Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 12 March 2026
Revised 26 April 2026
Accepted 18 May 2026

Natural Resources for Human Health 2026, Vol. 6 (1s): 542–557

KEYWORDS:
ConvNeXt V2, Swin Transformer, Feature Extraction, Attention Mechanism, Transfer Learning, Whole slide Imaging, Classification, Precision, Recall, F1-Score, ROC Curve.

Multi-Scale Deep Feature Fusion for Soft Tissue Sarcoma Subtype Classification From Histopathological Images

Vineela Madireddy[1]*, Dr. HariKrishna Bommala [2]

[1]Research Scholar, Department of Computer Science and Engineering, Bharatiya Engineering Science And Technology Innovation University, Gorantla, Anantapur, AP, 515231, India.
Assistant Professor, Department of Computer Science and Engineering(AI&ML), Vignan Institute of Technology and Science, Nalgonda,Telangana,508284 India.
vineela.madireddy@gmail.com

[2] Professor, Department of Computer Science and Engineering, KG Reddy College of Engineering and Technology, Moinabad, Ranga Reddy, India,501504.
haribommala@gmail.com

ABSTRACT: Soft tissue sarcoma is a heterogeneous and a rare malignant tumour that is formed in connective tissues such as fat, muscle and fibrous tissue. For an efficient treatment planning accurate subtype classification is very essential. Traditional diagnostics rely on histopathological works which is time consuming and also subject to inter-observer variability. This research proposes a novel deep learning hybrid transformer architecture combined with EfficientNet V2B0, ConvNeXt V2(Tiny) and Light weight Swin Transformer for automated multi class STS subtype classification based on histopathological images which classifies 10 different sub types of Soft Tissue Sarcoma. Fine Grained textures are captured by EfficientnetV2B0, Local spatial features are extracted using the model ConvNeXt V2 and Global contextual dependencies are captured by using Swin Transformer. The proposed hybrid model demonstrates high precision of 0.91, accuracy 0.88, recall 0.88 and f1-score is of 0.88. High precision will always indicate that the model predicates very few false positive predictions. Effectiveness of deep learning methods helps in assisting pathologists and improves diagnostic reliability.

1. INTRODUCTION

Soft Tissue Sarcoma is a rare heterogeneous group of malignant tumours that is formed in mesenchymal tissues such as fat, fibrous tissue, blood vessels, peripheral nerves and muscles. As it is less than 1% of all adult cancers and due to its complexity and diversity soft tissue sarcoma represent a significant clinical challenge. More than 50 histological subtypes of soft tissue sarcoma have been identified, each exhibiting different genetic, morphological and clinical characteristics. For an efficient prognosis estimation, treatment planning and personalized therapy accurate subtype classification of soft tissue sarcoma is very crucial. Leiomyosarcoma, Liposarcoma and synovial sarcoma are the subtypes that vary significantly in terms of metastatic potential aggressiveness and response to therapy. Due to the intra-class variability and because of many overlapping histopathological features distinguishing various subtypes is a challenging task. Traditional subtype classification methods are considered as time-consuming, labour-intensive and subject to inter observer variability among pathologists. Initial STS subtype classification is considered as the gold standard relies on histopathological examination of haematoxylin and eosin-stained tissue slides, supplemented by immunohistochemistry and molecular diagnostic techniques are time consuming. In case of rare and poorly differentiated tumours subtle differences in cellular morphology and tissue architecture can lead to misclassification.

Advances in AI, ML and DL have identified various techniques for automated medical image analysis. Extracting hierarchical features from histopathological images, enables accurate classification of various cancer subtypes using CNN have demonstrated extremely efficient results. Recent transformer-based architectures such as Swin Transformers and Vision Transformers are used to extract global contextual information and to capture long range dependencies in images enhances classification performance.

In spite of various advancements in Artificial Intelligence, Machine Learning and Deep Learning methods standalone CNN or transformer models have various limitations. As CNN primarily focus on local spatial features and fail require large datasets and high computational resources. To address these challenges, hybrid models that combine EfficentNetV2B0, ConvNeXt V2 with transformer-based model (Swin Transformer) have emerged as a promising approach. The proposed hybrid models leverage the strengths of all methods, enabling robust feature extraction and improved classification accuracy.

The proposed model focuses on the development of hybrid deep learning framework for subtype classification of soft tissue sarcoma using histopathological images. The hybrid approach integrates EfficientV2B0 captures fine grained morphological features, ConvNeXt V2(tiny) extracts local feature extraction, and Swin Transformer captures global context modelling, followed by feature fusion and classification. The aim of the proposed model mainly focuses on ten different types of STS subtype classification namely Adipose Tissue, Angiosarcoma, clear cell sarcoma, Dermatofibrosarcoma, protuberans, Spindle Cell Lipoma, Liposarcoma, Fibrous Histiocytoma, Fibrosarcoma, Epithelioid sarcoma, Desmoplastic small round cell tumor. And also to improve diagnostic accuracy, assist human pathologists and reduce human error in clinical decision making.

Accurate classification of soft tissue sarcoma remains a challenging task due to similarity of substantial morphological among various tumour subtypes, class imbalance and large intra-class variations and scarcity of high quality annotated histopathological datasets. To capture complex tissue characteristics were failed by existing machine learning methods as these rely heavily on handcrafted features. These limitations are the reason for the requirement of the robust classification of the hybrid model capable of learning local and global features. The proposed model integrates feature extraction mechanisms. The manuscript is organized as follows. **Section 1.1** focuses on related literature on soft tissue sarcoma classification using deep learning and machine learning techniques. **Section 2** represents the proposed methodology hybrid deep learning framework, including the dataset description, data preprocessing, model architecture, training strategy, and evaluation metrics. **Section 3** discusses the experimental setup and presents the results of performance metrics along with a comparative analysis against existing approaches and provides a detailed discussion of the findings, highlighting the strengths, limitations, and clinical implications of the proposed method. Finally, **Section 4** concludes the paper by summarizing the key contributions and suggesting directions for future research.

1.1 Related Work

In spite of many advances in artificial intelligence for histopathological image analysis several challenges remain in the automated classification. Most studies focus on binary tumour classification or broad tumour categorization, whereas accurate discrimination among multiple STS subtypes remains underexplored. Early studies mainly focused on gene expression profiling when combined with classical machine learning algorithms such as Support Vector Machine demonstrates efficient results in distinguishing STS subtype. Machine Learning models such as Random Forest, K-Nearest Neighbours and Navie Bayes are used to implement classification achieved a moderate accuracy approximately up to 80% on MRI datasets. Accurate subtype classification, prediction and tumour grading have been successfully applied to whole-slide images. CNN have significantly improved classification performance by automatically extracting discriminative features from histopathological and radiological images. ResNet and EfficientNet has enhanced performance in limited dataset scenarios has enhanced performance. The hybrid models proposed in this research are used to captured both local and global contextual features have shown superior performance compared to standalone models.

