iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** XI **Month of publication:** November 2021

DOI: https://doi.org/10.22214/ijraset.2021.38981

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue XI Nov 2021- Available at www.ijraset.com

Improvement of Automatic Diagnosis of Soft Tissue Tumours Using ML

P. Charitharyan[1], Anubhav Kar Bhowmik[2], Dilip Kumar[3], A. Vidyavani[4]
[1, 2, 3]*UG Scholar,* [4]*Asst. Professor, Department of Engineering and Technology, SRM Institute of Science and Technology, Ramapuram Campus, Chennai, India.*

Abstract: ***STTs are a type of sarcoma that develops in the tissues that connect, support, and surround bodily structures. Because of their scarcity and diversity, they are difficult to detect when seen using Magnetic Resonance Imaging (MRI). They are frequently mistaken with other disorders, and diagnostic errors have a significant negative impact on patients' medical care. Several methods for classifying these tumours have been presented by researchers, but none have satisfactorily addressed the problem of misdiagnosis. This is due to the fact that most research that have developed models for evaluating such tumours ignore the heterogeneity and magnitude of the data. Because of their scarcity and diversity, they are difficult to detect when seen using Magnetic Resonance Imaging (MRI). They are frequently mistaken with other disorders, and diagnostic errors have a significant negative impact on patients' medical care. Several methods for classifying these tumours have been presented by researchers, but none have satisfactorily addressed the problem of misdiagnosis. This is due to the fact that most research that have developed models for evaluating such tumours ignore the heterogeneity and magnitude of the data. As a result, we offer a machine learning-based strategy that incorporates a new pre-processing technique for features modification, resampling approaches to minimise discrepancies, and Decision Tree (DT) algorithms to eliminate discrepancies. Applying Machine learning processes could provide effective tools to aid in the automatic decision-making processes of STT diagnosis.***
Keywords: Machine Learning, CT scan, PET scan, data pre-processing, soft tissue tumors, radiological image analysis, MRI, classification tumors, classification of tumor texture, pathology, digital pathology, Decision Tree Algorithms

I. INTRODUCTION TO SARCOMA

Cancers are characterised in two ways: by the type of tissue from which it develops (histological type) and by the primary site in the body where it first appeared. This part will introduce you to the first method, which is based on histological type classification of cancer. The international institute for histological classification and nomenclature is the International Categorization of Diseases for Oncology.

There are several forms of cancer, which are grouped into six main categories based on histology:

1) Carcinoma
2) Sarcoma
3) Myeloma
4) Leukaemia
5) Lymphoma
6) Mixed types

Cancers that start in the muscles, fat, fibrous tissue, blood vessels, or other supportive tissues of the body are known as soft tissue sarcomas. Tumors can appear everywhere on the body, although they are most commonly found in the arms, legs, chest, and abdomen.

Lumps or swelling can be signs of soft tissue sarcoma. Until the tumour grows and crushes surrounding nerves or other parts of the body, no indications or symptoms may appear.

Soft tissue sarcoma was diagnosed by biopsy.

If your doctor thinks you may have a soft tissue sarcoma, a biopsy will be done. The size of the tumour and its location on the body will determine the sort of biopsy required. Three types of biopsies can be used:

a) *Open Biopsy:* Removal of part of a mass or tissue sample.
b) *Central Biopsy:* Use a wide needle to remove tissue.
c) *Resection Biopsy:* Resection of the entire mass or tissue area that is abnormal.

©IJRASET: All Rights are Reserved

II. PROBLEM STATEMENT

The lack of attention devoted to data size and learner validation was a prominent issue in the studies evaluated in this research work. In other words, there are a lot of research in which the experimental design has received less attention. A sufficiently big data collection that can be partitioned into disjoint training and test sets is a minimum condition for every machine learning project. Sarcomas were discovered, but there was no way of knowing if it was a sarcoma or another sort of cancer, leading to doctors prescribing the wrong therapy and victims dying.

III. OBJECTIVE

We propose a machine learning-based approach that combines a new technique for pre-processing data for features transformation, resampling techniques to eliminate bias and the deviation of instability, and performing classifier tests based on the Support Vector Machine (SVM) and Decision Tree (DT) algorithms so that the machine can accurately and beforehand detect features transformations in the body such as tumours so that cancer can be treated at an early stage. Implementing a computer-aided diagnostic system based on our model may also prove to be more efficient and successful than relying just on a radiologist's visual assessment for diagnosis.

