Soft tissue sarcoma diagnosis using machine and deep learning-survey

Vineela Madireddy 1*, Harikrishna Bommala 1*, Sirisha Yerraboina 2

1 Department of CSE, Bharatiya Engineering Science & Technology Innovation University, Anantapur, Andhra Pradesh.
2 Department of Information Technology, Matrusri Engineering College, Saidabad, Hyderabad, Telangana, India

Abstract. The collection of unique and diverse tumors known as soft-tissue sarcomas (STS) is hindered by a number of factors, such as delayed or inaccurate diagnosis, and a lack of clinical knowledge, and a restricted range of treatment alternatives. The tissues that surround, link, and support other body organs and structures are the target of a rare type of cancer known as soft tissue sarcomas. Muscle, fat, blood vessels, deep skin tissues, tendons, and ligaments are among the tissues that can be impacted by soft tissue sarcomas. Soft tissue sarcomas can arise in nearly every body component, including the arms, legs, and abdomen. The way that patients are treated medically is severely harmed by these diagnostic mistakes. Numerous machine learning models have been proposed by researchers to categorize cancers, but none of them have sufficiently addressed the issue of misdiagnosis. Furthermore, the majority of comparable research that has suggested models for the assessment of these malignancies do not take the heterogeneity and volume of the data into account. This research presents the comparison between machine and deep learning methods for the improved categorization of soft tissue sarcomas. This research further proposes on the early detection of STS. In the next stage of classification, an optimal Convolution Neural Network (CNN) is employed.

1 Introduction

body’s other

* Corresponding author: haribommala@gmail.com
tissue tumors (STT) are so difficult to detect, accurate diagnosis is the key obstacle, because of this reason various strategies are created for the detection. MRI can be used to examine the tumor's texture or other, less well-defined features intensity, form of tumor boundaries for a number of reasons: (1) textural characteristics are easily computed; (2) textural characteristics and tumor pathology are widely correlated; and (3) textural characteristics are robust to changes in MRI acquisition parameters, such as modifications in the tumor image's resolution and the image's corruption brought on by the heterogeneity of the magnetic field. Certain malignant tumors have a texture that is difficult to detect because to the heterogeneity of the magnetic field in MRI, and humans' ability to detect and distinguish these textures.

Malignant tumors known as STT can arise inside a variety of tissues, including blood vessels, muscles, fat, nerves, and fibrous tissues. Due to its infrequent occurrence and the difficulties doctors face when evaluating results; these obstacles have halted the creation of novel medicinal medicines. Additionally, doctors find it challenging to select an appropriate course of treatment due to the uneven MRI images. Furthermore, STT is frequently mistaken for various illnesses such as lymphadenopathy, fibroadenoma mammae, and struma nodosa. The course of the patient's treatment is significantly impacted by this diagnostic failure is limited. As a result, machine learning (ML) techniques are being used increasingly to analyze the MRI pictures and identifies more efficiently.

The idea put forth by Karanian and Coindrestates that there are four varieties of tumors: tumors with low metastatic potential, connective tumor evolution benign lesions, sarcomas and tumors with local potential. Sonographers and others may find use for an automated tool that can carry out simple categorization tasks, including differentiating between benign and malignant tumors. Radiologists by boosting diagnostic confidence and offering clinical decision support.

High accuracy scores in image identification tasks have been demonstrated by CNN. CNNs is implemented successfully in the classification of MRI images, including such as the detection of skin cancer, the characterisation of lung nodules at screening CT, the evaluation of tuberculosis at chest CT, and the detection of pneumonia through chest scan. Additionally, there is a growing corpus of research on the various ML application on the assessment of soft-tissue lesions.
2 Literature Review

A CNN model that showed images of lipoma are the masses with histologic diagnosis confirmed at biopsy or surgical excision or a vascular malformation which is a benign peripheral nerve sheath tumour. To identify the tumour is cancerous or non-cancerous, 227 images of patients were trained and assessed using CNN model who had a histologic diagnosis. The model was trained using 75%-25% training data with fourfold cross-validation and validation split on the rest of the examples were utilized, with 20% of the data is retained as a test dataset. The model's performance was contrasted with two musculoskeletal radiologists with experience who blinded themselves to the patient's clinical history and retrospectively interpreted the same information.

To train and assess an alternative model to identify the three prevalent benign masses found in 275 of the 419 patients, a second set of US pictures was utilized. Distinguish between the masses. Using a modified pretrained VGG16 network, the Keras machine learning platform (version 2.3.1) is used to train the models. Proposed a ML method The number of patients, labels, and metric scores were retrieved from research that were diagnosis-oriented in order to further analyze them for the synthesis. There were no discernible relationships between the measures and the average number of samples. Numerous research shown how ML might help the diagnosis of musculoskeletal tumours in specific circumstances based on imaging. To obtain results that are therapeutically useful, however, more data of higher quality and quantity are required. In contrast to the expertise of a radiology specialist, the investigations primarily contained a single type of data and used limited datasets.