Schmitz et al.,[1] is focused on classifying the tumor as low-grade or high-grade STS using ADC values, MRI radiomics features, conventional semantic imaging features with various machine learning algorithms. The model was developed for tumor grading before surgery or biopsy.

Madireddy, V., et al.,[2] has focused on to develop accurate AI-based system for STS classification as benign or a malignant tumour. A system was developed to improve diagnostics precision compared to standalone models. Model was used to assist pathologists and clinicians in automated diagnosis.

Voigtlander et al[3] has proposed Automated detection and classification of extremity STS. Using CNN based image classification has improved diagnostic efficiency and demonstrated CNNs adaptability of CNNs for clinical sarcoma workflows.

Schmitz et al[4] has proposed Tumor grade classification using machine and deep learning models Random Forest, SVM, KNN and SMOTE+RFE piplines. The hybrid models Randon forest, SMOTE and recursive Feature Elimination performed best with AUC ≈ 0.87. performance has significantly improved by combining radiomics, ADC, and semantic.

Maussion et al[5] has proposed on federated deep learning for prognosis classification using genomics, histopathology and clinical data for multi-centre STS analysis.

Gunarsson et al[6] has focused on classifying MRI based grading and classification by using radiomics and machine learning by combining imaging biomarkers with machine learning classifiers and also has improved accuracy of tumor grading from MRI scans.

He et al[7] has presented in his study using advanced deep learning methods ResNet50 has predicted 2-year survival with AUROC ≈0.937 and DenseNet121 achieved AUROC ≈ 0.944 for subtype classification.

Thiesen et al[8] has focused on subtype classification of pediatric STS using the advanced deep learning frame work with the help of digital histopathology images gathered from multiple medical centres. The research focused on improving diagnostic accuracy and generalizability in rare sarcomas.

Michot et al[9] has focused on DL models predicted metastatic relapse-free survival. High-risk patients were more accurately identified by AI than conventional grading.

Di,J., et al[10] has focused on classifying the tumor as benign or malignant evaluated applying SVM, Logistic Regression and decision tree and have achieved high accuracy, improved sensitivity for detecting malignant tumors.

Xu et al[11] has proposed in their study about Deep learning models to analyze WSIs to detect tor regions, to classify sarcoma subtype and to predict tumor grade. Deep Learning have achieved high accuracy comparable to expert pathologist.

V.Madireddy et.al[12] has compared between various machine and deep learning algorithms in their survey and has proposed early diagnosis using deep learning can improve treatment planning and survival rate and hybrid models provides better diagnosis performance.

Liu et. al[13] has proposed in their study that focused on to develop a model using deep learning radiomics to predict the pathological response of patients with oesophageal squamous cell carcinoma to develop a multi model to predict the pathological virus after neoadjuvant chemoradiotherapy.

Peeken et. al[14] has specified in the comparison between two imaging-based approaches for predicting outcomes in patients with high grade STS using semantic image analysis and radiomics.

He s et. al[15] has focused in his research about the DL models that has improved subtype recognition and automated diagnosis workflows on US clinical histopathology data has AUC of 0.89

Frankel et. al[16] has proposed a machine learning model achieved 80-85% accuracy in classifying MRI-based tissue classification using SVM, Random Forest, KNN, Navie Byes algorithms.

Wallner et.al[17] has proposed a model has classified 60+ sarcoma subtypes with high accuracy using molecular data implementing machine learning algorithms.

Foersch et.al[18] has successfully predicted molecular subtypes and clinically relevant characteristics using histopathology slides.

Arin Rosennberg et.al[19]. has proposed in their research for automated subtype detection on histopathology images using deep learning Convolution Neural Network has achieved high accuracy.

Hermessi et. al [20]. has proposed a model using CNN-based Radiomics and Transfer Learning by extracting deep features from MRI significantly improved classification accuracy of Soft Tissue Sarcoma with accuracy of 85%.

2. MATERIALS AND METHODS

A hybrid deep learning framework is the proposed methodology for automated classification of Soft Tissue Sarcoma subtypes using histopathology images. The local spatial features and global contextual features from microscopic tissue images are effectively captured by the proposed architecture by integrating CNN and transformer-based learning. The proposed system workflow is illustrated in the architecture diagram in Figure 1.

