Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Evaluating the Efficacy of Deep Learning Models for Distinguishing Spinal Hemangiomas from Common Metastatic Tumors in Across MRI Sequences

Serkan Bayram

dr.serkanbayram89@gmail.com

Istanbul University

Hakan Yılmaz

Konya Technical University

Ahmet Yasin Karkaş

Arnavutköy Devlet Hastanesi

Furkan Okatar

Istanbul University

Taha Bedir Demir

Istanbul University

Turgut Akgül

Istanbul University

Research Article

Keywords: spine metastasis, hemangioma, deep learning, T1 sequence

Posted Date: April 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9125445/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Purpose

This study aimed to develop and evaluate a deep learning approach for distinguishing spinal hemangiomas from thoracolumbar spinal metastases (breast, lung, prostate, renal cell carcinoma, and multiple myeloma) across MRI sequences and assess sequence-dependent sensitivity.

Method

Radiological data from a university hospital (2016–2023) were collected from 352 adults (mean age 61.3 years): 109 hemangiomas and 243 metastases (histopathologically confirmed). Lesions were labeled and regions of interest annotated using VGG Image Annotator, cropped, converted to grayscale, and resized to 256×256 pixels across four sequences (T1, T2, STIR, contrast-enhanced). A unified CNN (three convolutional layers with 32/64/128 filters, max pooling, batch normalization, 128-neuron fully connected layer, dropout 0.5, softmax output) was trained with downsampling-based class balancing and a 60/40 stratified train–test split (Adam, categorical cross-entropy, 40 epochs, batch size 16), with hyperparameter optimization over learning rate, batch size, epochs, and dropout. Performance varied markedly by sequence and comparison.

Results

T1 and T2 yielded the most robust discrimination; the best result occurred on T1 for hemangioma versus multiple myeloma (accuracy 0.89, AUC 0.93). Across all T2 datasets, performance was balanced and high (accuracy 0.84, sensitivity 0.93, specificity 0.75, AUC 0.87). In contrast, STIR and contrast-enhanced sequences showed reduced and inconsistent performance (overall STIR accuracy 0.57, AUC 0.54; contrast-enhanced accuracy 0.68, AUC 0.72), including occasional near-random classification and pronounced sensitivity–specificity imbalances.

Conclusion

Deep learning-based differentiation of hemangiomas from spinal metastases is feasible using conventional MRI, with T1/T2 sequences providing reliable features, while STIR and post-contrast imaging may need larger datasets and advanced architectures to improve robustness.

1. Introduction

Back pain is a prevalent phenomenon that affects a significant proportion of the global population. It accounts for approximately one-third of all daily outpatient visits and ranks second only to the common cold and upper respiratory infections[1]. Additionally, it is the third most common indication for surgical procedures and the fifth most common cause of hospitalization[2]. In the majority of low back pain cases, imaging is deemed necessary. Magnetic resonance imaging (MRI) is typically the preferred modality because of the absence of radiation exposure, superior soft tissue contrast resolution, and sensitivity to bone marrow abnormalities[3]. Intraosseous hemangiomas are incidentally identified during magnetic resonance imaging (MRI) scans of the spine. It has been documented that the incidence of haemangioma in spinal MRI is 27%[4].

Vertebral hemangiomas are typically readily discernible by radiologists and clinicians on magnetic resonance imaging (MRI) scans. However, atypical hemangiomas, which exhibit distinct signal patterns, can present as metastatic lesions, particularly in patients with a history of malignancy[5]. This may necessitate repeated MRI examinations, the utilization of advanced

imaging techniques (such as scintigraphy and positron emission tomography (PET)), and even biopsy, which can induce significant anxiety in patients.

In the present era, in addition to sophisticated examinations, evaluation can be conducted with the aid of artificial intelligence. This study aimed to utilize a deep learning-based algorithm to differentiate the diagnosis of spinal hemangioma from that of the most common spinal metastatic tumors (including breast, lung, and prostate cancers, multiple myeloma, and renal cell cancer) using conventional MRI and to investigate the sensitivity of the algorithm.

2. Materials and Method

A retrospective collection of radiological data was performed for patients diagnosed with spinal lesions at a university hospital between 2016 and 2023. All patients were required to provide informed consent prior to their participation in the clinical study as part of the process leading up to their radiological evaluation and subsequent admission to the outpatient clinic. The following inclusion criteria were adopted: Inclusion criteria were as follows: (1) age above 18 years, (2) presence of at least one bone metastasis lesion in the thoracolumbar spine, (3) diagnosis of primary cancer (lung, renal cell, multiple myeloma, breast and prostate cancers) in a patient with spinal metastases, (4) availability of complete radiological data including contrast-enhanced MRI, and (5) diagnosis of spinal metastasis or spinal hemangioma by both a radiologist and an orthopedist. Patients who had undergone MRI at least twice at separate times to diagnose hemangioma and for whom there had been no alteration in the recorded images were selected for this study. Patients diagnosed with metastatic spinal disease, who constituted the sample population of this study, were histopathologically diagnosed. MRIs of hemangiomas were obtained from patients without a pathological diagnosis of hemangioma.

The initial cohort comprised patients diagnosed with spinal hemangioma via magnetic resonance imaging (MRI) by an orthopaedic specialist and a radiologist with expertise in the musculoskeletal system. All images were reviewed and labeled by an experienced radiologist as either hemangioma or non-hemangioma.

The dataset consisted of MRI images acquired using four different imaging sequences: contrast-enhanced, T1-weighted, T2-weighted, and STIR sequences. To improve the generalizability and robustness of the proposed convolutional neural network (CNN) model, images were obtained from multiple anatomical regions and under various pathological conditions.

