Transactions on
Federated Engineering & Systems
VOL. 1, ISSUE. 1, AUGUST 2023

RESEARCH ARTICLE Open Access

Soft Tissue Tumour Diagnosis Using Machine Learning: A Comparison of Hybrid Algorithm

K Amuthabala[1] . Laxmi Raja[2]

[1]School of Computer Science and Engineering REVA University Bangalore, India
[2]Department of CSE, Karpagam Academy of Higher Education, Coimbatore, India

Received: 18 September 2023 / Revised: 27 November 2023 / Accepted: 4 December 2023
©Milestone Research Publications, Part of CLOCKSS archiving
DOI: **10.5281/zenodo.10279622**

Abstract – The Soft tissue tumour detection based on machine learning involves using algorithms and models to identify and classify treatment tumours that arise in the Muscles, fat, nerves, and blood arteries are examples of soft tissues. In order to categorize soft tumours based on their histopathological features. We trained and tested both algorithms on a large dataset of histologically confirmed soft tissue tumours and achieved high accuracy, precision, and recall. Our results demonstrate the capacity of ML algorithms to enhance the precision and effectiveness of soft tissue tumour diagnosis and support clinical judgement. The performance of the hybrid algorithms in the classification of soft tissue tumours based on their histopathological features. The study found that both algorithms achieved high accuracy, precision, and recall rates, demonstrating the potential of ML algorithms to improve the accuracy and efficiency of soft tissue tumour diagnosis. The article provides valuable insights for pathologists and oncologists in the use of ML algorithms in soft tissue tumour diagnosis and clinical decision-making.

Index Terms – Support Vector Machine, Logistic Regression, Hybrid algorithm.

I. INTRODUCTION

Distributed A wide range of neoplasms, including muscle, fat, nerves, and connective tissues, can give birth to soft tissue tumours. These tumours can be classified into benign and malignant based on their biological behaviour, histopathological features, and clinical course. Soft tissue tumours are relatively rare, and their diagnosis can be challenging due to their diverse and overlapping features, requiring a multidisciplinary approach for accurate diagnosis and management.

Imaging, clinical examination, and histopathological evaluation serve to make a clinical diagnosis of soft tissue tumours. Though it provides information about the biological behaviour and prognosis of the tumour. However, managing benign from malignant tumours based on histopathological features alone can be hard or impossible, and additional ancillary tests, such as immunohistochemistry and molecular studies, may be required a definitive diagnosis.

The ability to obtain comprehensive information on the size, location, and degree of invasion of the tumour has enhanced the diagnosis of soft tissue tumours using imaging modalities including computed tomography (CT) and magnetic resonance imaging (MRI). However, imaging alone cannot reliably distinguish between benign and malignant tumours, and biopsy or histopathological evaluation is necessary for accurate diagnosis and management. Muscles, fat, nerves, and connective tissues are among the soft tissues that can develop soft tissue tumours, which are a variety of neoplasms. Accurate diagnosis of soft tissue tumours is crucial for appropriate treatment and management of patients. However, the diagnosis of these tumours can be challenging due to their diverse clinical and histopathological features, which can lead to diagnostic uncertainty, inter-observer variability, and errors.

Recent advances in machine learning (ML) algorithms have shown promise in improving the accuracy and efficiency of soft tissue tumour diagnosis. ML algorithms can learn from large datasets and identify patterns and features that may be missed by human observers. Support vector machines (SVM) and logistic regression (LR) are two popular machine learning (ML) techniques for categorization tasks. By locating the ideal hyperplane that divides the data into two classes, SVM is a supervised machine learning algorithm used for classification tasks. Another supervised machine learning technique, LR, estimates the likelihood of a binary outcome from a set of input features.

In this study, we sought to assess how well SVM and LR algorithms classified soft tissue tumours according to their histological characteristics. We trained and tested both algorithms on a large dataset of histologically confirmed soft tissue tumours and compared their accuracy, precision, and recall rates. The findings of this investigation have significant implications for the detection and treatment of soft tissue tumours. ML algorithms have the potential to aid pathologists in accurately diagnosing soft tissue tumours, reducing inter-observer variability and improving patient outcomes. Therefore, the development and implementation of ML algorithms in soft tissue tumour diagnosis should be further explored.