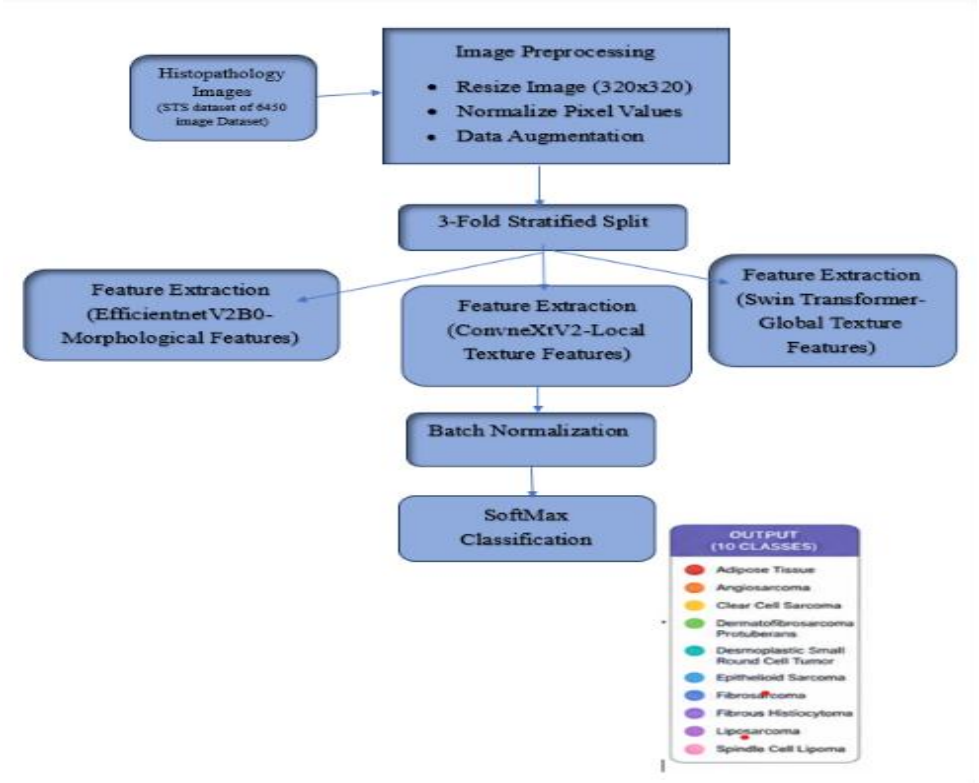


Figure 1: Architecture of the Proposed System

2.1 Histopathology Image Acquisition

The input dataset is collected from Soft Tissue Biopsies - v 1 No Augmentations > Images consisting of histopathology clinically validated STS dataset. A total of 6450 microscopic images with dimensions of 320x320 are used for classification are normalized using imageNet normalization statistics. Data augmentation techniques including random horizontal, and vertical flipping, random rotation, zooming and brightness adjustment were applied only to training images to reduce overfitting and to improve robustness. Validation and testing images were used without augmentation. Image level class annotations were provided in Open sarcoma dataset but does not include patient identifiers, slide identifiers, scanner information or other clinical meta-data. Therefore, patient-level or slide-level independence between the training, validation and testing subsets cannot be independently verified. In the proposed study the predefined dataset partitions were maintained without modification, and the test set remained completely unseen during hyperparameter optimization and model training. This limitation should be considered when interpreting the reported classification performance. The dataset contains multiple soft tissue sarcoma subtypes including Adipose Tissue, Angiosarcoma, Clear Cell Sarcoma, Dermatofibrosarcoma Protuberans, Desmoplastic Small

Round Cell Tumor, Epithelioid Sarcoma, Fibrosarcoma, Fibrous Histiocytoma, Liposarcoma, Spindle Cell Lipoma and other 42 different classes from which the proposed hybrid model classifies the above specified 10 different classes where the total 6450 images are divided into 71.6% Training Data(4618), 20.1% Testing Data(1297) and 8.3% are used for validation(535).

2.2 Image Preprocessing

The extracted images are preprocessed to improve the model generalization and robustness capability prior to feature extraction. The preprocessing steps comprises of

- Image resizing to 320x320 pixels.
- Pixel Intensity normalization.
- Noise reduction and enhancement.
- Data augmentation techniques.
- Data Augmentation including Rotation, Horizontal and vertical flipping, Zooming, Random cropping.

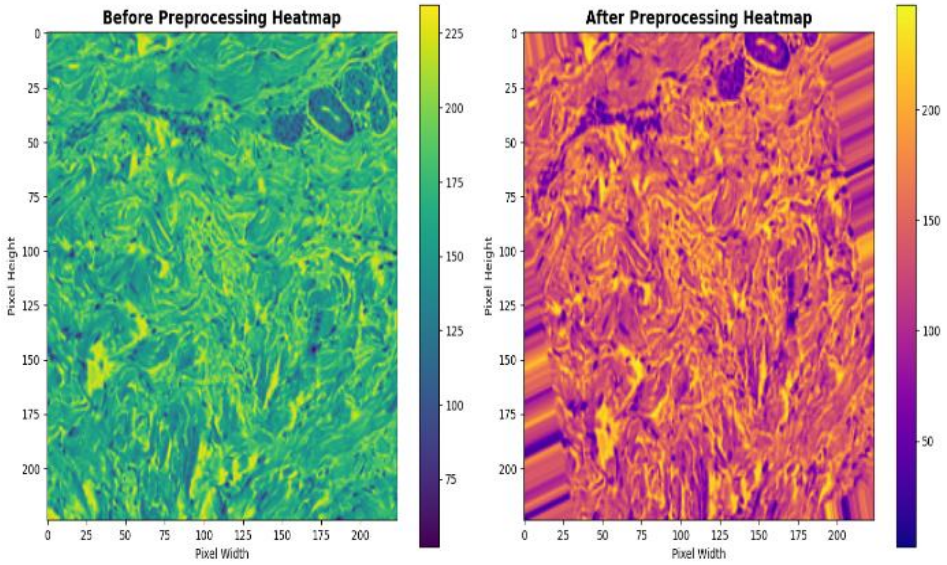


Figure2: Heatmap before and after preprocessing

The left side of Figure 2. depicts heat map before preprocessing. The original dataset contains uneven intesity distributions, certain cellular structures are not clearly highlighted and throughout the pixel intensity values vary irregular. After preprocessing onto the right side the tissue structures become more prominent and sharper, contrast between the regions is improved significantly. Pixel intesity becomes more uniform.

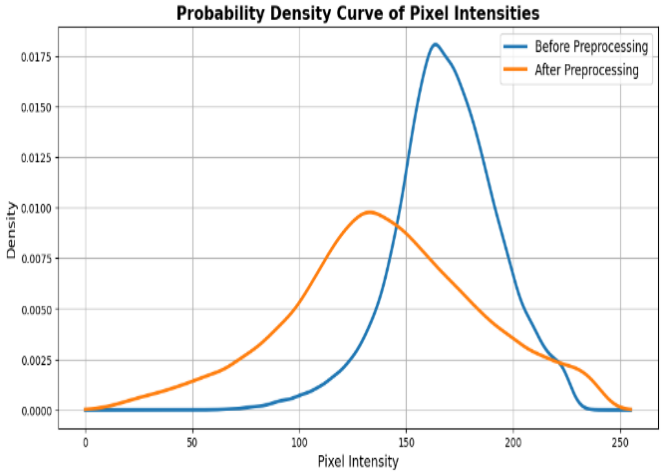


Figure3: Probability Density Curve

Figure 3 depicts the Probability Density Curve that higher peaks indicates that many pixels share similar intensity values. Probability Density Curve indicates the pixel distribution is uneven and highly concentrated. The narrow distribution indicates that actual dataset lacks balanced intensity representation which may reduce feature extraction capability. After preprocessing broader and smoother distribution represents enhanced contrast, reduced noise and improved tissue visibility and makes the dataset more suitable for subtype classification. Boxplot represents the outliers. In Figure 4 many outliers appear in lower intensity which represents staining, noise and artifacts after preprocessing outliers shift towards lower intensity and distribution becomes smoother and more controlled and preserves important tissue structures.