IV. EXISTING SYSTEM

The main diagnostic method for the identification and categorization of STT with well-characterized biological features such as cellular origins and tumour specimens used to identify malignancies is magnetic resonance imaging (MRI). The medical imaging technique of magnetic resonance imaging (MRI) is used by radiologists to create images of the body's architecture and physiological processes. The use of pathology in the medical unit has been substantial. The study of disease is known as pathology. It functions as a link between science and medicine. It serves as the foundation for all aspects of patient care, from diagnostic testing and treatment guidance to illness prevention and the utilisation of cutting-edge genetic technology. Pathology doctors and scientists are disease experts who apply their knowledge to all parts of health care, from teaching doctors how to treat common ailments to applying cutting-edge genetic technologies to treat patients with life-threatening disorders. Pathologists are essential in research, medical innovation, and the discovery of novel medicines for viruses, infections, and diseases like cancer. We've witnessed a substantial drop in diseases like polio around the world over the last 100 years, as well as significant advancements in blood transfusions, immunizations, and the treatment of hereditary ailments. Pathologists' pioneering work is responsible for all of this.
In addition to that, these are the two mandatory ways to find a cancer. They are:

A. CT SCAN

(Computed Tomography (CT) Scan)
Doctors utilise a computed tomography (CT) scan, often known as a CAT scan, to diagnose cancer. They could potentially utilise it to understand more about the malignancy once it has been discovered. They can learn the cancer's stage thanks to the scan. Knowing this will assist you and your doctor in selecting the most appropriate treatment options. It also aids doctors in predicting your recovery.

1) Steps are

Locate a suitable location for a biopsy.
Prepare for cancer treatment with radiation therapy.
During the therapy term, assess how effectively the treatment is functioning.
During follow-up care, look for new cancer growth after treatment has ended.
A CT scan employs X-rays to take several images of the inside of the body. These photos are combined by a computer into a detailed three-dimensional image. This image depicts all tumours and aberrant locations.
Before a test, some participants are given a special dye called contrast. The contrast substance is injected into a vein and passes through the circulation to assist create a crisper picture of particular sections of the body; it can also be administered as a beverage. Depending on whatever region of your body needs to be scanned, you may need to swallow.
The head, neck, chest, belly, pelvis, and extremities are all areas that are frequently tested for cancer. The chest, belly, and pelvis are frequently included in a "full body" CT scan. It's frequently used by doctors to determine the stage of cancer.
In most cases, the benefits of a CT scan exceed the risks: You are exposed to a little amount of radiation during a CT scan. There may be a slightly higher risk of cancer in the future if you have had many CT scans and X-rays.
In many circumstances, clinicians employ low-dose CT scans or limit the area scanned in youngsters. Ask your doctor about tests that will limit your radiation exposure if you undergo several CT scans and x-rays.

©IJRASET: All Rights are Reserved

B. PET Scan

PET scans create comprehensive three-dimensional images of the inside of the body.

The images can clearly reveal the part of the body that is being checked, including any abnormalities, as well as how well specific body functions are functioning.

To obtain even more comprehensive images, PET scans are frequently coupled with CT scans. This is referred to as a PETCT scan. They are occasionally paired with an MRI examination (so-called PETMRT examination). Instead of only showcasing your appearance, parts of your body work.

PET scans detect radioactivity generated as a radiotracer by a material injected into your arm as it accumulates in various places of your body.

Most PET scans employ a radiotracer called fluorodeoxyglucose (FDG), which is comparable to natural glucose (a form of sugar) and is processed similarly by your body.

By examining the places where radiotracer accumulates rather than accumulates, it is possible to evaluate how particular physiological functions are working.

FDG, for example, can be used to identify cancer cells in bodily tissues because cancer cells utilise glucose considerably faster than normal cells.

V. PROPOSED METHODOLOGY

A. Data pre-processing

The two key phases in our suggested data pre-processing technique are (1) pre-processing and (2) categorization. By dealing with each characteristic independently, the pre-processing stage reduces processing time and improves classification performance while staying relatively stable against variable scales.

B. Features Transformation

The raw data from the numerous tests and information sources is gathered and translated into a numerical representation in this pre-processing stage, allowing the Support vector machine classifiers to be applied. To do so, all categorical attributes (blood type, AGS-AS, gender, and disease) must first be categorised as nominal (blood type), binomial (gender and disease), or missing values, and then translated to numerical format according to the kind. I For Boolean functions like f3: We only have two options for this sort of feature: yes or no (1/0), male or female (1/0), or positive/negative (1/0).

VI. RESAMPLING TECHNIQUES

When a dataset has been captured in a specific time interval or number of samples, resampling is usually required. This method has an effect on data distribution, which makes it more ideal for classification algorithms. We wish to downsize and reorganise the data in this scenario to improve the model's performance. This is accomplished by removing the algorithms' bias and variance in order to construct a stable and realistic model without overfitting.