Hemangioma images constituted the positive class, whereas the non-hemangioma class included images from various organs and pathologies, namely the lungs, breast, prostate, renal cell carcinoma (RCC), and multiple myeloma. Images were available for all four MRI sequences for each anatomical region and pathology.

All images were stored in NumPy (. npy) format and resized to a uniform input dimension before being fed into the CNN model. The numerical distribution of images across categories and MRI sequences is summarized in Table-1.

The VGG Image Annotator (VIA), a platform-independent JavaScript application, was used to label the images and mark the regions of interest. The regions marked by the experts were cropped, and all images were resized to a resolution of 256 × 256 pixels. This process formed the dataset used to develop the deep learning model. A visual sample of the dataset is presented in Figure-1.

Programming Environment

The dataset was created, and preprocessing tasks on the images were performed using a computer with an Intel i7 processor and 16 GB of RAM, alongside the Python programming language.

Convolutional Neural Network (CNN)

Convolutional Neural Networks (CNNs) have proven to be a highly effective architecture within deep learning, particularly excelling in image analysis and classification tasks. A CNN is generally structured with multiple layers, each fulfilling a

unique role that enhances the overall performance of the network. Gaining an understanding of these layers is essential to comprehend how CNNs function and their applications across various domains, such as medical imaging, facial recognition, and object detection[6]

The core component of a CNN is the "convolutional layer." This layer applies convolution operations to the input data using a series of trainable filters. Each filter is crafted to identify particular features, such as edges or textures, by moving across the input image and calculating the dot products. The convolutional layer greatly reduces the number of parameters in the model compared to fully connected layers, as the same filter is applied to various regions of the input, taking advantage of weight sharing[7, 8].

After the convolutional layer, CNNs usually incorporate a “pooling layer.” Pooling is used to downsample the feature maps produced by the convolutional layers, reducing their dimensionality and computational complexity while preserving key information. The two most common forms of pooling are max pooling and average pooling, which involve selecting either the maximum or average value from a group of values in the feature map, respectively[8].

The Batch Normalization (BN) layer is another important component used in CNNs. In deep learning architectures, each layer acts as the input for the next, meaning that the learning of one layer must be completed before the next layer begins. While normalization is generally applied to standardize input values at the start, the intermediate layers do not benefit from this early normalization. Consequently, problems such as slower training, instability, or gradient vanishing, where learning progresses very slowly, can occur. BN was introduced to address these issues and enhance training efficiency and stability[6].

Following several convolutional, pooling, and batch normalization layers, the architecture moves into “fully connected layers.” These layers resemble those in standard neural networks, in which each neuron is linked to every neuron in the preceding layer. The primary function of the fully connected layers is to analyze the features extracted by the convolutional and pooling layers, ultimately generating the final output, such as class probabilities in classification tasks[9].

In addition to the primary layers, CNNs often include “non-linear activation functions,” such as Rectified Linear Units (ReLU) (Eq. 1) and softmax (Eq. 2), applied after each convolutional layer. These activation functions introduce non-linearity to the model, enabling it to capture and learn more complex patterns in the data [9]. In addition, dropout layers are frequently used to enhance training stability and prevent overfitting. By randomly deactivating a subset of neurons during training, dropout helps the model generalize better and reduces the risk of overfitting the training data[10].

$$ReLU\,(x) = \max(0, x)\ ^{6}$$

$$\sigma\,(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} or = 1,2,\ldots,K\ ^{7}$$

Metrics Used

In deep learning, especially for classification tasks, evaluating model performance using various metrics is crucial. These metrics provide valuable insights into the effectiveness of the model and enable comparisons between different models. The key performance indicators included accuracy, sensitivity (recall), specificity, F1 score, and AUC.

3. Results

A total of 352 patients were included in the study, comprising 187 men (53.1%) and 165 women (46.9%), with a mean age of 61.3 years. The hemangioma cohort had a mean age of 55.6 years and consisted of 63 men and 46 women (n = 109). The lung cancer cohort had a mean age of 60.2 years, with 6 males and 23 females (n = 29). The breast cancer cohort had a mean age of 53.6 years, including 1 male and 40 females (n = 41). The multiple myeloma cohort had a mean age of 64.2 years, with 34 males and 36 females (n = 70). The renal cell carcinoma cohort had a mean age of 61.6 years and comprised

16 men and 20 women (n = 36). The prostate cancer cohort, consisting entirely of males, had a mean age of 72.6 years (n = 67**) (Table-2)**.

A unified convolutional neural network (CNN) architecture was employed across all experimental datasets. The same model configuration was used without modifications to ensure a consistent comparison across the MRI sequences and dataset combinations. Each dataset corresponding to a specific MRI sequence and diagnostic category was independently processed. The CNN model was trained and evaluated using identical preprocessing steps, class balancing, and training protocols for each dataset. This resulted in multiple experimental runs equal to the prepared datasets, with the performance metrics calculated separately. To investigate robustness, hyperparameter optimization experiments were performed to explore variations in the learning rate, batch size, epochs, and dropout rate. The impact on the classification performance was analyzed, and the best-performing configurations were used in the final CNN model. The CNN architecture consisted of three convolutional layers with 32, 64, and 128 filters, followed by max-pooling layers, a fully connected layer with 128 neurons, and a dropout layer with a rate of 0.5. The output layer used softmax activation to classify the images into hemangioma and non-hemangioma classes. The experiments used grayscale images resized to 256 × 256 pixels. The deep learning network is illustrated in Fig. 2.

