IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 10,Iss 06, 2021

LEVERAGING MACHINE LEARNING TO ANALYZE AND AUTOMATE COMPLEXITY IN LIPOSARCOMA DIAGNOSIS

Bommala Nirmala Devi
Associate Professor
Annamacharya College of Pharmacy, Rajampet, AP, India
bommalanirmaladevi47@gmai.com

Anudeep Kotagiri
Robotics process Automation Lead
Invictus Infotech LLC
NC, USA
anudeepkotagiri03@gmail.com

Abhinay Yada
Senior BI Engineer
LendingTree
Charlotte, NC, USA
abhinay.yada@gmail.com

Manoj Kuppam
Site Reliability Engineering Lead
Medline Industries Inc
Mundelein, IL, USA
mcmanoj2001@gmail.com

Abstract: Soft Tissue Tumors (STT) are a form of sarcoma found in tissues that connect, support, and surround body structures. Because of their shallow frequency in the body and their great diversity, they appear to be heterogeneous when observed through Magnetic Resonance Imaging (MRI) giving diverse data collection and behavior patterns. They are easily confused with other disease dataset patterns, and these diagnostic errors have a considerable detrimental effect on the medical treatment process of patients. Researchers have proposed several machine learning models to classify tumors, but none have adequately addressed this misdiagnosis problem. Also, similar studies that have proposed models for evaluation of such tumors mostly do not consider the heterogeneity and the size of the data. Therefore, we propose a machine learning-based approach which combines a new technique of preprocessing the data for features transformation, resampling techniques to eliminate the bias and the deviation of instability and performing classifier tests based on the Support Vector Machine (SVM) and Decision Tree (DT) algorithms. The tests carried out on dataset collected in Nur Hidayah Hospital of Yogyakarta in Indonesia show a great improvement compared to previous studies. These results confirm that machine learning methods could provide efficient and effective tools to reinforce the automatic decision-making processes of STT diagnostics.

Keywords: Support Vector Machine, Logistic Regression, Random Forest.

I. INTRODUCTION

The term “soft tissue” refers to tissues that support, connect, or surround other structures and organs in the body such as fat, muscles, and blood vessels, deep cutaneous tissues, nerves, and tissues surrounding the joints. As the name suggests, these are sensitive tissues that can be affected by several infections, including tumors that can develop almost anywhere in the human body. The malignant types of these tumors, also known as Soft Tissue Sarcomas (STS), are grouped together because they share many microscopic features, exhibit the same symptoms, and are almost similarly treated. Yet, effective diagnosis of Soft Tissues Tumors (STT) is still a big challenge owing to the difficulty in detecting these cancers.

Several techniques have therefore been developed to strengthen the detection of such cancers, including Magnetic Resonance Imaging (MRI) analysis. MRI is currently considered the standard diagnostic tool for the detection and classification of STT with well characterized biological properties such as cellular origins and tumor specimens used to distinguish tumors.

MRI can be used to analyze textural characteristics or other less characterized tumor characteristics several reasons: ease of computation textural characteristics, wide correlation of textural characteristics to

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 10,Iss 06, 2021

tumor pathology and robustness to changes in MRI acquisition parameters such as changes in the resolution of the tumor image and the corruption of the MRI image due to heterogeneity of the magnetic field.

Hence there is an increasing use of Machine Learning (ML) algorithms to analyze MRI images more effectively and automatically detect cancers. It has become an essential tool for modern medicine today and has been strengthened by predictive automatic learning algorithms that improve the diagnostic performance of existing expert systems. Among these many applications, we have developed a machine learning-based technique for the automotive detection and diagnosis of tumors such as STT. STT are malignant tumors that develop within tissues such as fat, muscles, nerves, fibrous tissues, and blood vessels. Because of their low frequency and the difficulty physicians have interpreting results, these challenges have prevented the development of new therapeutic agents.

In addition, the inconsistent MRI images make it difficult for physicians to determine an effective treatment Besides, STT can easily be confused with other diseases such as fibro adenoma mammae, lymphadenopathy, and struma nodosa. This diagnostic failure has a significant impact on the patient treatment process.

EXISTING SYSTEM

Due to a lack of knowledge about data visualization, it is a bit difficult to deploy machine learning algorithms in the current system. In the current approach, constructing models is done by mathematical computations, which can be very difficult and time-consuming. We employ machine learning tools from the Scikit-Learn toolkit to get around all of this.

Disadvantages

- High complexity.
- Time consuming.