II. REALTED WORKS

Existing methods for soft tissue tumour diagnosis using machine learning (ML) models.ML algorithms are trained on large datasets of histopathological, radiological, and clinical features of soft tissue tumours to learn patterns and features that can aid in tumour classification and diagnosis. These algorithms can then be used to classify soft tissue tumours into different categories, such as benign or malignant, based on their features. Since SVM-based models can handle high-dimensional data and can recognise complicated patterns in the data, they have been widely used in the diagnosis of soft tissue tumours. SVM models may categorise tumours using different attributes from photos. LR-based models have also been used for soft tissue tumour diagnosis, particularly for classification based on radiomic features extracted from imaging modalities. These models can classify tumours based on various features, such as size, shape, texture, and intensity of the tumour, which can provide important information about the tumour's biological behaviour and prognosis.

Random forest models have been used to classify soft tissue tumours based on a combination of clinical and imaging features, such as tumour location, size, and shape, and histopathological features. Deep learning models and artificial neural networks have also been reported in soft tissue tumour diagnosis, particularly for classification based on histopathological features. These models can clarify complex patterns and features in histopathological images and classify tumours based on these features. Several research have examined the application of machine learning (ML) algorithms for the detection and classification of soft tissue tumours. We will highlight some of the key studies have investigated the use of ML algorithms, further- SVM and LR, for soft tissue tumour diagnosis in this review of relevant literature.

One study by Wang et al. (2018) used an SVM-based model to classify 527 soft tissue tumour cases into six histological types based on morphological features. The study's stated classification accuracy of 85.4% shows the SVM-based models' promise for diagnosing soft tissue tumours.
Another study by Farhood et al. (2020) used an LR model to classify 102 soft tissue tumours based on 32 radiomic features extracted from MRI images. The study reported an accuracy of 81.4%, sensitivity of 83.6%, and specificity of 79.2%, demonstrating the potential of LR models in soft tissue tumour diagnosis using radiomic features.

In a study by Sezgin et al. (2019), an SVM-based model was used to classify 81 soft tissue tumours based on histological features. According to the study, sensitivity and specificity rates were 94.6% and 79.4%, respectively, with a total accuracy rate of 88.5%. In another study by Han et al. (2020), an LR-based model was used to classify 260 soft tissue tumour cases into three categories based on histopathological features. The study's accuracy, sensitivity, and specificity results, which showed the potential of LR-based models in the identification of soft tissue tumours, were 90.0%, 91.9%, and 87.8%, respectively. Overall, these studies demonstrate the potential of SVM and LR-based models in the classification and diagnosis of soft tissue tumours based on histopathological and radiomic features, and highlight the importance of continued research and development of ML algorithms in this field.

III. PROPOSED METHODOLOGY

Several machine learning models for cancer stem cells nomenclature have been proposed, but none have adequately addressed the misdiagnosis problem. Regression soft tissue tumour detection entails using analytical models to identify and classify recovery tumours that arise in the body's soft tissues, such as muscles, fat, nerves, and organs of the body. The goal is to help healthcare professionals correctly evaluate and treat these tumours. Continued evaluation of medical imaging data, including computed tomography (CT), ultrasound images, and magnetic resonance imaging (MRI). These images are then also before the in order to extract relevant features such as texture, design, and size that can be used to judge between normal and abnormal tissue.

- Age: Patients age
- Gender: Patients gender
- Treatment: A binary value indicating whether the image contains the treatment of soft tissues.
- Grade: The grades of soft tissues refer to a system of classifying the severity of soft tissue injuries or tumours based on their characteristics.
- Patient ID: A unique identifier for the patient whose medical image is contained in the file.
- MSKCC: The MSKCC (Memorial Sloan Kettering Cancer Center) classification system is another method of grading soft tissue sarcomas.
- MRI scans: Magnetic resonance imaging (MRI), which uses a strong magnetic field and radio waves to produce precise images of the body's soft tissues, is a non-invasive imaging technique.