A downsampling strategy was used to address class imbalance before model training, ensuring equal samples from each class. The balanced datasets were divided into training and test sets using a 60%–40% stratified split. The model was trained for 40 epochs with a batch size of 16 using the Adam optimizer and categorical cross-entropy loss function. The performance was evaluated using a held-out test set. For each dataset, the accuracy, sensitivity, specificity, F1-score, and area under the ROC curve (AUC) were calculated. The quantitative results are summarized in Table-3. The CNN model's performance was evaluated across datasets representing different MRI sequences (T1-weighted, T2-weighted, STIR, and contrast-enhanced) and diagnostic comparisons between hemangiomas and non-hemangioma conditions using accuracy, sensitivity, specificity, F1-score, and AUC metrics.

The highest classification performance was observed for the T1- and T2-weighted MRI sequences. In T1-weighted images, the model achieved an accuracy of 0.89 and an AUC of 0.93 in the hemangioma versus multiple myeloma comparison, representing the best performance. High performance was also observed for hemangiomas versus RCC (AUC = 0.90) and hemangiomas versus breast lesions (accuracy = 0.85). The performance across all T1-weighted datasets showed an accuracy of 0.75 and an AUC of 0.79. T2-weighted images showed strong results across comparisons. The hemangioma versus breast and hemangioma versus lung datasets achieved accuracies of 0.88 and 0.87, respectively, with AUC values exceeding 0.88. Across all T2-weighted datasets, the model achieved an accuracy of 0.84, sensitivity of 0.93, and AUC of 0.87, indicating robust performance (Fig. 3–6). STIR and contrast-enhanced sequences showed reduced and variable performances. While some STIR comparisons, such as hemangioma vs. lung, yielded high sensitivity (0.92) and acceptable AUC values (0.82), the overall STIR performance was limited, with an accuracy of 0.57 and an AUC of 0.54. Some STIR datasets showed extreme imbalances between sensitivity and specificity, with sensitivity decreasing to 0.00 despite perfect specificity. The contrast-enhanced datasets produced inconsistent classification results. Although some comparisons achieved high sensitivity, this often came with low specificity, yielding a modest overall performance. The general contrast-enhanced dataset achieved an accuracy of 0.68 and an AUC of 0.72, while certain comparisons, such as hemangioma vs. RCC, showed an AUC below 0.50, indicating near-random classification.

4. Discussion

The discriminative performance of the proposed CNN model depends on both the MRI sequence and the diagnostic comparisons. T1- and T2-weighted sequences exhibited superior performance, offering stable morphological characteristics for automated hemangioma classification. The high AUC values in the T1- and T2-based comparisons indicate the effective capture of spatial and intensity patterns associated with hemangiomas. Deep learning algorithms have demonstrated strong performance in detecting spinal metastases across various tasks and imaging modalities. Motohashi et al. developed an AI model for detecting osteolytic spinal metastases using 475 CT scans from patients with and without metastases[11]. The

model showed sensitivity, accuracy, and F1-scores of 0.68, 0.68, and 0.72 per slice and 0.75, 0.36, and 0.48 per lesion. Studies indicate that the AI model improved residents' sensitivity while matching specialists' performance. The model shows promise for clinical practice, despite requiring accuracy improvements. Fan et al. evaluated AdaBoost and Chan-Vese algorithms using MRI images from 87 lung cancer patients with spinal metastases[12]. AdaBoost achieved 96.55% accuracy, whereas CV showed superior segmentation with a Dice index of 0.8591. Liu et al.'s study of 514 patients[13] found the expert-based features model outperformed the deep learning model in accuracy (88% vs. 69%) and AUC (0.80 vs. 0.71), particularly for lung and kidney metastases. The DL model exhibited better recall for rare sites. In our study of 352 patients' MRI scans, the model showed superior classification in T1- and T2-weighted sequences, with T1 being optimal for differentiating hemangiomas from multiple myelomas. T2 sequences showed consistently high accuracy, whereas STIR and contrast-enhanced sequences performed poorly, with imbalanced sensitivity and specificity.

Hemangiomas are the most prevalent spinal lesions and are often misdiagnosed as metastatic disease. Er et al. used the YOLO-v8 deep learning model to identify vertebral hemangiomas[14], achieving 73% accuracy in T1- and T2-weighted MRI. This lower accuracy was due to the limited training data and signal intensity variations. Hemangiomas present diagnostic challenges owing to their similarity to metastases and compression fractures. Yeh et al. developed a decision support system using ResNet50 to distinguish between benign and malignant spinal fractures[15]. The system, tested on 140 benign and 50 malignant cases, improved diagnostic performance, particularly for less-experienced clinicians, with one resident's sensitivity increasing from 78% to 94%. In our study, T1 and T2 sequences showed the highest accuracy (0.75–0.89) and AUC values for differentiating hemangiomas from tumors, while STIR and contrast-enhanced images performed worse. T1/T2 sequences achieved high sensitivity in distinguishing hemangiomas from multiple myeloma and breast cancer, although performance decreased for prostate cancer and RCC. The T2 sequence exhibited the most balanced performance overall, with an accuracy, 0.932 sensitivity, 0.750 specificity, and AUC of 0.841, 0.932, 0.750, and 0.865 AUC. Although high sensitivity indicates effective hemangioma detection, some cases showed reduced specificity, suggesting false-positive predictions.

The results demonstrate that the discriminative performance of the CNN model depends on the MRI sequence and diagnostic comparison. T1- and T2-weighted sequences showed superior performance, providing stable morphological features for the automated classification of hemangiomas. High AUC values in T1- and T2-based comparisons indicate effective capture of spatial and intensity patterns associated with hemangiomas. Although high sensitivity values suggest effective hemangioma detection, some cases showed reduced specificity with false-positive predictions. STIR and contrast-enhanced sequences showed reduced performances. STIR's fat signal suppression and emphasis on edema-related changes in STIR may increase the visual similarity between hemangiomas and other conditions, explaining the sensitivity and specificity instability. The heterogeneous enhancement patterns of contrast-enhanced images vary with lesion type, vascularity, and timing, challenging CNN feature learning. Despite hyperparameter optimization, the STIR and contrast-enhanced dataset performances remained below the T1- and T2-weighted results, suggesting limitations beyond the training parameters. The contributing factors include the reduced dataset size from class balancing, shallow CNN architecture insufficient for complex feature capture, and single-sequence inputs limiting cross-modality information. Future improvements may involve larger datasets, advanced augmentation, deeper CNN architectures, multimodal fusion, transfer learning, and attention-based mechanisms to enhance the classification performance and robustness.

