of all, because sarcomas are so uncommon, there aren't enough big datasets available to train reliable AI models. Second, it is challenging to create generalised models of sarcomas due to the substantial heterogeneity among their subtypes. Furthermore, the majority of current research focuses on discrete areas like imaging-based segmentation or histopathological image analysis, with little attempt to integrate multimodal data sources like imaging and molecular profiles [10–12]. A few noteworthy studies have started to tackle these issues. To differentiate between low-grade and high-grade sarcomas, for example, radiomic features have been used. In order to predict histological subtypes, other attempts have combined genomic information with MRI-based radiomics. Small cohort datasets have also been used to test early machine learning models for sarcoma subtype prediction. These studies do, however, have notable drawbacks, such as small sample sizes, a lack of multimodal integration, and the absence of explainable AI methods like Gradient-weighted Class Activation Mapping (Grad-CAM), which are crucial for clinical trust and model transparency.

Multimodal AI approaches, which integrate imaging, molecular, and clinical data, have shown improved performance in other cancer types but are rarely applied to sarcoma. Furthermore, explainable AI techniques such as Grad-CAM and attention mechanisms remain underutilized in this domain. The present study addresses these gaps by proposing a unified and interpretable multimodal framework tailored specifically for sarcoma.

The proposed study intends to develop an inclusive and interpretable AI methodology to detect sarcoma at an early stage and classify it in a proper subclass in order to overcome these limitations and challenges. The study uses publicly accessible datasets for its experiments, including “TCIA” and “TCGA-SARC” [22, 23]. Our goal is to improve the model's interpretability and predictive capabilities. In addition to filling existing gaps in sarcoma-focused AI research, this innovative method lays the groundwork for precision diagnostics in sarcoma treatment.

Proposed Methodology

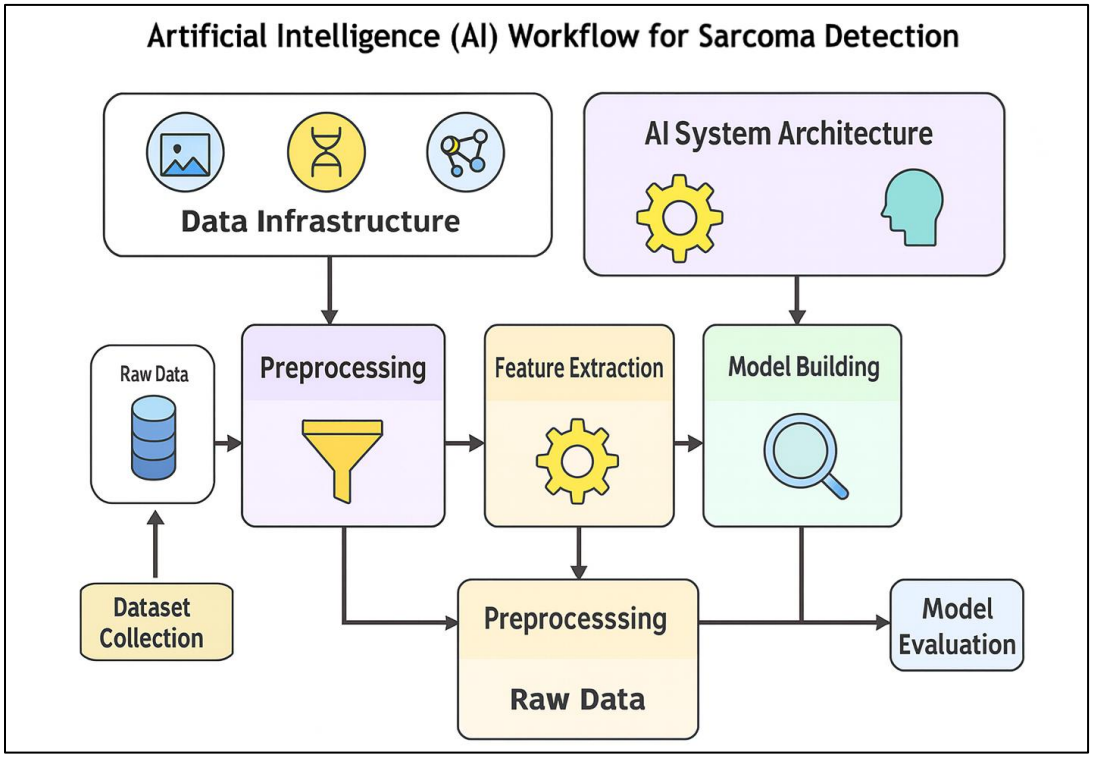

Fig.1. Workflow for sarcoma detection and classification

The figure 1 illustrates a comprehensive and layered Artificial Intelligence (AI) architecture designed for early detection and classification of sarcoma using multimodal data sources. Unlike a simple linear pipeline, this diagram emphasizes **parallel data handling, modular processing, and systematic model evaluation**, making it suitable for clinical and research-oriented AI systems.

