2021 8th International Conference on Signal Processing and Integrated Networks (SPIN) | 978-1-6654-3564-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/SPIN52536.2021.9566061

Machine Learning approach to classify and predict different Osteosarcoma types

Sanket Mahore
Dept. of Instrumentation and Control
College of Engineering Pune
Pune, India
sanketam19.instru@coep.ac.in

Kalyani Bhole
Dept. of Instrumentation and Control
College of Engineering Pune
Pune, India
kab.instru@coep.ac.in

Shashikant Rathod
Dept. of Instrumentation and Control
College of Engineering Pune
Pune, India
rsr18.instru@coep.ac.in

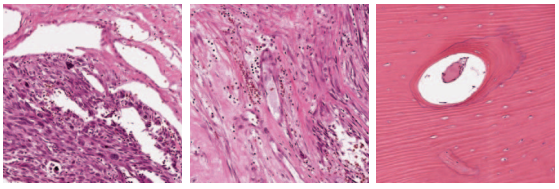
***Abstract*—Family physicians rarely see a malignant bone cancer because it is hard to find, and most of the time, bone cancer is benign. It is very time-consuming and complicated for the pathologist to classify Osteosarcoma histopathological images. Typically Osteosarcoma classifies into viable, Non-viable, and Non-tumor classes, but intra-class variation and inter-class similarity are complex tasks. This paper used the Random Forest(RF) machine learning algorithm, which efficiently and accurately classifies Osteosarcoma into Viable, Non-viable, and Non-tumor classes. The Random Forest method gives a classification accuracy of 92.40%, a sensitivity of 85.44%, and specificity 93.38% with AUC=0.95.**

***Index Terms*—Random Forest, Osteosarcoma, Machine Learning**

I. Introduction

Osteosarcoma is a malignant bone cancer tumor mainly found in cells that form Bone. Osteosarcoma grows in long bones near metaphyseal growth plates, commonly found in Humerus(10%), Tibia(19%), and femur(42%) [1].The cause of cancer is unknown, but it shows that Osteoblasts that form Bone get under mutation and develop new DNA sequences. For osteosarcoma diagnosis, Hematoxylin and Eosin(H&E) staining method use which color nuclei blue and tissue as pink in slides [2]. Evaluation of digital whole slides images(WSIs) done under microscope manually by the pathologist. During staining, normal cells and tumor cells stain in blue color, but their shape is irregular. Our experiment divides such cells into three categories such as 1)Viable, 2)Non-Viable, and 3)Non-Tumor. A viable tumor means the cancerous element is still present in a cell, while a Non-viable tumor means the cancerous element destroy. Non-Tumor cell means a cell which is not affected by cancerous element and have a proper cellular structure. non-tumor cells include Bone and cartilage, as shown in Fig. 1.

For classification and prediction, different Machine Learning, as well as Deep Learning methods, are used, such as Artificial Neural Network(ANN), Convolutional Neural Network(CNN), Recurrent Neural Network(RNN), K-Nearest Neighbor(KNN), Support Vector Machine(SVM), Random Forest(RF), Decision Tree(DT).



(a) Viable (b) Non-Viable (c) Non-Tumor

Fig. 1: Example showing the complexity of Histopathological images (a) Viable(Nuclei is densely aggregate together) (b) Non-viable(Necrosis, Fibrosis, and Osteoid) and (c)Non-Tumor (Bone, Cartilage)

For prostate cancer and Breast cancer detection CNN is beneficial with 99% sensitivity [2]. A researcher from SJTU(Shanghai Jiao Tong University) Team used a metabolic dataset of Osteosarcoma for classification where researchers used LR(Logistic regression, SVM(Support Vector Machine), and RF(Random Forest). For this specific metabolic dataset, get overall accuracy of 95% for RF model [3]. The author, Arunachalan et al. uses the K-means clustering technique uses along with the multi-segmentation otsu threshold technique to classify into the viable and non-viable Osteosarcoma, which gives an accuracy of 95.5% [4].