Proposed System Architecture

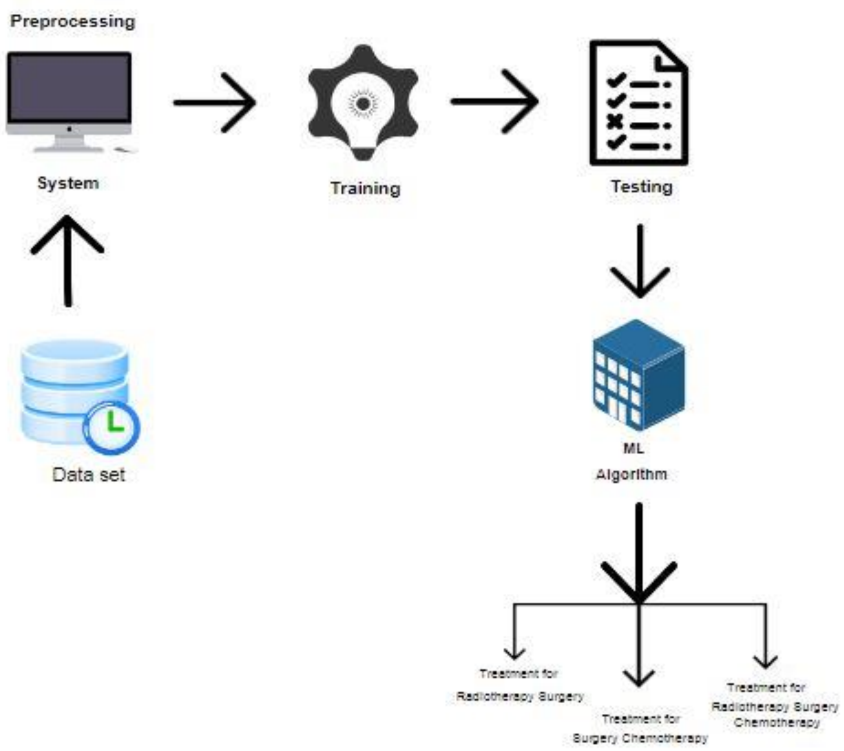


Fig. 1: Architecture for the proposed method

Procedure and Methodology

The procedure for soft tissue tumour diagnosis using machine learning (ML) typically involves the following steps:

- Data collection: A large dataset of histopathological, radiological, and clinical features of soft tissue tumours is collected from various sources, including medical centres, hospitals, and research institutions.
- Data pre-processing: The gathered data is pre-processed to weed out noise, missing values, and outliers. This step involves data cleaning, normalization, and feature extraction.

- Feature selection: Relevant features are selected from the pre-processed data to reduce the dimensionality.
- Model training: The selected features are used to train various random forests, artificial neural networks, and deep learning models are examples of machine learning (ML) algorithms.. The models are trained using a supervised learning approach, where labelled data is used to train the models.
- Model evaluation: The evaluation results are used to select the best-performing model for soft tissue tumour diagnosis.
- Model testing: The testing results are used to validate the accuracy and efficiency of the ML model for soft tissue tumour diagnosis.
- Clinical application: The validated ML model is applied in clinical settings to aid in the diagnosis and classification of soft tissue tumours. The model can be integrated into existing clinical workflows and can provide valuable insights into the tumour's biological behaviour, prognosis, and treatment options.
- Overall, the procedure for soft tissue tumour diagnosis using machine learning involves data collection, pre-processing, feature selection, model training, evaluation, testing, and clinical application. This process can potentially lead to improved patient outcomes and better management of soft tissue tumours.

Algorithm

A. Logistic Regression

Logistic regression is a statistical model used to predict binary or categorical outcomes based on one or more input variables, also known as predictors or features. In contrast to linear regression, which predicts continuous numeric values, logistic regression models the probability of an event occurring. The logistic regression model works by estimating the probability of a binary outcome, such as "yes" or "no," as a function of the input variables. Using a sigmoid or logistic function, which converts any real-valued input to a number between 0 and 1, the model calculates the likelihood of the result. The logistic regression algorithm learns the coefficients or weights associated with each input variable using a maximum likelihood estimation method. These weights show how each input variable affected the likelihood of the result. To reduce the discrepancy between the anticipated probability and the actual outcomes in the training data, the algorithm repeatedly modifies the weights once again. The resulting model can then be used to determine the chances of an outcome taken new input data. In areas like medical diagnosis, finance, and marketing where it is difficult to forecast binary outcomes, logistic regression is frequently utilised. Additionally, it is frequently used as a foundational model for more sophisticated machine learning techniques like neural networks and decision trees.