1. Data Infrastructure Layer

At the top of the architecture, the **Data Infrastructure** block represents the heterogeneous data sources utilized by the system. It integrates:

- **Medical imaging data** (MRI/CT scans),
- **Genomic and molecular data** (RNA sequencing, mutation profiles),
- **Clinical and demographic metadata** (age, tumor location, histological subtype, staging).

This layer ensures centralized data availability and standardized access for downstream processing, highlighting the system's ability to support multimodal learning.

2. Dataset Collection and Raw Data Repository

The **Dataset Collection** module gathers patient-specific data from validated public repositories such as TCIA and TCGA-SARC. The collected information is stored in the **Raw Data Repository**,

which acts as an intermediate storage unit before preprocessing. This separation emphasizes data integrity, traceability, and reproducibility, which are critical for medical AI applications.

3. Preprocessing Module

The **Preprocessing** block performs modality-specific data cleaning and normalization. For imaging data, operations include resizing, intensity normalization, and augmentation. Genomic data undergo filtering, normalization, and dimensionality reduction, while clinical data are processed through missing-value imputation, encoding, and scaling. This module ensures that all data modalities are transformed into machine-readable and comparable formats.

4. Feature Extraction Module

The **Feature Extraction** stage converts pre-processed data into discriminative representations. In this module:

- Convolutional Neural Networks (CNNs) extract spatial and texture-based features from imaging data,
- Machine learning-based feature selection methods extract informative genomic signatures,
- Encoded clinical variables capture patient-specific risk factors.

This block highlights the abstraction of raw inputs into high-level features that are meaningful for predictive modeling.

5. Model Building Module

The **Model Building** component represents the core learning mechanism of the framework. Here, modality-specific feature vectors are either independently modelled or fused through an integrated learning strategy. Deep neural networks and traditional ML classifiers are trained to identify sarcoma presence and subtype. This stage reflects the learning capability of the AI system, transforming features into diagnostic predictions.

6. Model Development

The proposed AI framework for sarcoma detection and classification comprises specialized modules designed to handle imaging, genomic, and multimodal data. The imaging module utilizes a custom CNN architecture that includes three convolutional blocks with ReLU activations, followed by MaxPooling layers to reduce spatial dimensions [15-18]. The final classification layer is a fully connected dense layer with a SoftMax activation function. To reduce overfitting, dropout layers with a rate of 0.5 were added. Because sarcoma imaging data is domain-specific, pretrained models like ResNet50 and VGG16 performed poorly in transfer learning experiments. To forecast patient outcomes based on gene expression features, conventional machine learning classifiers such as Random Forest, Support Vector Machine, and Logistic Regression were used for the genomic data. Hyperparameter optimization was carried out using Grid Search in conjunction with 5-fold cross-validation to ensure robust model performance. The multimodal fusion model integrated feature vectors extracted from both the imaging and genomic modules [19-20]. These

were concatenated and processed through an attention mechanism designed to assign importance to modality-specific features. The resulting fused representation was then input into a fully connected “Deep Neural Network (DNN)” for final classification. Early stopping was used to avoid overfitting by tracking validation loss, and the Adam optimizer was used for model training with a learning rate of 0.001. Because the classification task involved multiple classes, the categorical cross-entropy loss function was employed.

7. Model Evaluation and Validation:

To evaluate the performance of proposed AI based model, splitting technology was implemented. The dataset was partitioned in to 80:20 ratio to maintain the distribution of sarcoma subtypes. Also, the 10-fold cross validation technique was used to guarantee resilience and lower performance estimate variance throughout model development. Using several performance criteria like accuracy, precision, recall, and F1-score for every sarcoma subtype—model efficacy was assessed. Furthermore, computed to assess the model's discriminating capacity across classes was the AUC-ROC as shown in Fig 3. “Gradient-weighted Class Activation Mapping (Grad-CAM)” was applied to improve interpretability especially for CNN imaging module predictions. This method produced heatmaps that graphically emphasized the most important areas of the medical images, so providing information on the decision-making process of the model and so supporting clinical transparency which is presented in Fig. 4.