Suggested a DL method using digital pathology on diagnosis and prognosis prediction of STS. An extra test data consisted of 51 patients from a second multicenter cohort. Nine pathologists with varying degrees of experience assessed the usage of the Deep Learning Model as a clinical decision support system. Regarding prognosis Overall, 139 leiomyosarcoma (LMS) slides from 85 patients were utilized. AUROC is the primary outcome metric. The 5 most prevalent STS subtypes with an accuracy of 79.9% (~6.1%) and a mean AUROC of 0.97 (~0.01) was diagnosed with DLM. The accuracy of the pathologist was increased dramatically with the DLM, from 46.3% (about 15.5%) to 87.1% (about 11.1%). Moreover, their diagnosis was made substantially faster and with greater certainty. Within LMS, the average survival prediction after the diagnosis of STS was 88.9% (~9.9%) with an AUROC of 0.91 (~0.1). In these patients, Cox regression demonstrated that the DLM prediction outperformed other risk factors as a significant independent prognostic predictor (P = 0.008, hazard ratio 5.5, 95% confidence interval 1.56-19.7).

Proposed Deep Learning Methodology to analyze images could provide a different approach to describe STS tissue. STS employing MR-imaging-based DL methods, the techniques used are two separate retrospective cohorts to gather contrast-enhanced T1FSGd and T2FS MRI images with training data of 148 patients and testing data of 158 patients. In pre-therapeutic biopsies, tumour grading was established in accordance with the FNCLCC. DL models were developed on the DenseNet 161 architecture using Transfer Learning. AUC values of 0.75 and 0.76 are used for the T1FSGd and T2FS-based deep learning models, respectively. Out of all the models, T1FS Gd has the best F1-score (0.90).

Proposed a Artificial Intelligence model with total of 134 images, including both high-grade and low-grade T1 and T2 images, were employed in the creation of the artificial intelligence model. The best representative image of the tumour at any slice was selected. This resulted in almost 36 million pixels that the Landing AI software examined. To use magnetic resonance imaging and artificial intelligence approaches to assess the level of,
total of 182 radiographs from 58 distinct pediatric (≤18 years old) patients were gathered (118 healthy, 20 osteomyelitis, 44 Ewing sarcoma). To address the challenges, the algorithm was divided into 2 phases, and two distinct classifiers were constructed and integrated via a Transfer Learning methodology with a very limited data. Phase 1 involved distinguishing between abnormal and healthy findings. Osteomyelitis was differentiated from Ewing sarcoma in phase 2. The use of data augmentation and median frequency balance was made. For training, validation, and hold-out testing, a data split of 70%, 15%, and 15% was used, respectively. In phase 1, the algorithm's accuracy was 90.6% on test data and 94.4% on validation data. Phase 2 results showed an accuracy of 86.7% on test data and 90.3% on validation. Regions identified by Grad-CAM results were important for the algorithm's decision-making.

[7] proposed a DL-CNN-based diagnostic method that measures the likelihood of diagnosis among trained subtypes of STS based on entire tissue slides intended to serve as a screening tool for pre-pathologists. With accuracy ROC and AUC values above 0.889 for all the subtypes of sarcoma, this CNN model successfully classified the withheld testing population. After that, we classified an externally-sourced cohort of data using the CNN model. Samples of rhabdomyosarcoma from human alveoli and embryos, as well as 318 tissue sections from histopathology obtained from genetically modified mice models of the disease.

[8] proposed machine-learning techniques to monitor changes following radiation therapy by autonomously identifying tissue compartments in STS. Multi-parametric MRI was used to examine 18 individuals with retroperitoneal sarcoma; 8/18 of these patients had a follow-up imaging examination after 2-4 weeks following pre-operative radiation. Eight popular supervised machine-learning approaches were refined to categorize pixels into five tissue subtypes as a gold standard, utilizing expert-defined regions of interest and a thorough cross-validation procedure. Using a MRF prior distribution on the ML models, the final pixel classification was smoothed. There was no discernible difference in the high median cross-validation accuracy of 5/8 machine-learning algorithms (82.2%, range 80.5–82.5%). The Naïve-Bayes technique was chosen because of its comparatively quick prediction and training periods (median 0.73 and 0.69 ms) [9] proposed a machine-learning-based strategy that combines a novel feature transformation preprocessing method, resampling techniques to reduce instability, and classifies tests using SVM and Decision Trees. The performance was measured using two distinct and more effective models, DT and SVM. The difference between the SVM and DT algorithms was marginal. AUC was 96.4%, accuracy was 99.0%, and the DT-based model performed best with f1-measure: 99.3%. Therefore, in terms of accuracy, f1-measure, AUC, the DT-based model outperformed the SVM-based model by 2.0%, 2.6%, and 1.4%, respectively.
predict survival rates in osteosarcoma and STS1. Soft Tissue sarcoma is one of the rare types of cancer which can be diagnosed through X-rays, CT scans and other imaging techniques used to image within the body to determine the size and location of soft tissue sarcoma.

Bone scans use radioactive probes to examine the bones. The patient receives only a small amount of radiation. The patient is given an injection of the tracer. It converges at specific bony sites and it is marked by a specialized camera. The camera also captures areas where cancer causes the loss and makes healthy bone appear smaller.