5. Conclusion

This study employed artificial intelligence analysis of magnetic resonance images to differentiate between the five most prevalent types of cancer that metastasize to the spine and the most common benign lesion, hemangiomas. In this study, T1- and T2-weighted sequences demonstrated superior discriminatory performance, with the T1 sequence exhibiting a particular efficacy. The T1 sequence achieved an accuracy rate of 89% and an AUC of 0.93 in differentiating between hemangioma and multiple myeloma, while also producing robust results for renal cell carcinoma and breast metastases. In

contrast, STIR and contrast-enhanced sequences exhibited significantly lower performance (STIR AUC 0.54; contrast-enhanced AUC 0.72) and, in several instances, yielded results similar to random classification.

Declarations

Conflict of interests:

The authors declare no competing interests.

Ethics Approval:

This study was approved by the Institutional Ethics Committee of our hospital and conducted in accordance with the ethical standards of the 1964 Declaration of Helsinki and its later amendments.

Informed Consent:

Due to the retrospective design of the study, the requirement for individual informed consent was waived by the Institutional Ethics Committee.

Author Contribution

S.B. and H.Y. drafted the paper, H.Y., T.B.D, F.O. and A.Y.K. carried out acquisition of data, S.B. and H.Y.writen the paper, T.A. conceived of the study, T.A. and S.B. carried out critical revision of the paper. All authors read and approved the final manuscript.

Acknowledgement

None

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable

References

1. Ma K, Zhuang ZG, Wang L, Liu XG, Lu LJ, Yang XQ, Lu Y, Fu ZJ, Song T, Huang D, Liu H, Huang YQ, Peng BG, Liu YQ (2019) The Chinese Association for the Study of Pain (CASP): Consensus on the Assessment and Management of Chronic Nonspecific Low Back Pain. Pain Res Manag. 2019:8957847.
2. Andersson GB (1999) Epidemiological features of chronic low-back pain. Lancet 354(9178):581-5.
3. *Teh J, Iman A, Watts C (2005) Imaging of back pain. Imaging. 17:171–207.*
4. Barzin M, Maleki I (2009) Incidence of vertebral hemangioma on spinal magnetic resonance imaging in Northern Iran. Pakistan J Biol Sci 12:542–4.
5. Matrawy K.A., El-Nekeidy A.A., Gaber El-Sheridy H (2013) Atypical hemangioma and malignant lesions of spine: can diffusion weighted magnetic resonance imaging help to differentiate? Egypt J Radiol Nucl Med. 44:259e63.
6. Varlı,M, Yılmaz H (2023) Multiple classification of EEG signals and epileptic seizure diagnosis with combined deep learning. Journal of Computational Science 67: 101943.
7. Liu J, Zhao H (2021) Application of convolution neural network in medical image processing. Technol Health Care. 29(2):407-17.
8. Purwono, P., Ma'arif, A., Rahmaniar, W., Fathurrahman, H., Frisky, A., & Haq, Q (2023) Understanding of Convolutional Neural Network (CNN): A Review. International Journal of Robotics and Control Systems,,2(4):739-48.

9. Yang W, Luo W, Zhang Y (2020) Classification of odontocete echolocation clicks using convolutional neural network. J Acoust Soc Am.147(1):49.
10. Zhang X, Li J, Wu W, Dong F, Wan S (2023) Multi-Fault Classification and Diagnosis of Rolling Bearing Based on Improved Convolution Neural Network. Entropy (Basel) 25(5):737.
11. Motohashi M, Funauchi Y, Adachi T, Fujioka T, Otaka N, Kamiko Y, Okada T, Tateishi U, Okawa A, Yoshii T, Sato S (2024) A New Deep Learning Algorithm for Detecting Spinal Metastases on Computed Tomography Images. Spine (Phila Pa 1976) 49(6):390-7.
12. Fan X, Zhang X, Zhang Z, Jiang Y (2021) Deep Learning on MRI Images for Diagnosis of Lung Cancer Spinal Bone Metastasis. Contrast Media Mol Imaging. 2021:5294379.
13. Liu K, Ning J, Qin S, Xu J, Hao D, Lang N (2025) Identifying Primary Sites of Spinal Metastases: Expert-Derived Features vs. ResNet50 Model Using Nonenhanced MRI. J Magn Reson Imaging. 62(1):176-86.
14. Er H, Tören M, Asan B, Kaba E, Beyazal M (2025) Deep Learning-Based Differentiation of Vertebral Body Lesions on Magnetic Resonance Imaging. Diagnostics (Basel). 15(15):1862.
15. Yeh LR, Zhang Y, Chen JH, Liu YL, Wang AC, Yang JY, Yeh WC, Cheng CS, Chen LK, Su MY (2022) A deep learning-based method for the diagnosis of vertebral fractures on spine MRI: retrospective training and validation of ResNet. Eur Spine J. 31(8):2022-30.

Tables

Table-1: Numerical distribution of the dataset across MRI sequences and diagnostic categories.