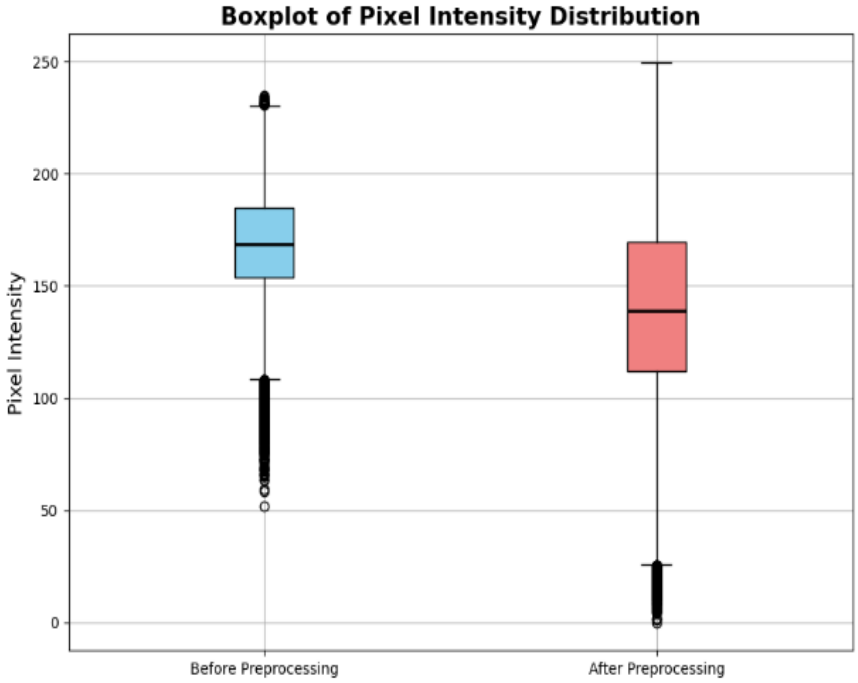


Figure 4: Boxplot of pixel Intensity Distribution

The boxplot demonstrates preprocessing that significantly improves the statistical distribution of pixel intensity.

2.3 Deep Feature Extraction

The proposed model extracts discriminative features from histopathology images using multiple deep learning architectures. The various feature extractors are CNN, Efficient V2B0, Reset-50, DenseNet-121, ConvNextV2(Tiny), Swin Transformer. Because of their superior representation learning capability, the hybrid model Efficient V2B0, ConvNext V2 and Swin Transformer are utilized as the primary backbone networks.

2.3.1 EfficientnetV2B0

The model captures discriminative morphological patterns such as tissue texture, nuclear a typia, cellular arrangement, stromal characteristics. Through compound scaling and progressive feature extraction the model captures subtle differences among various STS subtypes. The convolutional operation is represented as

$$Y = f(W * X + b) \qquad (1)$$

- X=input image feature map
- W=convolutional kernel weights
- b=bias term
- f=activation function
- y=output feature map

Swich activation Function is represented as:

$$f(x) = x.\sigma(x) \qquad (2)$$

2.3.2 ConvNeXt V2

Fine-grained local texture information and morphological characteristics of tumour cells are effectively captured through convolutional operations. Depthwise convolutions, Feature scaling and normalization are used to improve representation learning. Local spatial features are extracted from histopathology images using:

$$Y(i,j) = \sum_m \sum_n X(i+m, j+n).K(m,n) \qquad (3)$$

Where:

- X=input feature map
- K=convolution kernel
- Y=output feature map
- i,j=spatial coordinates

2.3.2.1 Layer Normalization

To stabilize the training data convNext-V2 uses normalization. Feature consistency is improved by normalization.

$$\hat{x} = \frac{x - \mu}{\sqrt{\sigma^2 + \epsilon}} \qquad (4)$$

Where

- X=input feature
- μ=mean
- σ^2=variance
- €=small constant for stability

2.3.2.2 GELU Activation Function

ConvNeXt-V2 uses GELU activation function to enhance non linear function learning.

$$GELU(x) = x\phi(x) \quad (5)$$

- $\phi(x)$=cumulative distribution of Gaussian function

2.3.2.3 Global Response Normalization

Global Response Normalization is used by convNeXt V2 for feature competition. GRN improves representation learning and feature diversity.

$$y_i = \gamma \frac{x_i}{\sqrt{\sum_j x_j^2 + \epsilon}} + \beta + x_i \quad (6)$$

where

- x_i=activation function
- γ,β= learnable parameters

2.3.3 Swin Transformer

To understand better tissue structures and spatial relationships global contextual dependencies are extracted by swin Transformer using shifted window self attention mechanisms.

2.3.3.1 Self Attention Mechanism

Self Attention Mechanism learns relationship between tissue regions.

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (7)$$

Where

- Q=Query Matrix
- K=Key Matrix
- V=Value Matrix
- d_k=dimensionality matrix

2.3.3.2 Widow-Based Multi -Head Self Attention

To reduce computational complexity Attention inside local windows is computed by swin Transformer.

$$W - MSA(X) = Concat(head_1, \ldots . head_h)W^0 \quad (8)$$

Where

- $head_h$= individual attention heads
- W^O= output projection matrix

2.3.3.3 Shifted Window Mechanism

SWA enables cross-window information exchange.

$$\widehat{x^l} = SW - MSA\left(LN(X^{l-1})\right) + X^{l-1} \quad (9)$$

Where

- SW-MSA=shifted window multi-head self-attention
- LN= Layer Normalization

2.3.3.4

Multi- Layer Perceptron(MLP)

MLP refines learned representation. The feed- forward network is:

$$MLP(X) = W_2\sigma(W_1x + b_1) + b_2 \quad (10)$$

Where

- W_1,W_2=weight matrices
- B_1,b_2=biases
- σ =activation function

2.4 Hybrid Feature Fusion

The Hybrid Feature Fusion is the combination of ConvNext-V2+Swin Transformer. The hybrid model improves subtype classification performance. Feature Fusion is expressed as:

$$Feature_{fusion} = Concat(F_{EfficientNetV2B0,}F_{ConvNeXt,}F_{swin}) \quad (11)$$

Where:

- $F_{ConvNeXt}$= local Feature representation
- F_{swin}=Global Contextual representation

2.5. Softmax Classification

Class with highest probability is predicted as STS subtype. The final classification probability is computed using softmax:

$$P(y_i) = \frac{e^{z_i}}{\sum_{j=1}^{c} e^{z_j}} \quad (12)$$

Where

z_i=class score

C=Total number of classes

2.6. Proposed Algorithm

Proposed Hybrid Algorithm(EfficientNetV2Bo+ConvNeXtV2(Tiny)+Swin Transformer) for soft tissue sarcoma classification

Output: Predicted sarcoma subtype

1. Collect and preprocess the images.
2. Resize the images and perform data augmentation.
3. Apply 3-fold stratified cross validation to preserve class distribution across all 10 subtypes.
4. For each fold use two folds for training and one-fold for validation/testing.
5. Extract morphological features using EfficientNetV2B0.
6. Extract local texture features using ConvNeXtV2-Tiny.
7. Extract global features using swin transformer.
8. Concatenate the extracted feature vectors to obtain a hybrid feature representation.
9. Apply batch Normalization and dropout for regularization.
10. Pass fused features through fully connected layers.
11. Generate subtype probabilities using softmax activation function.

12. Compute categorical cross entropy loss and update model parameters using Adam optimizer.
13. Evaluate performance using Accuracy, precision, Recall, F1-score, ROC_Curve, confusion Matrix
14. Output the final predicted subtypes among the 10 soft Tissue sarcoma Classes.
15.