For analysis of histopathological images of Osteosarcoma, Rashika Mishra and co-researchers used DCNN having seven layers which give higher accuracy of 92% for classification [5]. Adding more parameters in the convolutional layer for feature extraction and the last two fully connected layers boost the performance using data augmentation [6]. A fully convolutional layer combines with multiple supervised side layers at the output side, which gives an average sensitivity of 86.66% and F1-measure 0.908 [7]. For region and object detection, Faster R-CNN is very famous, having two parts where Region Proposal Network (RPN) give region and detector network gives object with an accuracy of 0.97 [8]. DS-Net is another customize architecture of CNN where the deep model combines with the Siamese network [9] composed of Auxiliary Supervision Network (ASN) and Classification Network with an average accuracy of 95.1% [10]. Olivia Alge et al. used RNA-seq and X-ray database features [11] to train the RF

model, which classifies Benign and Osteosarcoma tumor with an accuracy of 96%, the sensitivity of 100% and specificity of 33% using three-fold-cross validation. [12].

Our paper used the Random Forest Machine Learning algorithm for the classification and prediction of Osteosarcoma. The paper's organization is as follows: Section II gives a brief idea about Feature selection, Random forest ML model, and Evaluation matrix. Section III shows our experimental results and comparative analysis with previous studies. Section IV gives the conclusion of our research work.

II. Methodology

A. Block Diagram

Our research flow is represented by a block diagram as shown in Fig.2, which includes Dataset, feature extraction, Pre-processing, machine learning model, and performance evaluation metrics. The data collected from TCIA [14] public repository and two types of features are extracted from images. Training and testing of ML model done by splitting data into train and test sets. To evaluate the performance of the ML model, we used performance metrics.

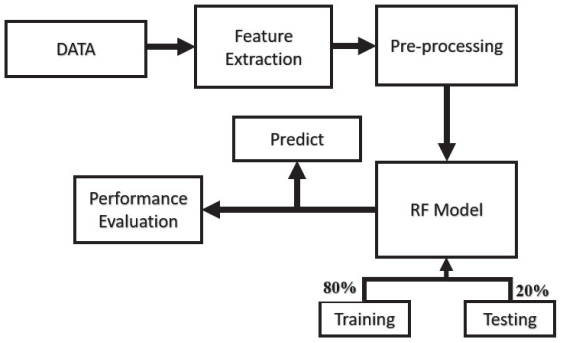


Fig. 2: Our Random Forest Machine Learning Model block diagram

B. Dataset

We used the open-source Osteosarcoma Histopathological Image dataset, created by the Research Team of the University of Texas, Dallas(U.S) [13] and available on a public repository, i.e., The Cancer Imaging Archive (TCIA) [14]. The researcher from UT, Dallas, used Hematoxylin and Eosin (H&E) stain histopathological images, which are observed under the microscope and converted into digital WSI(Whole Slide Image). The H&E stain gave Blue color to nuclei and pink color to tissue. From microscopic evaluation by an expert pathologist, they classify images into three types, i.e., Non-Tumor, Viable and Non-Viable. Studying WSI of OST, two types of features is extracts from OST images, i.e., expert-guided features and Cellprofilier features. We use both types of features to train and test the ML model. The dataset is split into 80% for training and 20% for testing.

C. Feature Extraction

Features are observable, and individual measurable property use to classify data. It may be statistical features(Variance, Mean, Correlation) or pathological features(Nuclei count, Texture, Number of a cell). Similarly, for Osteosarcoma Histopathological Images, the two forms of features extracted from the OST Images, i.e., Expert-guided features and Cellprofilier features [13]. Expert-guided features are the cell's observable property needed by a pathologist to analyze images and extract them using in-house software. Expert-guided features are Total clusters, Red & Blue count, Red & Blue percentage, Circularity and Area [10]. While CellProfiler features extracts using Cellprofiler Software [15] which include Entropy, Variance, Gabor. We use both types of features to categorize Osteosarcoma into three categories, i.e., Viable, Non-Viable, and Non-Tumor.

TABLE I: List of Expert-guided features and Cellprofilier features extracted from OST images

Feature type	Feature names
Expert-Guided	Total clusters,Red & Blue count Red & Blue percentage, Circularity and Area.
Cellprofiler	Gabor, sum Entropy, sum variance, variance sum average, correlation and angular second moment.