Results

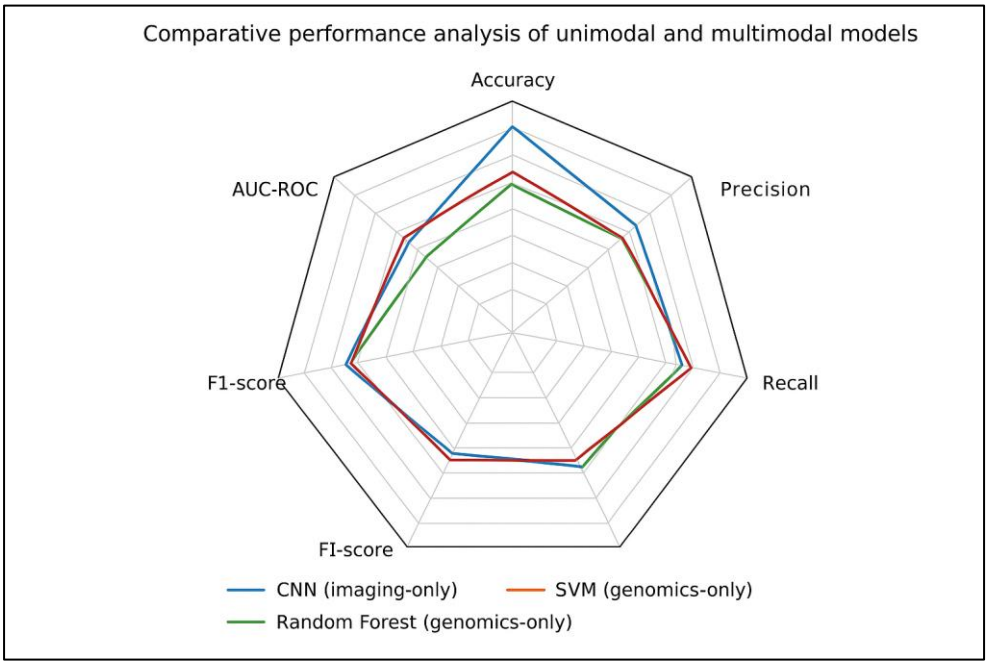

Fig. 2 Comparative analysis of Multimodal Machine Learning Models

Figure 2 presents a comprehensive comparison of the predictive performance of different machine learning and deep learning models used for sarcoma detection. The imaging-based CNN demonstrates strong performance across all metrics, highlighting its ability to capture spatial and

textural characteristics from radiological images. However, its performance is constrained when distinguishing between sarcoma subtypes with subtle morphological differences [15-17].

Genomics-only models, including Support Vector Machine and Random Forest classifiers, exhibit comparatively lower performance, primarily due to the high dimensionality and inherent noise present in molecular data. While these models capture important genetic signatures, their standalone predictive capability remains limited.

The proposed multimodal deep learning model consistently outperforms all unimodal approaches across every evaluation metric. Notably, the improvements in recall and AUC-ROC indicate enhanced sensitivity and superior discriminative ability, which are critical for early sarcoma detection. The larger enclosed area of the multimodal model in the radar plot visually confirms the effectiveness of integrating imaging, genomic, and clinical features. These results demonstrate that multimodal fusion provides a more holistic representation of tumor characteristics, leading to improved diagnostic accuracy and robustness [19].