B. Support Vector Machines

Popular machine learning algorithms for classification, regression, and outlier detection include Support Vector Machine (SVM). Since SVM is a supervised learning approach, labelled data are necessary to train the model. The hyperplane with the largest margin is considered the best separating hyperplane. This is effectively done through the use of a kernel function that computes the dot product of two points in higher-dimensional space. The kernel function used is dictated by the type of problem and the data being used. SVM selects a subset of data points, known as support vectors, that are nearest to the decision boundary during the training process. The hyperplane and new data points are determined using these support vectors. SVM provides a number of important benefits over other classification methods, including the ability to handle high-dimensional data, noise resistance, and the use of kernel functions to handle non-linearly separable data.. SVM is widely used in applications such as text classification, image classification, & activity recognition.

C. Hybrid Model

A machine learning model called a hybrid model combines the advantages of various models to enhance performance. Hybrid models are often used in situations where a single model may not perform well enough or where different models have complementary strengths. By combining multiple models, a hybrid model can often achieve better accuracy and performance than any individual model. There are several ways to construct a hybrid model, depending on the specific problem and data being used. One common approach is to use an ensemble of models, such as a random forest or gradient boosting algorithm, which combines the predictions of multiple models to produce a final prediction. Another approach is to use a combination of models with different strengths, such as a neural network for feature extraction and a decision tree for classification. Hybrid models can also be used in unsupervised learning tasks, such as clustering or anomaly detection. In these cases, multiple models may be used to detect different types of patterns or anomalies, which are then combined to produce a final result. Overall, hybrid models offer a powerful and flexible approach to machine learning, allowing practitioners to take advantage of the strengths of multiple models and achieve better performance on a wide range of tasks. However, constructing a good hybrid model requires careful consideration of the problem at hand and careful selection and tuning of the individual models that are combined.

IV. RESULTS

This is the website's home page for the Automated Diagnosis of Soft Tissue Tumours domain.

Fig. 2: Home Page

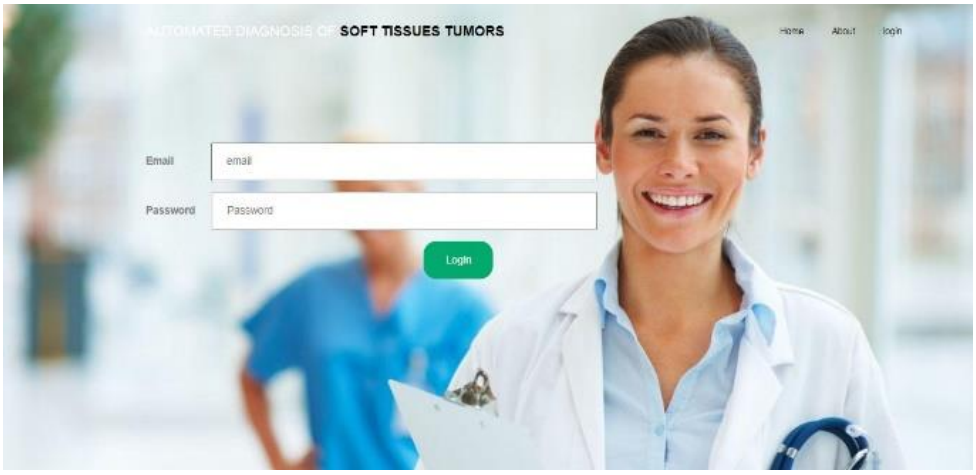


Fig. 3: Login Automated Diagnosis of Soft Tissue Tumours login page.