VII. CLASSIFICATION

Following the pre-processing processes, the next step in our challenge is to use appropriate classification algorithms to distinguish between non-STT and STT. These pre-processing methods allowed us to clean the data properly while reducing outliers and multicollinearity that might affect the classifiers' performance. Our method relies on the optimization of Support vector machine methods, which we will discuss briefly in the next sections.

In the management of patients, radiological imaging are becoming increasingly vital. The use of radiographic imaging for diagnostic and therapeutic purposes is rapidly expanding. The demand for faster, more precise, less expensive, and less intrusive therapy is driving the rapid expansion. Images were also welcomed by the teams. These advancements include the capacity to acquire pictures with increasingly greater resolutions, allowing for the visibility of tiny anatomical features and aberrations. A growing average number of photos per patient goes hand in hand with improved resolution. These images must be interpreted by radiologists, and as the quantity of images grows, so does the burden of radiologists. Radiologists are at risk of becoming overwhelmed by the growing amount and complexity of pictures. Automated and intelligent picture collection and processing are becoming more important in many radiological operations. Image segmentation, registration, and computer-aided diagnosis and detection. and treatment, for example, have a huge market and are employed in areas such as radiation therapy planning and automatic identification of image biomarkers from radiological pictures of particular disorders, among other things. The techniques and software that enable for computer-assisted diagnosis, prognosis, and therapy are based on machine learning algorithms.

©IJRASET: All Rights are Reserved

Radiology is a discipline of medicine that diagnoses and treats disease using radiation and imaging technologies. Advances in physics, electrical engineering, and computer science have all aided it immensely. Over the previous few decades, diagnostic radiology has progressed. The most often used modalities in hospitals and medical institutes nowadays include radiography, fluoroscopy, computed tomography (CT), ultrasound, magnetic resonance imaging (MRI), and positron emission tomography (PET).

Radiologists read and analyse medical pictures from many modalities on a regular basis in their work. In most cases, radiologists must analyse and interpret these pictures in a short amount of time. However, as medical technology advances, the amount of imaging data generated continues to increase. For example, CT scans are now performed with thinner parts than in the past. The number of CT slices increases the reading and interpreting time for the radiologist.

Machine learning is a powerful tool for automating medical picture analysis and diagnosis. It has the ability to relieve radiologists' workload in the radiology office. Machine learning applications in radiology include segmenting medical images (e.g., brain, spine, lungs, liver, kidney, colon), recording medical images (e.g., recording of images of organs in various modalities or time series), computer-aided recognition and diagnosis systems for CT or MRI imaging (e.g., mammography, CT, colonography, and CAD of CT pulmonary nodules), and brain function analysis.

The study of computer algorithms that can understand complicated correlations or patterns from empirical data and make exact conclusions is known as machine learning. This multidisciplinary field encompasses artificial intelligence, pattern recognition, data mining, statistics, probability theory, optimization, statistical physics, and theoretical computing. .

Machine learning algorithms may be classified into several types based on their principles. They may be classed into supervised learning, semi-supervised learning, and unsupervised learning algorithms based on how they employ training pattern identifiers.

Pathology digitization would be a nice bonus. Digital pathology is the collecting, management, exchange, and analysis of pathological data, such as slides and data, in a digital setting. Glass slides are scanned to create a high-resolution digital image that may be seen as digital slides on a computer screen or mobile device. What are the benefits of using digital pathology? There's a reason why glass slides don't work. The pathology starts with a tissue that has been removed. Even if the slides are later converted to a digital scan, glass slides are required. Today's pathology, however, is about more than just tissue and scans; it's about increasing quality, productivity, and more.