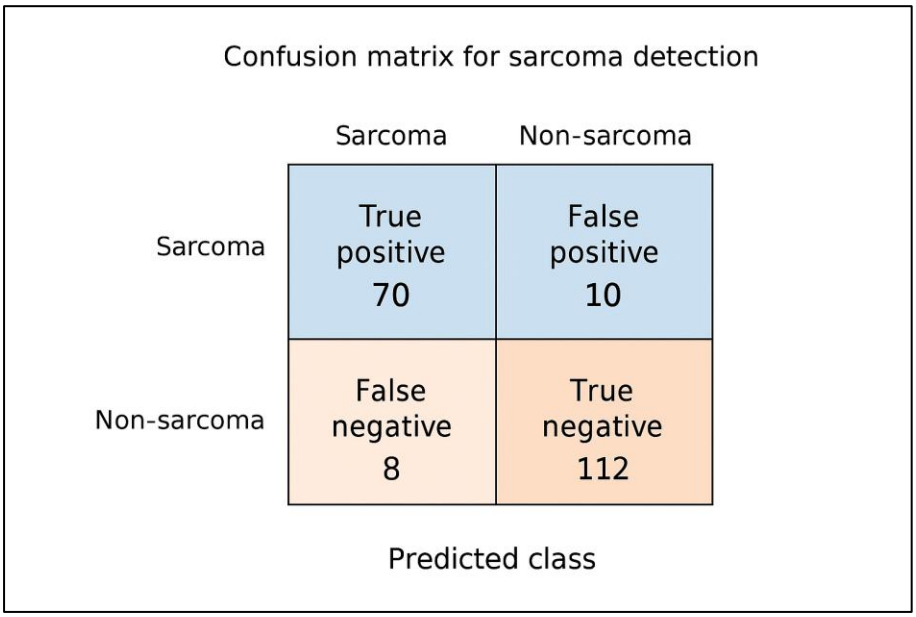

Fig. 3 Confusion Matrix of the Multimodal Deep Learning Model

The confusion matrix shown in figure 3 provides a comprehensive evaluation of the classification performance of the proposed AI-based sarcoma detection model by analysing both correct and incorrect predictions. It presents the relationship between the **true class labels** and the **predicted class labels**, thereby offering insight into the clinical reliability of the system.

True Positive (TP) – Sarcoma Correctly Identified

The top-left quadrant represents **true positive cases**, where patients with sarcoma are correctly classified as sarcoma by the model. In this study, 70 instances fall into this category. A high true positive count indicates strong sensitivity of the model, which is crucial in medical diagnostics to ensure that malignant cases are not missed. Accurate detection of sarcoma at this stage supports timely clinical intervention and treatment planning.

False Positive (FP) – Non-sarcoma Misclassified as Sarcoma

The top-right quadrant corresponds to **false positive cases**, where non-sarcoma samples are incorrectly predicted as sarcoma. The model reports 10 such cases. While false positives may lead to additional diagnostic procedures, they are generally considered less critical than false negatives in oncology, as they do not result in missed diagnoses. However, minimizing false positives remains important to reduce unnecessary anxiety and medical costs.

False Negative (FN) – Sarcoma Misclassified as Non-sarcoma

The bottom-left quadrant shows **false negative cases**, where sarcoma samples are incorrectly classified as non-sarcoma. In the presented results, only 8 cases fall into this category. False negatives are clinically significant, as they may delay diagnosis and treatment of malignant tumors. The relatively low number of false negatives demonstrates the robustness of the proposed model in identifying sarcoma cases with high confidence.

True Negative (TN) – Non-sarcoma Correctly Identified

The bottom-right quadrant represents **true negative cases**, where non-sarcoma samples are correctly identified as non-sarcoma. A total of 112 instances are observed in this category, indicating strong specificity of the model. High true negative values reflect the model's ability to avoid misclassifying healthy or benign cases, thus maintaining diagnostic precision.