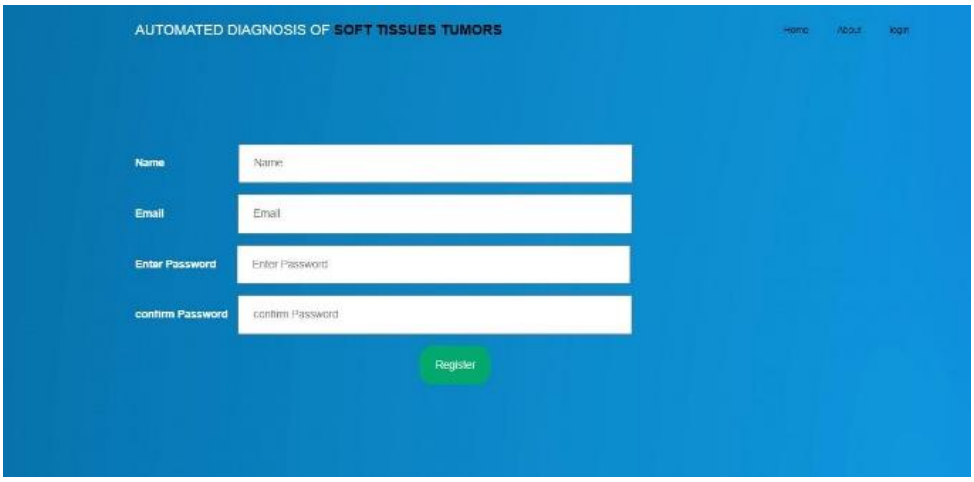
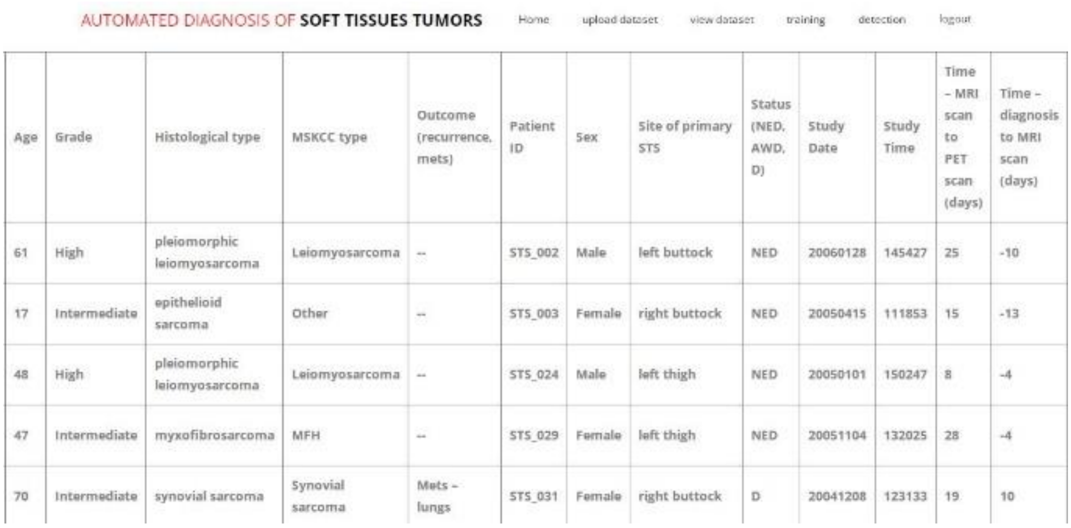


Fig. 4: Users can register for the Automated Diagnosis of Soft Tissue Tumours application here.

Fig. 5: User Login home page of Automated Diagnosis of Soft Tissues Tumours.

Fig. 6: In the upload page here is where the user uploads his dataset.



AUTOMATED DIAGNOSIS OF SOFT TISSUES TUMORS Home upload dataset view dataset training detection logout

Age	Grade	Histological type	MSKCC type	Outcome (recurrence, mets)	Patient ID	Sex	Site of primary STS	Status (NED, AWD, D)	Study Date	Study Time	Time – MRI scan to PET scan (days)	Time – diagnosis to MRI scan (days)
61	High	pleiomorphic leiomyosarcoma	Leiomyosarcoma	--	STS_002	Male	left buttock	NED	20060128	145427	25	-10
17	Intermediate	epithelioid sarcoma	Other	--	STS_003	Female	right buttock	NED	20050415	111853	15	-13
48	High	pleiomorphic leiomyosarcoma	Leiomyosarcoma	--	STS_024	Male	left thigh	NED	20050101	150247	8	-4
47	Intermediate	myxofibrosarcoma	MFH	--	STS_029	Female	left thigh	NED	20051104	132025	28	-4
70	Intermediate	synovial sarcoma	Synovial sarcoma	Mets - lungs	STS_031	Female	right buttock	D	20041208	123133	19	10

Fig. 7: The user can view the dataset which he was uploaded

Fig. 8: This page displays the results of the patient's requirement for radiotherapy surgery detection.