END

3. RESULTS AND DISCUSSION

The proposed model was implemented in Googlecolab and trained on GPU's in python3.X. The deep learning framework was developed using pyTorch/TensorFlow. The implementation used the libraries numpy, pandas, matplotib, scikit-learn, OpenCV, Torchvision. The performance of various deep learning baseline models including CNN, EfficientNetV2B0, ResNet-50, DenseNet-121, and the proposed hybrid fusion model EfficientNetV2B0-ConvNextV2-Swin Transformer were trained using the same predefined training, validation and testing partitions, with identical input image resolution(320x320) of soft tissue sarcoma using histopathology images, AdamW optimizer(learning rate=$1x10^{-4}$, weight decay=$1x10^{-5}$), batch size of 4, and a maximum 15 epochs. The learning rate was adaptively reduced using the ReduceLROnPlateau scheduler factor=0.5, patience=3), while early stopping(patience=5) restored the best performing model. The learning rate was adaptively reduced using the ReduceLROnPlateau scheduler (factor = 0.5, patience = 3), while Early Stopping (patience = 5) restored the best-performing model. Balanced class weights were computed from the training set weights were computed from training set and incorporated into loss function to mitigate class imbalance. Accuracy, Precision, Recall, F1-Score and ROC-AUC score were used to assess the performance metrics.

3.1 CNN Performance Analysis

The CNN model has demonstrated very less classification accuracy of 0.65. Low-level spatial features and tissue textures were highly sensitive to staining variations and complex tissue morphology are responsible for increased misclassification among similar subtypes.

3.2 Efficient Net V2B0

To achieve high classification accuracy while reducing training time and computational complexity an advanced CNN architecture Efficient Net V2B0 is designed. To extract both local and global histopathological features from whole slide or patch- based microscopic images. To capture both It combines both MBConv and Fused MBConv blocks to efficiently extract both local and global histopathological features.

3.3 ResNet-50 Performance Analysis

By utilizing residual learning and deeper feature extraction mechanisms ResNet-50 there is a slight variation in classification performance. The model has achieved 0.70 of accuracy. The model was unable to focus on sufficient global contextual features.

3.4 DensNet-121 performance Analysis

Improved feature reuse and dense feature propagation has improved the accuracy. DensNet-121 has achieved an accuracy of 0.72. Tumor The model was not able to capture global spatial dependencies. Tumor structures and stromal tissues were effectively discriminated by the model when compared with CNN and ResNet-50.

3.5 ConvNeXtV2 performance Analysis

Cellular textures and Local morphological structures are effectively captured by the model and accuracy is increased to 0.87. The model has enhanced feature representation and learning efficiency.

3.6. Swin Transformer Analysis

The model has implemented using shifted window self-attention mechanism enabling effective learning of global contextual relationships between tissue regions. The model has achieved an accuracy of 0.84. The model was not able to extract extremely local texture details and required higher computational resources.

3.7. Hybrid Model(Convolutional Feature Extraction+ Swin Transformer)

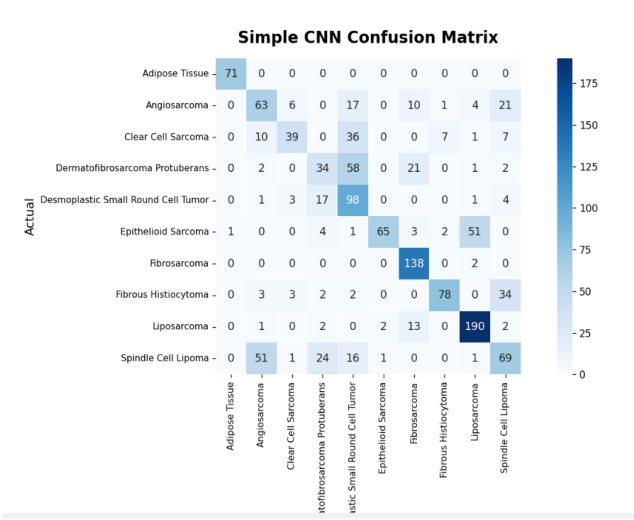
The hybrid model was developed by integrating convolutional feature extraction with swin Transformer-inspired attention mechanisms. To improve generalization and classification robustness data augmentation, class-weight balancing, batch

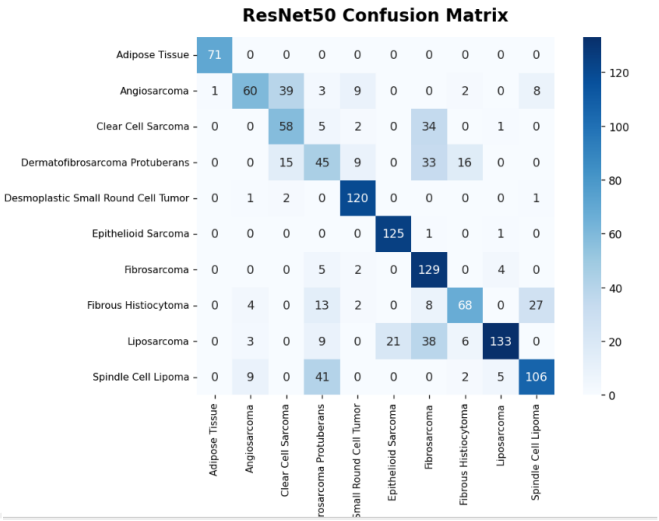
normalization, dropout regularization and AdamW optimization techniques were incorporated. When compared with single CNN the accuracy was increased and with swin Transformer the performance was reduced and accuracy was 0.77

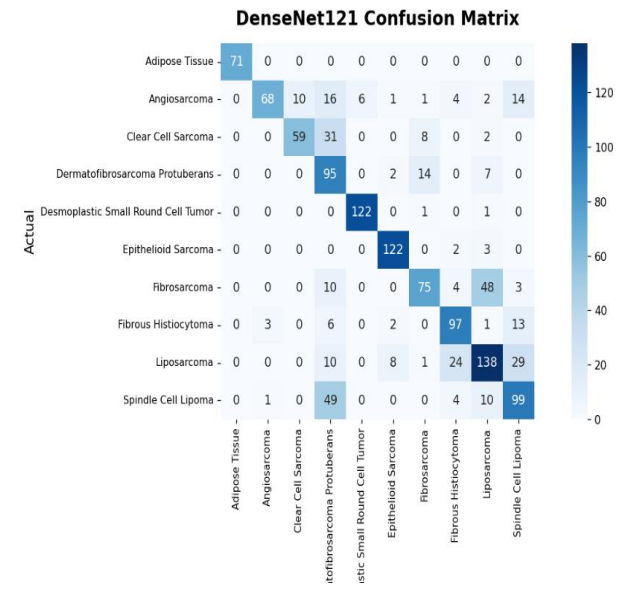
3.8 Proposed hybrid Fusion Model EfficientV2B0+ConvNeXtV2(Tiny)+Swin Transformer