D. Pre-processing Step

We used the label binarize [17] from sklearn. preprocessing library, where Non-tumor, Viable, and Non-Viable classes are denoted as an integer like 0, 1, and 2, respectively. Converting string values into numeric form is readable for the machine learning algorithm. It is a significant pre-processing step while training supervised machine learning algorithms.

E. Random Forest Algorithm

Leo Breiman proposed Random Forest in 1999 [18], an ensemble technique used for Classification and Regression, and uses a Decision Tree(DT) as the base classifier [19]. Random forest train via Bagging method where the multi-decision tree gets random data in Row Sampling and Feature Sampling. This process is called feature bagging, which uses to train and test sub-tree and gives the result. The result depends on the majority of votes given by DT then RF predicts the output. The two criteria are uses to choose the root of the RF model from the feature sample, i.e., Gini index and Information gain. We use the Gini index criterion to choose the vital feature as the root. The highest value of the Gini index where choose as the root of the RF model. The Gini index is calculated by

$$G = \sum_{i=1}^{C} p(i) * (1 - p(i)) \tag{1}$$

where,
C= Number of classes
p(i)= Probability of choosing the data from i class

The error rate of the random forest depends on two things, i.e., Correlation between two classes and the strength of each

class. Increasing the correlation value higher be the chances of Error rate, and increasing the value of strength decreases the error rate. Following the Algorithm I give a brief idea about the Random forest algorithm.

Algorithm 1 Pseudo code for Random Forest

H

Training set $T = (p_1, q_1), ..., (p_n, q_n)$, Features F, and D is number of Decision Tree
function RandomForest(T, F)
    $H \leftarrow \theta$
for $i \epsilon 1, \ldots, B$ **do**
    $T^{(i)} \leftarrow$ A bootstrap sample from S
    $h_i \leftarrow$ Randomized Tree Learn($T^{(i)}, F$)
    $H \leftarrow H \cup h_i$
end for
return H
end function
Randomized Tree Learn(T, F)
    At each node:
    $f \leftarrow$ very small subset of F spilt on best feature in f
return The tree learn

F. Evaluation

The performance metrics like Accuracy, Sensitivity, and Specificity, ROC curve (receiver operating characteristic curve) were used to evaluate ML models based on confusion matrix.

A confusion matrix is a table used to describe the classifier's performance on a training and testing dataset where the actual value is known.

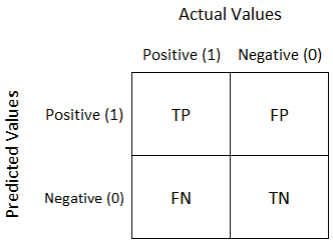


Fig. 3: Confusion Matrix

The confusion matrix is a tabular plot between predicted value versus actual values as shown in Fig.3, where TP= True positive, FP=False positive, FN=False negative, and TN=True negative.

Accuracy is defined using how our model predicts True positive and True negative values. Below formula used to calculate the accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

Sensitivity is the ability to detect True Positive values from the given set of samples. Mathematically the Sensitivity calculate using

$$Sensitivity = \frac{TP}{TP + FN} \tag{3}$$

Specificity detects True Negative values, which are calculated by the given formula

$$Specificity = \frac{TN}{TN + FP} \tag{4}$$

ROC curve (receiver operating characteristic curve) [20] is another performance evaluation metric that represents the Sensitivity vs. (1-Specificity) plot. The AUC(Area Under Curve) values range between 0 and 1. The AUC=0 means model predicts 100% wrong values, and AUC=1 stated that the model predicts 100% correct values [21]. The above performance matrix parameter gives a detailed statistical evaluation of our RF model.

III. Results and Discussion

A. RF Model Analysis

For training and testing the ML model, we use 80% data for training and 20% for testing. We observe the training and testing performance by a confusion matrix.