Category / Diagnosis	**T1**	**T2**	**Contrast**	**STIR**	**Total**
Hemangioma	109	109	109	109	436
Lung cancer	29	29	29	29	116
Breast cancer	41	41	41	41	164
Multiple Myeloma	70	70	70	70	280
Prostate cancer	67	67	67	67	268
Renal Cell Carcinoma (RCC)	36	36	36	36	144
Total	**352**	**352**	**352**	**352**	**1408**

Table-2: Distribution of age and gender between the groups.

	Mean age ± SD	Female/Male
Hemangioma group	55,6 ± 13	63/46
Lung cancer group	60,2 ± 9,3	6/23
Breast cancer group	53,6 ± 10,1	1/40
Multiple Myeloma group	64,2 ± 10,6	34/36
Renal cell cancer group	61,6 ± 11,4	16/20
Prostate Group	72,6 ± 8,4	0/67

Table-3: Performance metrics of the CNN model across MRI sequences and datasets.

Dataset Comparison	MRI Sequence	Accuracy	Sensitivity	Specificity	F1 Score	AUC
Hemangioma vs. Multiple Myeloma	T1	0.893	0.964	0.821	0.900	0.929
Hemangioma vs. Breast	T2	0.879	0.875	0.882	0.875	0.912
Hemangioma vs. Lung	T2	0.875	0.833	0.917	0.870	0.882
Hemangioma vs. Breast	T1	0.848	1.000	0.706	0.865	0.794
All Sets	**T2**	**0.841**	**0.932**	**0.750**	**0.854**	**0.865**
Hemangioma vs. RCC	T1	0.828	0.929	0.733	0.839	0.905
Hemangioma vs. Prostate	T1	0.815	0.741	0.889	0.800	0.866
Hemangioma vs. Multiple Myeloma	T2	0.804	0.714	0.893	0.784	0.883
Hemangioma vs. Prostate	T2	0.796	0.926	0.667	0.820	0.855
All Sets	**T1**	**0.750**	**0.773**	**0.727**	**0.756**	**0.786**
Hemangioma vs. Lung	T1	0.750	0.917	0.583	0.786	0.764
Hemangioma vs. RCC	T2	0.724	0.714	0.733	0.714	0.729
Hemangioma vs. Lung	STIR	0.708	0.917	0.500	0.759	0.819
Hemangioma vs. Prostate	STIR	0.704	0.778	0.630	0.724	0.701
All Sets	**Contrast-enhanced**	**0.682**	**0.682**	**0.682**	**0.682**	**0.717**
Hemangioma vs. Prostate	Contrast-enhanced	0.667	1.000	0.333	0.750	0.705
Hemangioma vs. Multiple Myeloma	Contrast-enhanced	0.643	0.714	0.571	0.667	0.626
Hemangioma vs. Lung	Contrast-enhanced	0.625	0.333	0.917	0.471	0.590
Hemangioma vs. RCC	Contrast-enhanced	0.621	0.214	1.000	0.353	0.490
Hemangioma vs. Multiple Myeloma	STIR	0.607	0.964	0.250	0.711	0.536
Hemangioma vs. Breast	Contrast-enhanced	0.606	0.750	0.471	0.649	0.596
All Sets	**STIR**	**0.568**	**0.705**	**0.432**	**0.620**	**0.537**
Hemangioma vs. RCC	STIR	0.552	0.429	0.667	0.480	0.467
Hemangioma vs. Breast	STIR	0.515	0.000	1.000	0.000	0.482