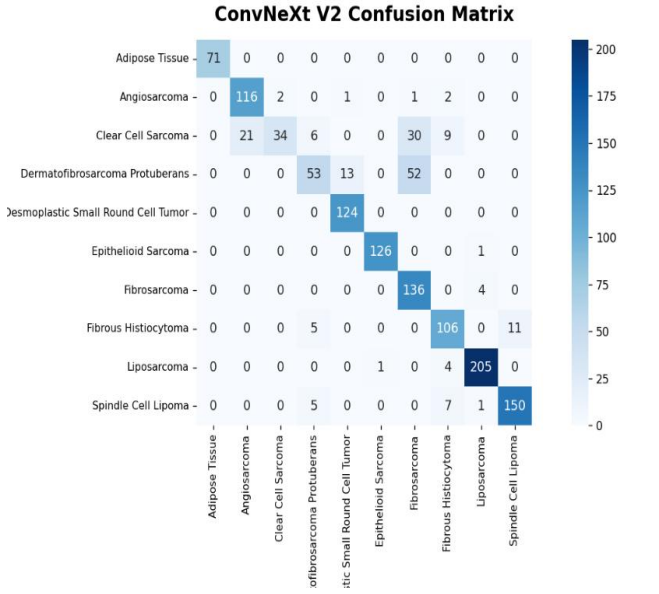
The hybrid fusion model has achieved superior performance when compared to all the other CNN models. Local spatial features were extracted from ConvNeXtV2 and Global contextual features were extracted by the swin transformer hybrid fusion model. The hybrid model has used 3-Fold stratified cross validation by splitting the dataset into 3 folds. Compute average predictions and calculate final metrics. When compared with all the models the proposed hybrid fusion model has increased the precision to 0.91 indicating that the model is making highly reliable predictions when assigning a histopathological subtype. The fusion model has minimized false-positive predictions which helps in efficient treatment planning techniques and avoids unnecessary interventions. The model has improved classification robustness and subtype discrimination. The hybrid model has extracted complex morphological patterns, tissue heterogeneity and intra class variability and achieved highest precision.

3.9 Confusion Matrix









Natural Resources for Human Health
eISSN: 2583-1194

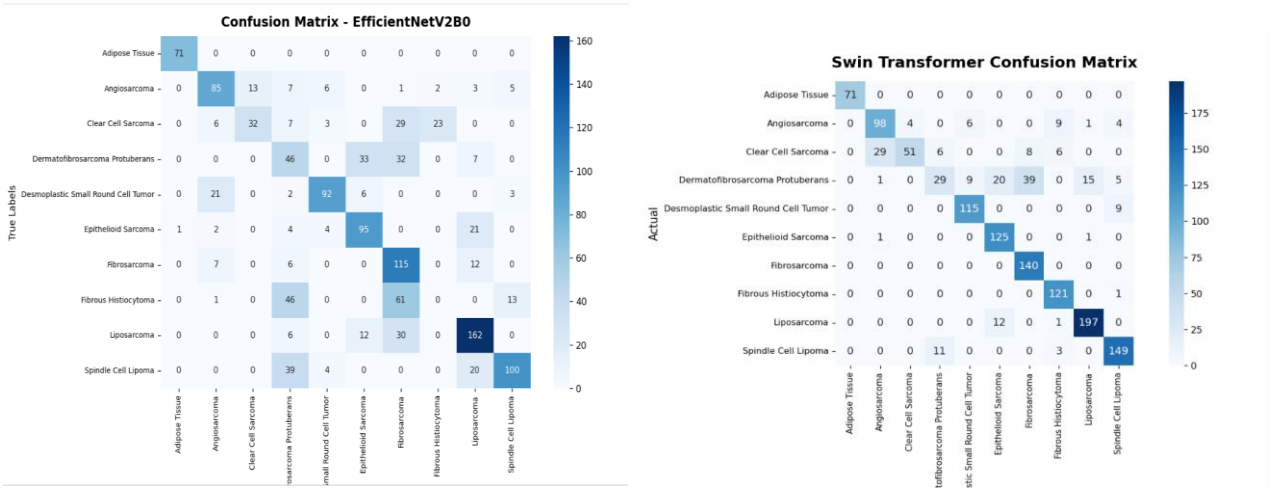


Figure 5: Confusion Matrix of CNN, ResNet-50, EfficietnetV2B0, DensNet-121, ConvNeXtV2, Swin Transformer

The confusion matrix is depicted in Figure 5. The confusion matrix of various models are compared with the proposed hybrid model. The proposed hybrid model has achieved high positive rates, minimal false classifications and improved sensitivity.

3.10 ROC Curve Analysis

The Roc Curve Analysis of various CNN and ML models are depicted in Figure 6. The proposed hybrid model achieved the highest Area Under Curve hybrid model has high sensitivity and specificity.