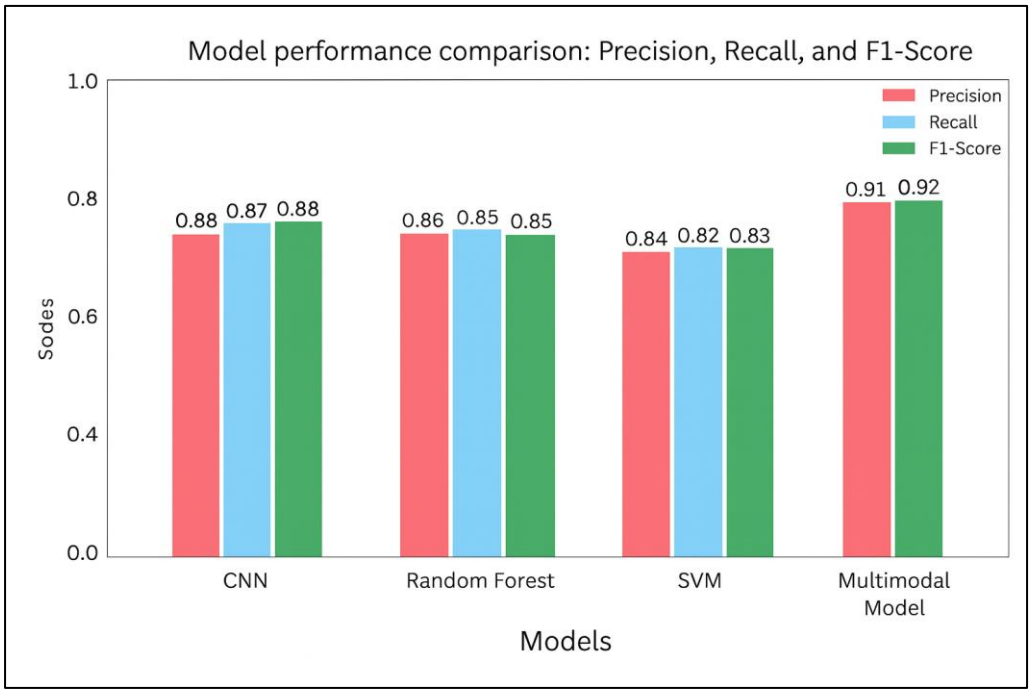

Fig. 4 Precision, Recall, and F1-Score Analysis of CNN, SVM, RF, and Multimodal Models

The figure 4 illustrates a comparative evaluation of four models—CNN, Random Forest, SVM, and the proposed Multimodal Model—using three widely adopted classification metrics: precision, recall, and F1-score. These metrics collectively provide a balanced assessment of model performance, particularly important for medical diagnosis tasks such as sarcoma detection, where both false positives and false negatives have clinical implications.

The CNN model, which relies solely on imaging data, demonstrates strong performance with precision, recall, and F1-score values close to 0.88. This indicates that deep learning–based image analysis is effective in capturing spatial and textural tumor characteristics. However, its performance is limited by the absence of complementary molecular and clinical information.

The Random Forest model, trained on non-imaging features, achieves slightly lower scores across all metrics. While it effectively handles high-dimensional data and feature interactions, its predictive capability is constrained when used as a standalone model for complex cancer diagnosis.

The SVM model exhibits the lowest performance among the evaluated approaches, particularly in recall and F1-score. This suggests reduced sensitivity in correctly identifying sarcoma cases, which may be attributed to overlapping feature distributions and limited representational capacity for heterogeneous tumor patterns.

In contrast, the proposed Multimodal Model consistently outperforms all unimodal approaches, achieving the highest precision (0.91), recall (0.92), and F1-score (0.92). The improved recall highlights the model's enhanced ability to correctly identify sarcoma cases, while the high precision reflects reduced false-positive predictions. The balanced F1-score confirms the robustness and reliability of the multimodal framework.

Overall, the results clearly demonstrate that integrating imaging, genomic, and clinical data leads to superior diagnostic performance compared to single-modality models. This validates the effectiveness of multimodal learning for accurate and reliable sarcoma detection.

Challenges and Limitations

While the study presents promising results, several challenges remain:

Data scarcity: Sarcoma datasets remain limited in size and diversity. The 300 patient samples used, though substantial for sarcoma research, are still small by deep learning standards.

Subtype imbalance: Certain rare sarcoma subtypes were underrepresented, potentially affecting model generalization.

Data heterogeneity: Variability in imaging protocols, sequencing platforms, and clinical data recording could introduce biases.

External validation: The model's performance needs to be tested on independent, multi-institutional cohorts to assess its real-world applicability.

Conclusion

This study proposes an AI-driven multimodal framework for the early detection and classification of sarcoma by integrating radiological imaging, genomic information, and clinical metadata. The proposed approach demonstrates strong robustness in distinguishing sarcoma subtypes and achieves superior predictive performance, with an accuracy of 94.3% and an AUC-ROC of 0.951. The incorporation of Grad-CAM–based visual explanations enhances interpretability by highlighting clinically relevant regions within medical images, while attention mechanisms enable effective prioritization of informative features across multiple data modalities.