II. Proposed System

Many machine learning models have been put out to categorise tumours, but none have sufficiently addressed the issue of incorrect diagnoses. Moreover, comparable research that have suggested methods for evaluating these tumours typically do not take the heterogeneity and magnitude of the data into account. Because of this, we suggest a machine learning-based strategy that incorporates a novel method of preprocessing the data for feature transformation, resampling approaches to remove bias and the divergence of instability, and running classifier tests based.

Advantages

- Highest Accuracy
- Reduces Time Complexity

A. IMPLEMENTATION

Step :1 First we have taken the Soft Tissue Tumors of CSV file.

Step :2 Load the dataset into work environment and made a check for null values if any.

Step :3 After checking the null values split the data in to train data and test data.

Step :4 After splitting apply the algorithm and fit the train data and test data.

Step :5 We got the best accuracy score for Logistic Regression.

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 10,Iss 06, 2021

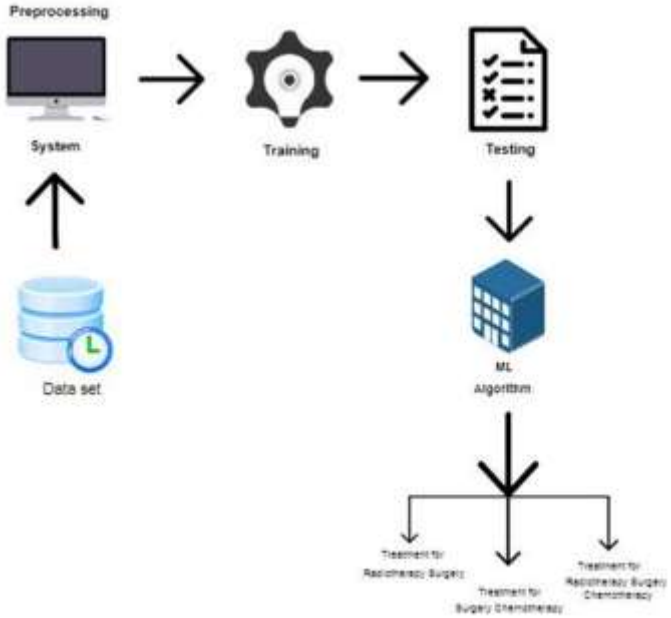
Step :6 Later, the entire work is done with flask framework

Step :7 User can view the home, about, login, register, login homepage, upload data, view data, training,

detection and logout pages.

Step :8 After detection, we will get to know that whether a patient need which type Treatment for Soft Tissue Tumors.

ARCHITECTURE:



III. ALGORITHMS

A. Logistic Regression:

One of the most often used Machine Learning algorithms, within the category of Supervised Learning, is logistic regression. Using a predetermined set of independent factors, it is used to predict the categorical dependent variable. In a categorical dependent variable, the output is predicted via logistic regression. As a result, the result must be a discrete or categorical value. Rather of providing the exact values of 0 and 1, it provides the probabilistic values that fall between 0 and 1. It can be either Yes or No, 0 or 1, true or false, etc. With the exception of how they are applied, logistic regression and linear regression are very similar. While logistic regression is used to solve classification difficulties, linear regression is used to solve regression problems.

In logistic regression, we fit a "S" shaped logistic function, which predicts two maximum values, rather than a regression line. The logistic function's curve shows the possibility of several things, including whether or not the cells are malignant, whether or not a mouse is obese depending on its weight, etc. Because it can classify new data using both continuous and discrete datasets, logistic regression is a key machine learning approach. When classifying observations using various sources of data, logistic regression can be used to quickly identify the factors that will work well.

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 10,Iss 06, 2021

The logistic function is displayed in the graphic below. A dataset is provided that includes data on various users collected from social networking sites. A new SUV vehicle was just introduced by an automobile manufacturer. The business therefore needed to determine how many consumers in the dataset were interested in buying a car. Using the logistic regression approach, we will create a machine learning model for this issue. The graphic below displays the dataset. We will use age and salary to forecast the purchased variable in this problem.

B. Support Vector Classifier:

SVMs are capable of handling both classification and regression issues. The decision boundary for this method's hyperplane needs to be defined. A decision plane is required to divide a collection of objects into their many classes. If the objects cannot be separated linearly, kernels—complex mathematical functions—must be used to separate the objects that belong to various classes. The goal of SVM is to correctly identify the objects using examples from the training data set. These are some benefits of SVM: It can manage structured and semi-structured data, and if the right kernel function can be determined, it can manage complex functions.

Less likelihood of overfitting exists since SVM adopts generalization. With large-scale data, it can scale up. It does not become trapped in regional optimum. The following are SVM's drawbacks: due to the longer training times required for large data sets, its performance suffers. Finding an adequate kernel function will be challenging. SVM performs poorly when the dataset is noisy. SVM doesn't offer probabilities in its output. It's challenging to comprehend the final SVM model.