Figures

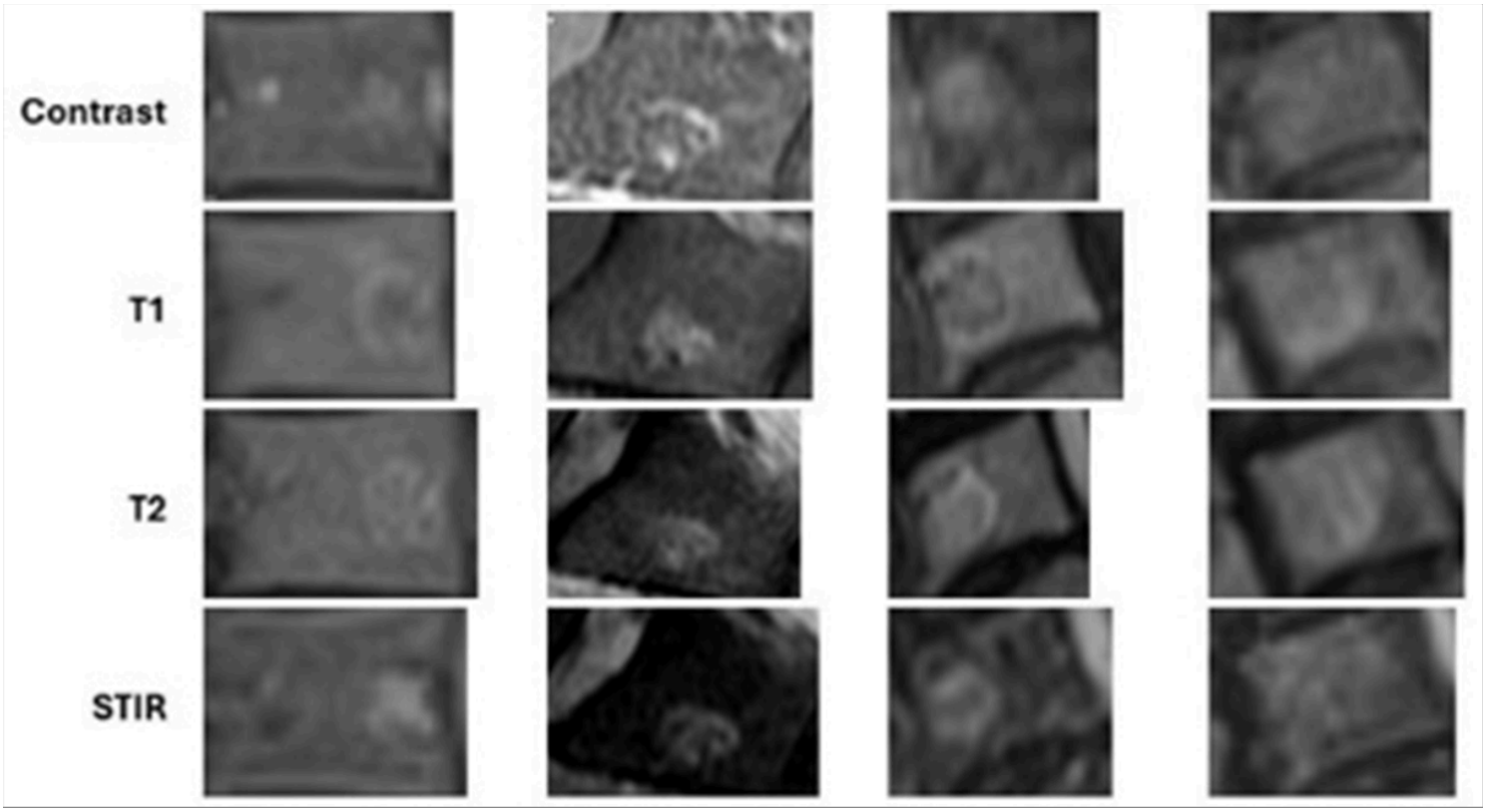


Figure 1

Legend not included with this version.

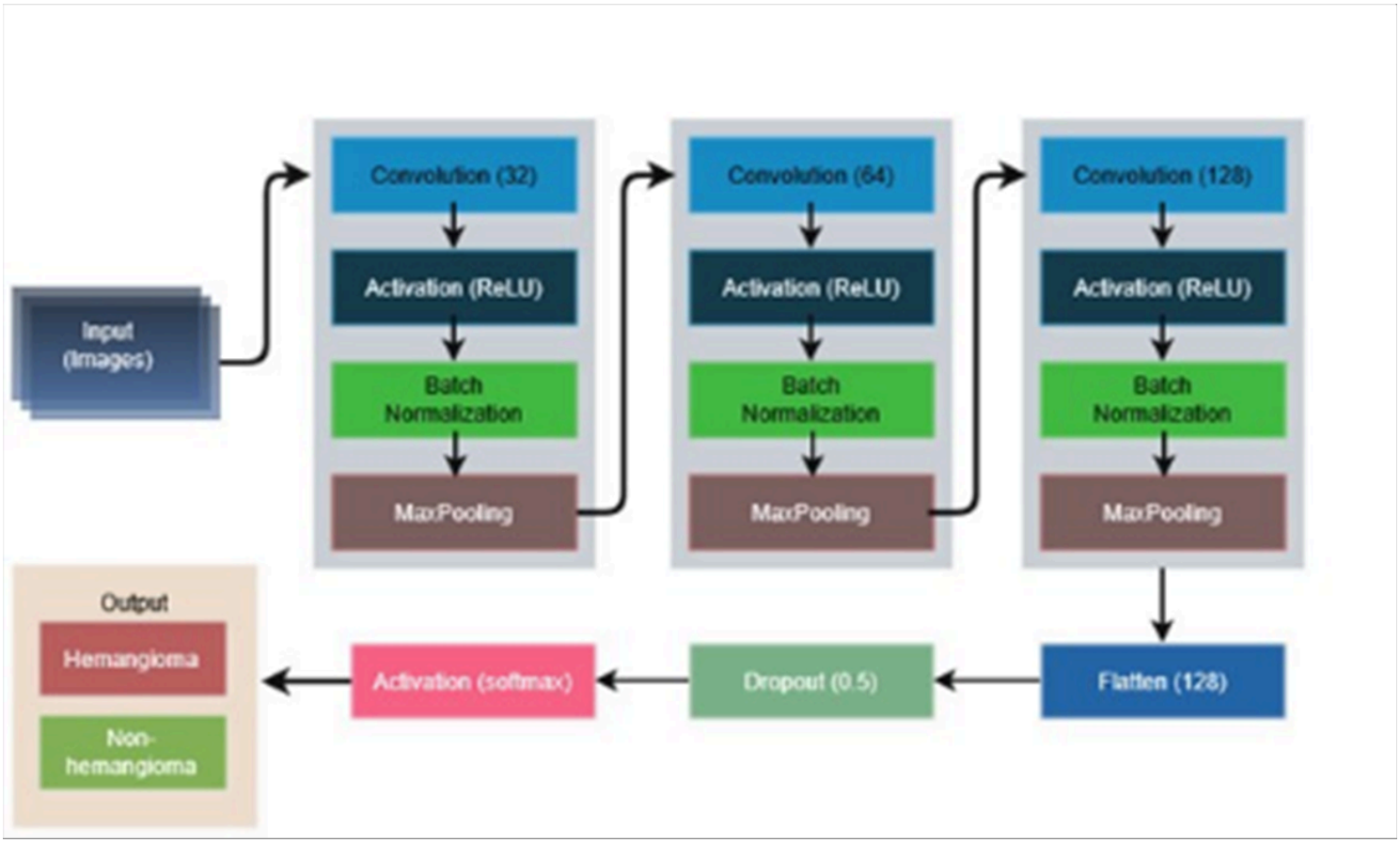


Figure 2

Legend not included with this version.