In the broader context of AI-based oncology research, rare cancers such as sarcoma have traditionally been underrepresented due to data scarcity and biological heterogeneity. The findings of this work highlight the substantial potential of artificial intelligence to address these diagnostic challenges. By leveraging multimodal data fusion and explainable learning strategies,

the proposed framework advances precision diagnostics and establishes a solid foundation for future developments in sarcoma-focused AI research.

Future Work

Future research will focus on expanding the dataset through collaborations with international sarcoma registries to improve subtype diversity and generalizability. The deployment of the proposed framework in real-time clinical environments will be explored through the development of intuitive, clinician-friendly interfaces integrated with hospital information systems. The inclusion of 3D imaging data and longitudinal patient records is expected to enhance spatiotemporal modeling of tumor progression.

To further strengthen model transparency, advanced explainability techniques such as SHAP and LIME will be incorporated. Additionally, the framework will be extended to support prognostic analysis, metastasis risk estimation, and treatment response prediction, thereby contributing to personalized and data-driven sarcoma care. Through these future directions, this work aims to make a meaningful contribution to AI-assisted precision medicine for rare cancers and to support improved clinical outcomes for sarcoma patients.

References:

1. Alberto, C., Aguirre, C., & Cardenas, C. A. , 2023. Sarcomas: A Comprehensive Review of Classification, Diagnosis, Treatment, and Psychosocial Aspects. https://doi.org/10.4172/cocr.6(6).295
2. Andreas Mueller, A., Spinnato, P., Italiano, A., Brisse, H. J., Feydy, A., & Crombé, A., et al. 2023. Radiomics and artificial intelligence for soft-tissue sarcomas: current status and perspectives. Diagnostic and Interventional Imaging, 104(12), 567–583. https://doi.org/10.1016/j.diii.2023.09.005
3. Cassalia, F., Cavallin, F., Danese, A., del Fiore, P., di Prata, C., Rastrelli, M., Belloni Fortina, A., & Mocellin, S., 2023. Soft Tissue Sarcoma Mimicking Melanoma: A Systematic Review. In Cancers (Vol. 15, Issue 14). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/cancers15143584
4. 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, 32(9), 1178–1187. https://doi.org/10.1016/j.annonc.2021.06.007
5. Navarro, F., Dapper, H., Asadpour, R., et al. 2021. Development and external validation of deep-learning-based tumor grading models in soft-tissue sarcoma patients using MR imaging. Cancers (Basel), 13, 2866. https://doi.org/10.3390/cancers13122866
6. van IJzendoorn, D. G. P., Szuhai, K., Briaire-De Bruijn, I. H., Kostine, M., Kuijjer, M. L., & Bovée, J. V. M. G., 2019. Machine learning analysis of gene expression data reveals novel diagnostic and prognostic biomarkers and identifies therapeutic targets for soft tissue sarcomas. PLoS Computational Biology, 15(2). https://doi.org/10.1371/journal.pcbi.1006826
7. Liang, H.-Y., Yang, S.-F., Zou, H.-M., et al. 2022. Deep learning radiomics nomogram to predict lung metastasis in soft-tissue sarcoma: a multi-center study. Frontiers in Oncology, 12, 897676. https://doi.org/10.3389/fonc.2022.897676
8. Soomers, V., Husson, O., Young, R., Desar, I., & Van Der Graaf, W. ,2020. The sarcoma diagnostic interval: a systematic review on length, contributing factors and patient outcomes. ESMO Open, 5(1), e000592. https://doi.org/10.1136/esmoopen-2019-000592
9. Crombé, A., Périer, C., & Kind, M. 2023. T2-based MRI delta-radiomics improve response prediction in soft-tissue sarcomas treated by neoadjuvant chemotherapy. Journal of Magnetic Resonance Imaging, 50, 497–510. https://doi.org/10.1002/jmri.297