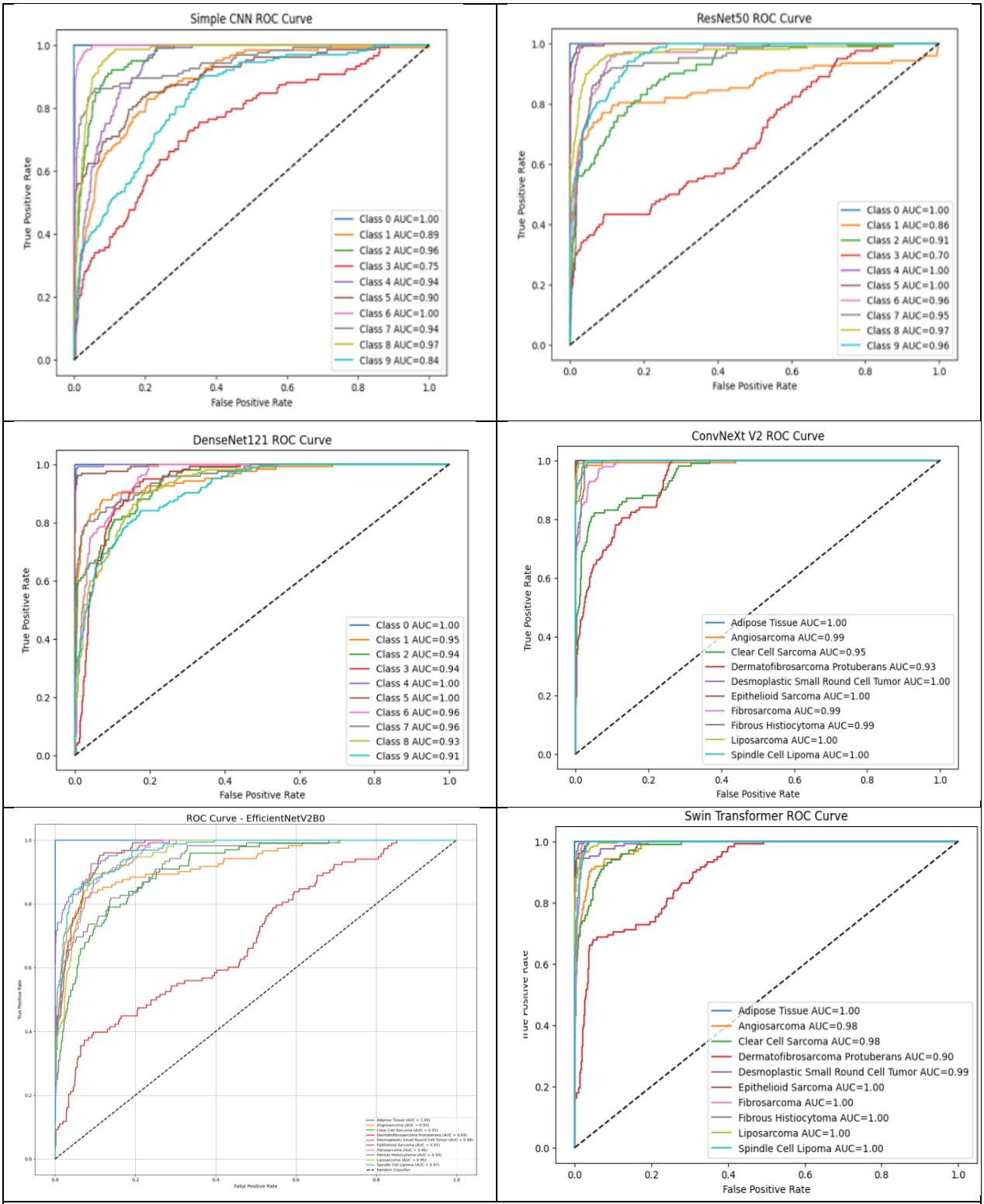


Figure 6: RoC Curve of CNN, ResNet, Efficient net V2B0,DensNet-121, ConvNeXtV2, Swin Transformer

across all soft Tissue Sarcoma subtypes. The hybrid model has effectively differentiated morphologically similar tumor classes with high reliability and misclassification was reduced.

3.11 Comparison Analysis

To identify the best model of all the available models the comparison between ResNet-50, Densnet-121, ConvNeXt V2, Swin Transformer models are compared with the hybrid model and the comparison is represented in Table 1. Comparison among various algorithms represented on a bargraph shown in Figure 7.

Table 1: Comparison Analysis

Models	Accuracy	Precision	Recall	F1
CNN	0.65	0.67	0.64	0.67
EfficientNet-V2	0.66	0.69	0.66	0.66
ResNet-50	0.70	0.73	0.70	0.70
DensNet-121	0.72	0.76	0.72	0.72
ConvNeXt- V2	0.87	0.86	0.86	0.87
Swin Transformer	0.84	0.84	0.84	0.82
Hybrid Model (EfficientNetv2 +ConNeXtV2+ Swin Transformer)	0.88	0.91	0.88	0.88

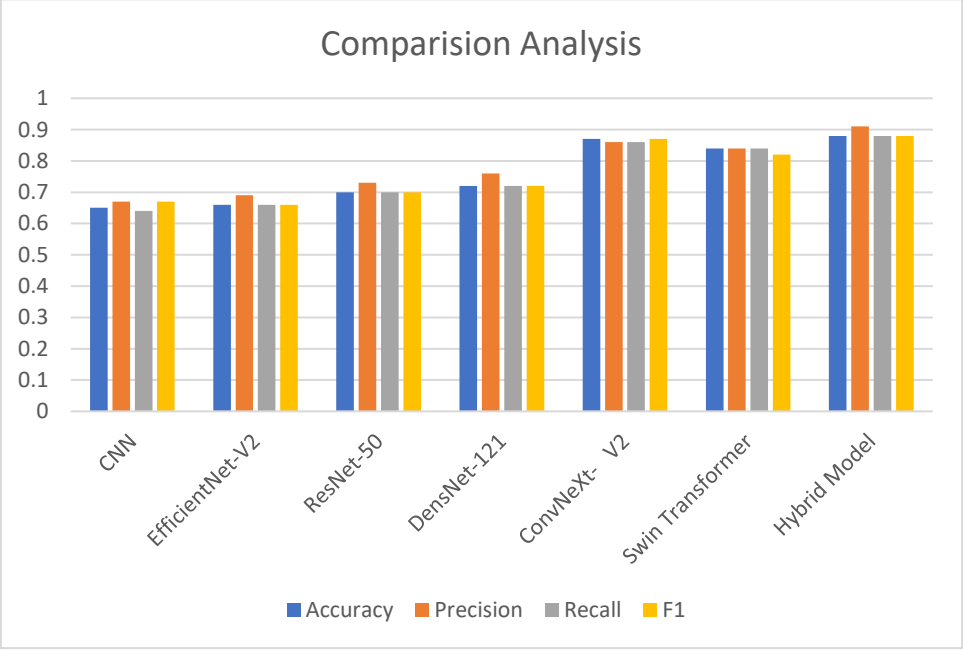


Figure 7: Comparison among various Models on a Bargraph

4. CONCLUSION

To accurately classify the subtypes of soft tissue Sarcoma and to improve the image quality and feature visibility. Morphological features were captured by efficientNetV2B0, Local texture features were extracted by convneXtV2 and long range contextual dependencies were captured by Swin Transformer using shifted Window Self-Attention mechanisms. The hybrid features were extracted using SoftMax based classification and Global Average Pooling. The proposed framework can be considered a precision-oriented diagnostic model that prioritizes prediction reliability while maintaining competitive overall classification performance.

4.1 Future Scope

The proposed hybrid model using EfficientNetV2B0, ConvNeXtV2 and Swin Transformer has provided automated model for subtypes classification of soft tissue sarcoma. Further improvements can be explored on predictions on disease prognosis and the treatment planning techniques suitable for patients.

ACKNOWLEDGEMENTS

The authors thank all individuals who contributed to this research through their valuable support and guidance.

AUTHOR CONTRIBUTIONS

All authors contributed to the conception and design of the study. Data collection, analysis, interpretation of results, and manuscript preparation were carried out collaboratively. All authors reviewed, revised, and approved the final version of the manuscript.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

ETHICS STATEMENT

Ethical approval was not required for this study because it did not involve human participants, animals, or any procedures requiring institutional ethical review.

DATA AVAILABILITY STATEMENT

The datasets analysed during this study are publicly available from the respective data repositories cited in the manuscript.