Biopsy (removal of few cells examined under a microscope) will confirm that the tumor is a soft tissue sarcoma, and the biopsy result is a basis to confirm.

Heart evaluation: Cardiac examination including Echocardiogram (ECHO) and Electrocardiogram (EGG or ECG) will identify for structural abnormalities in the organ and look for changes in heart wall motion. This test is used to identify cardiac sarcoma.

A physical examination of the physician will confirm the tumor is malignant. But further improvements, advances in technology, such as machine learning, offer compelling possibilities for the STS leap.

Soft Tissue Sarcoma can be diagnosed by using Machine and Deep Learning Techniques.

3.1 Algorithms and Techniques in the detection of Soft Tissue Sarcoma

3.1.1 Machine Learning

Machine learning which is a sub-domain of AI. Training data is used to learn and improve the system with supervised learning and unsupervised learning techniques. Machine learning (ML) represents data in different patterns and it is used to create its predictions. ML algorithms and models are often learned through experience. To extract patterns from data and associate them with a small sample size, Machine learning algorithms and models are used.

Through the implementation of Support Vector Machine (SVM) and artificial neural networks (ANN), naive Bayes classifiers, and AdaBoost algorithms a robust model for prediction of Soft Tissue Sarcoma with high performance is developed. Dimensionality is reduced on by principal component analysis. Studies showed that decision trees, regression trees and other In contrast to methods, artificial neural networks (ANNs) are the most widely used method. A reliable method of forecasting its predictions in real time are provided by ANN techniques.

3.1.2 Machine Learning Classifier

There are many machine learning frameworks readily available but since ML methods are based on a given set of data, it is impossible to say which is better than what the other researchers use ML classifications.

3.1.3 Machine Learning Classifier

There are many machine learning frameworks readily available but since ML methods are based on a given set of data, it is impossible to say which is better than what the other researchers use ML classifications.
3.1.3 Logistic Regression

Logistic Regression is one of the supervised machine learning techniques applied to a classification problem (binary or multiclass). LR is based on the concept of probability. LR is represented by the following graph.

![Logistic Regression Graph]

Fig. 3. Logistic Regression

3.1.4 Support Vector Machine (SVM)

SVM is a supervised machine learning method that uses classification and inference of regression topics. Hyperplanes help SVM to explicitly classify data points into N distinct classes. Figure 3 shows the graph representation of the SVM. Among all ML classifiers, the SVM classifiers achieved the highest accuracies of 96.25 and 96.5% simultaneously.

![Support Vector Machine Graph]

Fig. 4. Support Vector Machine

3.1.5 Decision Tree

The concept of regression and concept of distribution are found in decision tree. The decision tree consists of root node and leaf nodes. Classification takes place in leaf nodes, while decision nodes are used for testing. Figure 4 shows a tree showing the decision tree. A straightforward CART algorithm and J48 were used to achieve an accuracy.
of 98.13 percent. J48 is a continuous and longitudinal analysis tool based on the Iterative Dichotomiser 3 (ID3). CART is based on the Gini index.

Fig. 5. Decision Tree

3.2 Deep-Learning Based Classification Framework

The development of a CNN model for characterizing images of soft-tissue masses in the musculoskeletal system is an area of research to be explored. In this study, we present two CNN models that automatically classify soft-tissue masses depicted on images. One model was trained to distinguish between benign and malignant masses.

Purpose: To train convolutional neural network (CNN) models to classify benign and malignant soft-tissue masses.

Fig. 6. Detection of Soft Tissue Sarcoma using CNN

3.2.1 Convolution Neural Network

CNNs are generated by neurons that learn to adapt, traditional artificial neural networks are resembled by CNN. Multiple ANNs receiving inputs and performing...
operations will still serve each neuron. From the raw image vectors of the inputs to the network to the class score that is finally produced, the network will continue a perceptual scoring task.

All traditional ANN techniques and techniques will still apply, and there will be problems associated with different groups of layers. CNN requires less preprocessing in comparison with other ML classifiers.

To predict a malignant tumor, Artificial Intelligence, Machine Learning, Deep Learning techniques use information from sources such as screening scans and medical histories. Modern Imaging techniques when combined with artificial intelligence (AI) enabled the real-time detection of tumors. Furthermore, this technique successfully detected cancerous tissue, demonstrating exceptional capabilities to distinguish between malignant and normal tissues.

Machine learning algorithms have been used to detect mismatch repair deficits (dMMR) in colorectal screening.

Fig. 7. AI, ML, DL techniques in Identifying Sarcoma

4 Conclusion

On surveying and analyzing various research papers on the soft tissue sarcoma using machine and deep learning techniques, it is concluded that the accuracy of 71% is obtained from CNN, 84.3% from AI, 82.2% from Naive Bayes, and highest accuracy 99.0% have obtained from SVM. It is analyzed by combining both the Machine and Deep Learning techniques an efficient model can be developed in prediction of soft tissue sarcoma at an initial stage by using MRI images as input.

References