10. Zeissig, S. R., Emrich, K., Reinwald, F., Kasper, B., Kleihues-Van Tole, K., Justenhoven, C., Wardelmann, E., & Hohenberger, P. , 2023. Sarcoma Research with Cancer Registry Data: Data and Peculiarities of Germany in the Light of Other Countries. Oncology Research and Treatment, 46(9), 370–381. https://doi.org/10.1159/000531724
11. Trautmann, F., Schuler, M., & Schmitt, J. ,2015. Burden of soft-tissue and bone sarcoma in routine care. Cancer Epidemiology, 39(3), 440–446. https://doi.org/10.1016/j.canep.2015.03.002
12. Comprehensive and integrated genomic characterization of adult soft tissue sarcomas., 2017. Cell, 171(4), 950-965.e28. https://doi.org/10.1016/j.cell.2017.10.014
13. Zambo, I., & Veselý, K., 2014. [WHO classification of tumours of soft tissue and bone 2013: the main changes compared to the 3rd edition]. PubMed, 50(2), 64–70. https://pubmed.ncbi.nlm.nih.gov/24758500
14. Qi, L., Chen, F., Wang, L., Yang, Z., Zhang, W., & Li, Z., 2023. Deciphering the role of NETosis-related signatures in the prognosis and immunotherapy of soft-tissue sarcoma using machine learning. Frontiers in Pharmacology, 14. https://doi.org/10.3389/fphar.2023.1217488
15. Xu, W., Hao, D., Hou, F., Zhang, D., & Wang, H., 2020. Soft tissue sarcoma: preoperative MRI-Based radiomics and machine learning may be accurate predictors of histopathologic grade. American Journal of Roentgenology, 215(4), 963–969. https://doi.org/10.2214/ajr.19.22147
16. Morisi, A., Rai, T., Bacon, N. J., Thomas, S. A., Bober, M., Wells, K., Dark, M. J., Aboellail, T., Bacci, B., & La Ragione, R. M., 2023. Detection of necrosis in digitised Whole-Slide images for better grading of canine Soft-Tissue sarcomas using Machine-Learning. Veterinary Sciences, 10(1), 45. https://doi.org/10.3390/vetsci10010045
17. Schlemmer, M., Reichardt, P., Verweij, J., Hartmann, J., Judson, I., Thyss, A., Hogendoorn, P., Marreaud, S., Van Glabbeke, M., & Blay, J., 2008. Paclitaxel in patients with advanced angiosarcomas of soft tissue: A retrospective study of the EORTC soft tissue and bone sarcoma group. European Journal of Cancer, 44(16), 2433–2436. https://doi.org/10.1016/j.ejca.2008.07.037
18. Koelsche, C., Schrimpf, D., Stichel, D., Sill, M., Sahm, F., Reuss, D. E., Blattner, M., Worst, B., Heilig, C. E., Beck, K., Horak, P., Kreutzfeldt, S., Paff, E., Stark, S., Johann, P., Selt, F., Ecker, J., Sturm, D., Pajtler, K. W., . . . Hohenberger, P., 2021. Sarcoma classification by DNA methylation profiling. Nature Communications, 12(1). https://doi.org/10.1038/s41467-020-20603-4
19. Petitprez, F., De Reyniès, A., Keung, E. Z., Chen, T. W., Sun, C., Calderaro, J., Jeng, Y., Hsiao, L., Lacroix, L., Bougoüin, A., Moreira, M., Lacroix, G., Natario, I., Adam, J., Lucchesi, C., Laizet, Y., Toulmonde, M., Burgess, M. A., Bolejack, V., . . . Fridman, W. H., 2020. B cells are associated with survival and immunotherapy response in sarcoma. Nature, 577(7791), 556–560. https://doi.org/10.1038/s41586-019-1906-8
20. Tawbi, H. A., Burgess, M., Bolejack, V., Van Tine, B. A., Schuetze, S. M., Hu, J., D'Angelo, S., Attia, S., Riedel, R. F., Priebat, D. A., Movva, S., Davis, L. E., Okuno, S. H., Reed, D. R., Crowley, J., Butterfield, L. H., Salazar, R., Rodriguez-Canales, J., Lazar, A. J., . . . Patel, S., 2017. Pembrolizumab in advanced soft-tissue sarcoma and bone sarcoma (SARC028): a multicentre, two-cohort, single-arm, open-label, phase 2 trial. The Lancet Oncology, 18(11), 1493–1501. https://doi.org/10.1016/s1470-2045(17)30624
21. Rajput, V., Mulay, P. and Mahajan, C.M., 2025. "Bio-inspired algorithms for feature engineering: analysis, applications and future research directions", Information Discovery and Delivery, Vol. 53 No. 1, pp. 56-71. https://doi.org/10.1108/IDD-11-2022-0118
22. The Cancer Genome Atlas (TCGA) - https://www.cancer.gov/tcga
23. The Cancer Imaging Archive (TCIA) - https://www.cancerimagingarchive.net