Fig. 4 shows the confusion matrix for the training set and testing set. For training set, RF model correctly classify osteosarcoma type i.e Class0= 417, Class1= 249 and Class2= 201. Similarly, for the testing set Class0=101, Class1=64, and Class2=36 samples correctly.

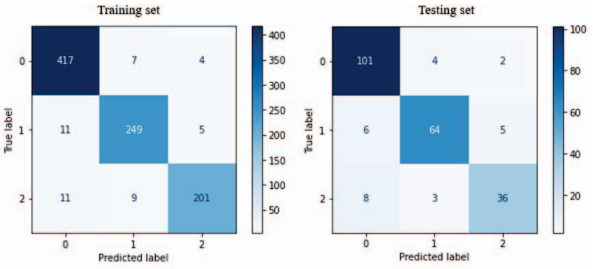


Fig. 4: Confusion matrix of training and testing set

TABLE II shows training results, where the accuracy of Class 0 is 96.40% with sensitivity and specificity of 97.40% and 95.50%, respectively. For Class 1, accuracy is 96.40%, sensitivity 93.20%, and specificity 97.50%. Where Class 2 gets an accuracy of 96.80%, a sensitivity of 90.90%, and specificity of 93.38%. The average accuracy, sensitivity, and specificity of the RF model for training are 96.53%, 93.83%, and 97.23%, respectively.

TABLE II: Class-wise performance for training set

Attributes	Class 0	Class 1	Class 2	Average
Accuracy	96.40%	96.40%	96.80%	96.53%
Sensitivity	97.40%	93.20%	90.90%	93.83%
Specificity	95.50%	97.50%	98.70%	97.23%

TABLE III shows a class-wise testing performance of the RF model, which shown outstanding results. The accuracy of class 0 is 91.86%, sensitivity 94.40%, and specificity 88.53%. For class 1 accuracy is 92.64%, sensitivity=85.33% and specificity 95.45%. Class 0 shows higher sensitivity as compared to class 1 because of data imbalance. Class 2 accuracy is 92.64%, and specificity 96.15%. Class 2 gives better results in terms of accuracy and specificity as compared to class0 as well class1. The sensitivity of class 2 is 76.59%, which is acceptable.

TABLE III: Class-wise performance for testing set

Attributes	Class 0	Class 1	Class 2	Average
Accuracy	91.86%	**92.64%**	**92.64%**	92.40%
Sensitivity	**94.40%**	85.33%	76.59%	85.44%
Specificity	88.53%	95.45%	**96.15%**	93.38%

The AUC value of class 0 is 0.960, which is higher than class 1 and class 2. Class 1 and class 2 shows the AUC values of 0.949 and 0.935, respectively. The overall performance of the RF model is good with accuracy 92.40%, sensitivity 85.44%, specificity 93.38%, and AUC score of 0.948.

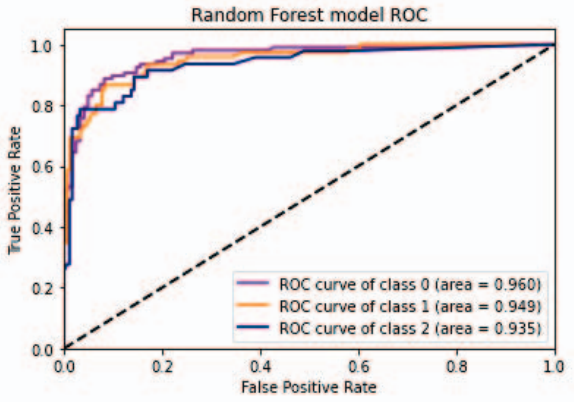


Fig. 5: Receiver Operating Characteristic (ROC) curve and AUC (Area Under Curve)

B. Comparative Analysis

Proposed RF model performance is very well compared to the previous study, as shown in TABLE IV.