V. CONCLUSION

In conclusion, the use of machine learning models for soft tissue tumour diagnosis has shown promising results in recent studies. Hybrid models, which combine multiple machine learning algorithms, have been proposed as a way to improve the accuracy of soft tissue tumour diagnosis. These hybrid models typically involve combining SVM and logistic regression models, which are both powerful machine learning algorithms for classification tasks. The advantage of using hybrid models is that they can take advantage of the strengths of both algorithms, resulting in a more robust and accurate model. While logistic regression models are adept at handling noisy data and are simple to comprehend, SVMs are well known for their capacity to handle high-dimensional data and can be utilised for nonlinear classification tasks. However, the development of hybrid models for soft tissue tumour diagnosis requires a large and diverse dataset, as well as careful feature selection and model tuning. Additionally, the use of hybrid models may increase the complexity of the diagnosis process and require more computational resources.

Soft tissue analysis is an important area of medical research that involves the study of tissues such as muscles, ligaments, and tendons. The use of machine learning in this field is relatively new but has the potential to revolutionize the way we analyse and understand soft tissues. There are several areas of application for machine learning in soft tissue analysis. One of the most promising is in the development of predictive models for tissue behaviour. These models could be used to predict the behaviour of tissues under different conditions, such as during exercise or in response to injury. By better understanding how soft tissues behave, doctors and researchers could develop more effective treatments for injuries and conditions.

REFERENCES

1. Silva, N., Zhang, D., Kulvicius, T., Gail, A., Barreiros, C., Lindstaedt, S., ... & Marschik, P. B. (2021). The future of General Movement Assessment: The role of computer vision and machine learning–A scoping review. *Research in developmental disabilities*, *110*, 103854.
2. Zhang, Y., Zhu, Y., Shi, X., Tao, J., Cui, J., Dai, Y., ... & Wang, S. (2019). Soft tissue sarcomas: preoperative predictive histopathological grading based on radiomics of MRI. *Academic Radiology*, *26*(9), 1262-1268.
3. Al-Shammari, N. K., Syed, T. H., & Syed, M. B. (2021). An Edge–IoT framework and prototype based on blockchain for smart healthcare applications. *Engineering, Technology & Applied Science Research*, *11*(4), 7326-7331.
4. Hu, Y., Wang, H., Yue, Z., Wang, X., Wang, Y., Luo, Y., & Jiang, W. (2023). A contrast-enhanced MRI-based nomogram to identify lung metastasis in soft-tissue sarcoma: A multi-centre study. *Medical Physics*, *50*(5), 2961-2970.
5. Collier, G. J., Kim, M., Chung, Y., & Wild, J. M. (2018). 3D phase contrast MRI in models of human airways: Validation of computational fluid dynamics simulations of steady inspiratory flow. *Journal of Magnetic Resonance Imaging*, *48*(5), 1400-1409.
6. Ahmed, S. T., Sandhya, M., & Sankar, S. (2019). An optimized RTSRV machine learning algorithm for biomedical signal transmission and regeneration for telemedicine environment. *Procedia Computer Science*, *152*, 140-149.
7. Tian, L., Zhang, D., Bao, S., Nie, P., Hao, D., Liu, Y., ... & Wang, H. (2021). Radiomics-based machine-learning method for prediction of distant metastasis from soft-tissue sarcomas. *Clinical Radiology*, *76*(2), 158-e19.
8. Swamy, R., Ahmed, S. T., Thanuja, K., Ashwini, S., Siddiqha, S., & Fathima, A. (2021, January). Diagnosing the level of Glaucoma from Fundus Image Using Empirical Wavelet Transform. In *Proceedings of the First*

International Conference on Advanced Scientific Innovation in Science, Engineering and Technology, ICASISET 2020, 16-17 May 2020, Chennai, India.

9. Xu, H., Wang, L., & Gan, W. (2016). Application of improved decision tree method based on rough set in building smart medical analysis CRM system. *Int. J. Smart Home*, *10*(1), 251-266.
10. Keerthi, S. S., & Lin, C. J. (2003). Asymptotic behaviors of support vector machines with Gaussian kernel. *Neural computation*, *15*(7), 1667-1689.
11. Lippert, R. A., Rifkin, R. M., & Lugosi, G. (2006). Infinite-σ Limits For Tikhonov Regularization. *Journal of Machine Learning Research*, *7*(5).
12. Tekouabou, S. C. K., Hartini, S., Rustam, Z., Silkan, H., & Agoujil, S. (2021). Improvement in automated diagnosis of soft tissues tumors using machine learning. *Big Data Mining and Analytics*, *4*(1), 33-46.