A. Consider How Current Pathology is Changing

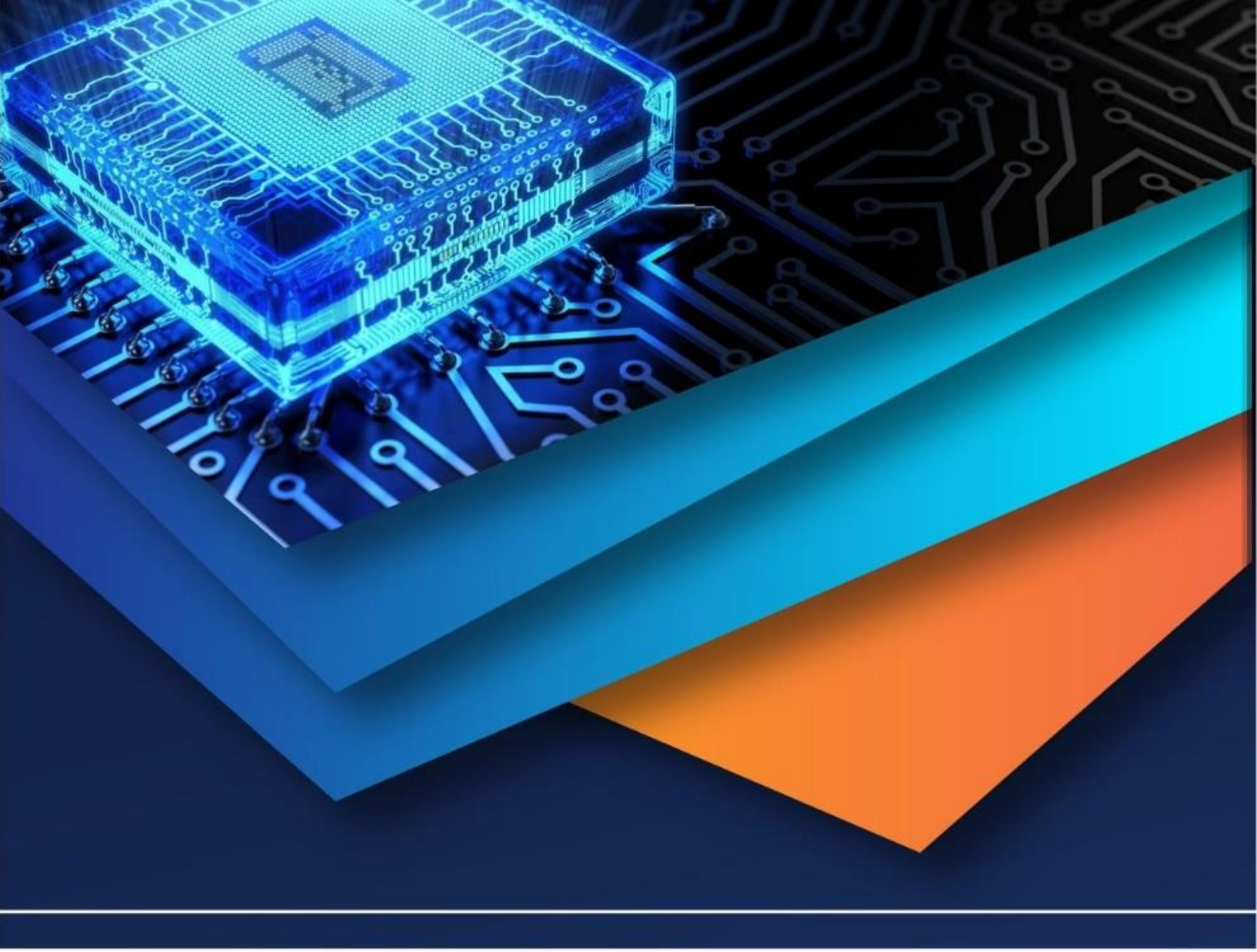
Pathologists are in short supply in the workforce, since more pathologists are retiring than joining the field.

Decision Tree is one of the predictive modelling techniques used in statistics, data mining, and machine learning, according to its definition. Decision trees are created using an algorithm that determines multiple ways to segment a data set depending on certain factors. It's a common and helpful supervised learning algorithm. Decision Trees are a non-parametric supervised learning approach that can be used for both classification and regression. Tree models in which the aim variable can take a discrete set of values are known as classification trees. Regression trees are decision trees with a continuous target variable (usually real numbers). Classification And Regression Tree is what this is called (CART).

REFERENCES

[1] J. Juntu, A. M. De Schepper, P. Van Dyck, D. Van Dyck, J. Gielen, P. M. Parizel, and J. Sijbers, Classification of soft tissue tumors by machine learning algorithms, in Soft Tissue Tumors, F. Derbel, ed. London, UK: IntechOpen, 2011, pp. 53–69
[2] G. Castellano, L. Bonilha, L. M. Li, and F. Cendes, Texture analysis of medical images, Clin. Radiol., vol. 59, no. 12, pp. 1061–1069, 2004.
[3] H. Farhidzadeh, B. Chaudhury, M. Zhou, D. B. Goldgof, L. O. Hall, R. A. Gatenby, R. J. Gillies, and M. Raghavan, Prediction of treatment outcome in soft tissue sarcoma based on radiologically defined habitats, in Proc. SPIE 9414, Medical Imaging 2015: Computer-Aided Diagnosis, Orlando, FL, USA, 2015, p. 94141U.
[4] P. D. Afonso and V. V. Mascarenhas, Imaging techniques for the diagnosis of soft tissue tumors, Rep. Med. Imaging, vol. 8, pp. 63–70, 2015.
[5] D. Komura and S. Ishikawa, Machine learning methods for histopathological image analysis, Comput. Struct. Biotechnol. J., vol. 16, pp. 34–42, 2018.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)