REFERENCES

[1] Schmitz, F., Strauss, D., Gengenbacher, J., Rehnitz, C., Schlemmer, H. P., Jang, H., & Sedaghat, S. (2026). Performance Comparison of Machine Learning Models for Predicting Tumor Grade in Soft-Tissue Sarcoma using Radiomics. *European Journal of Radiology Artificial Intelligence*, 100077. https://doi.org/10.1016/j.ejrai.2026.100077

[2] Madireddy, V., & Bommala, H. (2025, September). Accurate Classification of Soft Tissue Sarcoma using Hybrid Models. In *2025 6th International Conference on Smart Electronics and Communication (ICOSEC)* (pp. 2048-2054). IEEE. https://doi.org/10.1109/ICOSEC67334.2025.11459573

[3] Voigtländer, H., Schmitz, F., Strauss, D., Kauczor, H. U., Voigtländer, S., Sauerwein, S., & Sedaghat, S. (2026). Automated detection of primary soft tissue sarcomas of the extremities using artificial intelligence and ChatGPT. *Frontiers in Oncology*, *16*, 1674509. https://doi.org/10.3389/fonc.2026.1674509

[4] Schmitz, F., Strauss, D., Gengenbacher, J., Rehnitz, C., Schlemmer, H. P., Jang, H., & Sedaghat, S. (2026). Performance Comparison of Machine Learning Models for Predicting Tumor Grade in Soft-Tissue Sarcoma using Radiomics. *European Journal of Radiology Artificial Intelligence*, 100077. https://doi.org/10.1016/j.ejrai.2026.100077

[5] Maussion, C., Coindre, J. M., Blay, J. Y., Schutte, K., Camara, A., Loarer, F. L., ... & 6.Karanian, M. (2025). Multimodal prediction of metastatic relapse using federated deep learning in soft-tissue sarcoma with a complex genomic profile. *Scientific Reports*, *15*(1), 36588. https://doi.org/10.1038/s41598-025-20495-8

[6] Gunnarsson, A. E., Correra, S., Sánchez, C. T., Recenti, M., Jónsson Jr, H., & Gargiulo, P. (2025). Toward New Assessment in Sarcoma Identification and Grading Using Artificial Intelligence Techniques. *Diagnostics*, *15*(13), 1694. **https://doi.org/10.3390/diagnostics15131694**

[7] He, S., Chen, J., Liu, Y., Xiao, Y., Li, M., Zhang, Q., ... & Xiao, J. (2025). Deep learning algorithms from histopathological images stratify molecular subtypes for leiomyosarcoma: a proof-and-concept diagnostic study. *International Journal of Surgery*, *111*(9), 5959-5969. https://doi.org/10.1097/JS9.0000000000002667

[8] Thiesen, A., Domanskyi, S., Zhang, J., Sheridan, T. B., Neuhauser, S. B., Stetson, A., ... & Rubinstein, J. C. (2025). Multicenter Histology Image Integration and Multiscale Deep Learning for Machine Learning-Enabled Pediatric Sarcoma Classification. *medRxiv*. https://doi.org/10.1101/2025.06.10.25328700

[9] Michot, A., Le, V. L., Coindre, J. M., Velasco, V., Soussi, M., Mesli, N., ... & Crombé, A. (2025). Prognostic prediction in soft-tissue sarcomas using deep learning and digital pathology of tumor and margin areas. *Scientific Reports*, *15*(1), 38534. https://doi.org/10.1038/s41598-025-20804-1

[10] Di, J., Hickey, C., Bumgardner, C., Yousif, M., Zapata, M., Bocklage, T., ... & Qasem, S. A. (2024). Utility of artificial intelligence in a binary classification of soft tissue tumors. *Journal of Pathology Informatics*, *15*, 100368. https://doi.org/10.1016/j.jpi.2024.100368

[11] Xu, R., Tang, J., Li, C., Wang, H., Li, L., He, Y., ... & Li, Z. (2024). Deep learning-based artificial intelligence for assisting diagnosis, assessment and treatment in soft tissue sarcomas. *Meta-Radiology*, *2*(2), 100069. https://doi.org/10.1016/j.metrad.2024.100069

[12] Madireddy, V., Bommala, H., & Yerraboina, S. (2024). Soft tissue sarcoma diagnosis using machine and deep learning-survey. In *MATEC Web of Conferences* (Vol. 392, p. 01138). EDP Sciences. https://doi.org/10.1051/matecconf/202439201138

[13] Liu, Y., Wang, Y., Hu, X., Wang, X., Xue, L., Pang, Q., ... & Hui, Z. (2024). Multimodality deep learning radiomics predicts pathological response after neoadjuvant chemoradiotherapy for oesophageal squamous cell carcinoma. https://doi.org/10.1186/s13244-024-01851-0

[14] Peeken, J. C., Neumann, J., Asadpour, R., Leonhardt, Y., Moreira, J. R., Hippe, D. S., ... & Combs, S. E. (2021). Prognostic assessment in high-grade soft-tissue sarcoma patients: a comparison of semantic image analysis and radiomics. *Cancers*, *13*(8), 1929. **https://doi.org/10.3390/cancers13081929**

[15] He, S., Chen, J., Liu, Y., Xiao, Y., Li, M., Zhang, Q., ... & Xiao, J. (2025). Deep learning algorithms from histopathological images stratify molecular subtypes for leiomyosarcoma: a proof-and-concept diagnostic study. *International Journal of Surgery*, *111*(9), 5959-5969. https://doi.org/10.1097/JS9.0000000000002667

[16] Frankel, A. O., Lathara, M., Shaw, C. Y., Wogmon, O., Jackson, J. M., Clark, M. M., ... & Keller, C. (2022). Machine learning for rhabdomyosarcoma histopathology. *Modern Pathology*, *35*(9), 1193-1203. https://doi.org/10.1038/s41379-022-01075-x

[17] Wallner, C., Alam, M., Drysch, M., Wagner, J. M., Sogorski, A., Dadras, M., ... & Behr, B. (2021). A highly reliable convolutional neural network based soft tissue sarcoma metastasis detection from chest X-ray Images: a retrospective cohort study. *Cancers*, *13*(19), 4961. **https://doi.org/10.3390/cancers13194961**

[18] Foersch, S., Eckstein, M., Wagner, D. C., Gach, F., Woerl, A. C., Geiger, J., ... & 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

[19] Rosenberg, A., Kochanny, S., Dolezal, J., Wang, C., Chen, H., Kather, J. N., & Pearson, A. T. (2020). Prediction of histologic and molecular subsets of soft tissue sarcoma using deep learning. DOI: 10.1200/JCO.2020.38.15_suppl.e23529

[20] Hermessi, H., Mourali, O., & Zagrouba, E. (2019). Deep feature learning for soft tissue sarcoma classification in MR images via transfer learning. *Expert Systems with Applications*, *120*, 116-127. 10.1016/j.eswa.2018.11.025