Support The practical use of vector machines includes text classification, handwriting identification, face detection, credit card fraud detection, and cancer diagnosis. Therefore, the first technique to try will be the logistic regression approach, followed by the decision trees (Random Forests) to see if there is a noticeable improvement. The third approach to try is the SVM strategy. SVM can be tested when there are lots of observations and features.

C. Random Forest:

A random forest is a machine learning method for tackling classification and regression issues. It makes use of ensemble learning, a method for solving complicated issues by combining a number of classifiers.

In a random forest algorithm, there are many different decision trees. The random forest algorithm creates a "forest" that is trained via bagging or bootstrap aggregation. The accuracy of machine learning algorithms is increased by bagging, an ensemble meta-algorithm.

Based on the predictions of the decision trees, the (random forest) algorithm determines the result. It makes predictions by averaging or averaging out the results from different trees. The accuracy of the result grows as the number of trees increases.

The decision tree algorithm's shortcomings are eliminated with a random forest. It improves precision and decreases dataset overfitting. Without requiring numerous configurations in packages, it generates forecasts (like Scikit-learn).

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 10,Iss 06, 2021

IV. EXPERIMENTAION RESULTS

V. CONCLUSION

Applications in many fields, including medicine, can benefit from high precision calculations enhanced by ML algorithms. The performance of computer-aided diagnostic systems has improved greatly thanks to these technologies in recent years, but integrating them continues to be difficult for contemporary healthcare organizations. Based on data gathered from the Nur Hidayah Hospital in Bantul, Yogyakarta, Indonesia, we built a solid and realistic model in this study that enables automatic predictive classification of STT and non-STT. After including a fresh data pretreatment method, we contrasted the SVM and LR classifiers. This comparison revealed that the LR model is substantially more sensitive to the number of variables than the SVM model, even though the LR algorithm is marginally more efficient than the SVM algorithm.

VI. REFERENCES

[1]. F. Collin, M. Gelly-Marty, M. B. N. Binh, and J. M. Coindre, "Sarcomas of the Muscle Tissue: Current Anatomicopathological Reports," Cancer/Radiotherapy, vol. 10, nos. 1 & 2, pp. 7–14, 2006.

[2]. Biological characterization of soft tissue sarcomas, Annals of Translational Medicine, vol. 22, no. 3, p. 368, 2015. T. Hayashi, A. Horiuchi, K. Sano, Y. Kanai, N. Yaegashi, H. Aburatani, and I. Konishi.

[3]. S. L. Salzberg, book review of J. Ross Quinlan's C4.5: Programs for Machine Learning. Machine Learning, vol.16, no. 3, 1993, pp. 235–240, Morgan Kaufmann Publishers, Inc.

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 10,Iss 06, 2021

[4]. Inability of humans to differentiate between visual textures that agree in second-order statistics—Revisited, Perception, vol. 2, no. 4, pp. 391-405, 1973. B. Julesz, E. N. Gilbert, L. A. Shepp, and H. L. Frisch.

[5]. Prediction of treatment outcome in soft tissue sarcoma based on radiologically defined habitats, Proc. SPIE 9414, Medical Imaging 2015: Computer-Aided Diagnosis, Orlando, FL, USA, 2015, p. 94141U. H. Farhidzadeh, B. Chaudhury, M. Zhou, D. B. Goldgof, L. O. Hall, R. A. Gatenby, R. J. Gillies.

[6]. Imbalanced client classification for bank direct marketing, G. Marinakos and S. Daskalaki, J. Mark. Anal., vol. 5, no. 1, pp. 14–30, 2017.

[7]. Application of improved decision tree approach based on rough set in developing smart medical analysis CRM system, International Journal of Smart Homes, vol. 10, no. 1, pp. 251-266, 2016. H. S. Xu, L. Wang, and W. L. Gan.

[8]. Asymptotic behaviour of support vector machines with Gaussian kernels, Neural Comput., vol. 15, no. 7, pp. 1667–1689, 2003. S. S. Keerthi and C. J. Lin.

[9]. Infinite-limits for Tikhonov regularisation, by R. A. Lippert and R. M. Rifkin, J. Machine Learning Research, vol. 7, pp. 855-876, 2006.

[10]. C. G. L. Guillou, Tumeurs des tissus mous: Role of the pathologist in the diagnostic approach, Review of Medical Switzerland, vol. 3, p. 32473, 2007.

[11]. Manoj Kuppam, Anudeep Kotagiri, Abhinay Yada, "Revolutionizing Attendance Management with Automated Face Recognition Technology" ISSN PRINT 2319 1775, Vol.10, issue 02,2021.