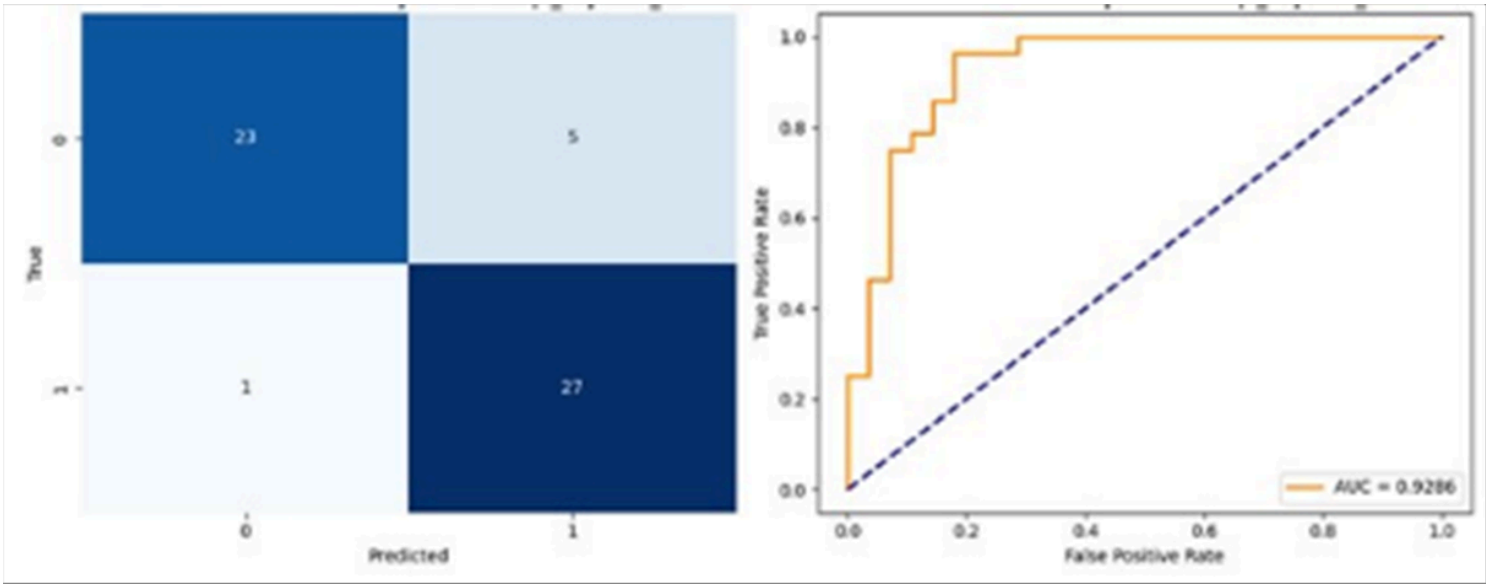


Figure 3

Legend not included with this version.

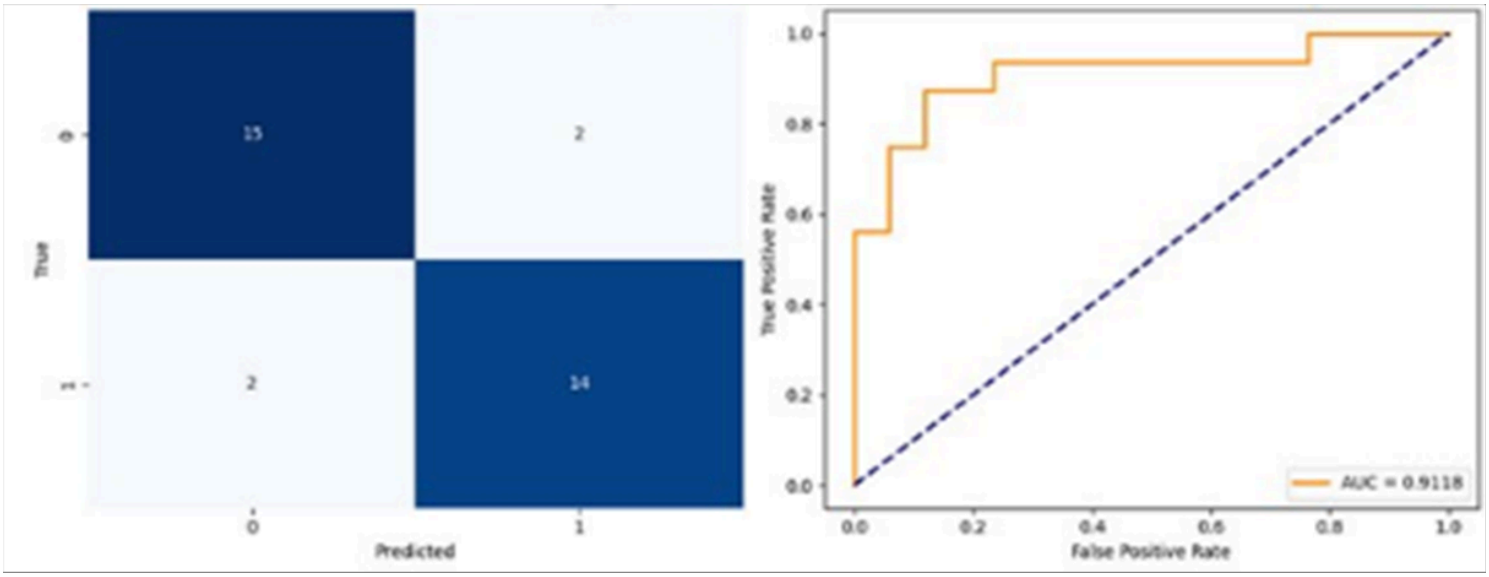


Figure 4

Legend not included with this version.

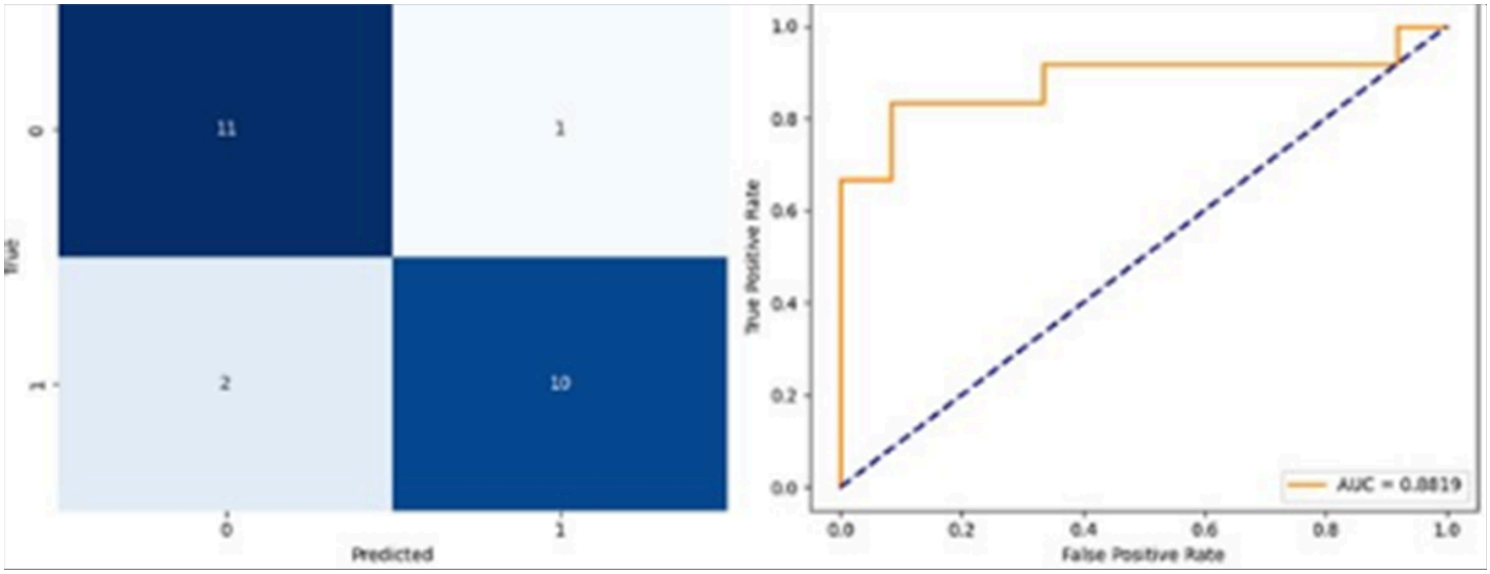


Figure 5

Legend not included with this version.

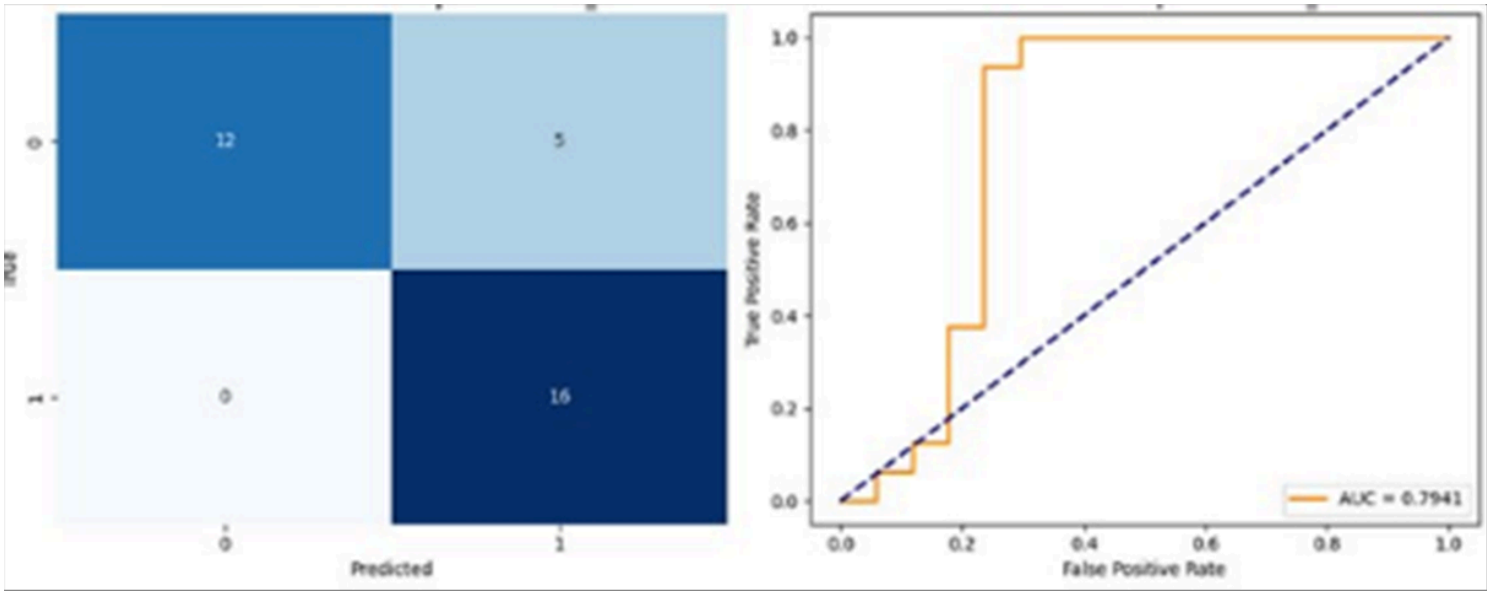


Figure 6

Legend not included with this version.