TABLE IV: Comparative analysis with exiting studies, where HPI= Histopathological Image data

Models	Database	ACC	SEN	SPC
CNN [5]	HPI data [13]	**92%**	**94%**	-
VGGNet [5]	HPI data [13]	84%	84%	-
AlexNet [5]	HPI data [13]	73%	75%	-
LeNet [5]	HPI data [13]	67%	67%	-
Complex tree [10]	HPI data [13]	81%	-	-
SVM [10]	HPI data [13]	90%	-	-
Ensemble Learners [10]	HPI data [13]	87%	-	-
Our RF Model	HPI data [13]	**92.40%**	85.44%	**93.38%**

Researchers use Histopathological Image data [13] to train dominant Deep learning algorithms like VGGNet, CNN, AlexNet, LeNet to classify Osteosarcoma where they achieved outstanding results. The proposed 9-layer Convolutional Neural Network by [5] has an input image size of (128*128*3) where (128*128) is (width*height) and 3 is the RGB channel of an image. For downsampling, the MaxPooling layer uses, and the final output gives to Fully-connected layer. The final layer gives an accuracy of 92% with a sensitivity of 94% for the image dataset. Even pre-trained models like VGGNet, AlexNet, and LeNet give accuracy of 84%, 73%, and 67% with the sensitivity of 84%, 75%, and 67%, respectively.

Arunachalam et al. use machine learning algorithms like Complex tree, SVM(Support Vector Machine), and Ensemble Learners, giving a better result than pre-trained DL models. The accuracy of the SVM model is 89.9%, where CT=80.9% and Ensemble learner=86.8%. The SVM gets higher Accuracy than Complex tree and Ensemble learner [10].

We choose Random Forest because the base learner of RF is the Decision tree which reduces overfitting and improves the model's accuracy. The optimization of the hyperparameters is carriying out before training the RF model to increase the model performance. The hyperparameters like max_depth=9, n_base estimator=8 were optimized. The proposed Random Forest model achieves 92.40% accuracy, which is high compared to previous Deep Learning and Machine learning models except for CNN. Proposed CNN model [5] gives 92% accuracy where the RF model gets an accuracy of 92.40% which slightly greater. TABLE IV shows our RF model's comparative analysis with the previous study, where our RF model outperforms all DL and ML models except CNN.

IV. Conclusion

This paper has used a Random Forest Machine Learning algorithm to classify and predict different Osteosarcoma types. RF model train and test using features extracted from histopathological images. Two types of features are used to train and test the RF model, i.e., Expert-guided and Cellprofilier features. The result obtains in the present study shows that the Random Forest algorithm classifies the dataset very efficiently and categorizes Osteosarcoma into different classes. The final result demonstrates that the proposed RF model classifies the dataset very well. Our RF model gives an Accuracy of 92.33%, sensitivity of 86.66%, Specificity 93.66%,

and AUC 0.963. The AUC result shows that the proposed RF model correctly predicts the class of Osteosarcoma cancer.

ACKNOWLEDGMENT

The authors would like to thanks the College of Engineering, Pune, for mentoring throughout the project. We are thankful to UT Southwestern/UT Dallas for their Open-source Osteosarcoma Histopathological Image dataset.

REFERENCES

[1] Jerome, T.J., Varghese, M., Sankaran, B., Thomas, S. and Thirumagal, S.K., 2009. Tibial Chondroblastic Osteosarcoma—Case Report. Foot and ankle surgery, 15(1), pp.33-39.

[2] Litjens, G., Sánchez, C.I., Timofeeva, N., Hermsen, M., Nagtegaal, I., Kovacs, I., Hulsbergen-Van De Kaa, C., Bult, P., Van Ginneken, B. and Van Der Laak, J., 2016. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. Scientific reports, 6(1), pp.1-11.

[3] Li, Z., Soroushmehr, S.R., Hua, Y., Mao, M., Qiu, Y. and Najarian, K., 2017, July. Classifying osteosarcoma patients using machine learning approaches. In 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 82-85). IEEE.

[4] Arunachalam, H.B., Mishra, R., Armaselu, B., Daescu, O., Martinez, M., Leavey, P., Rakheja, D., Cederberg, K., Sengupta, A. and NI'SUILLEABHAIN, M.O.L.L.Y., 2017. Computer aided image segmentation and classification for viable and non-viable tumor identification in osteosarcoma. In Pacific Symposium on Biocomputing 2017 (pp. 195-206).

[5] Mishra, R., Daescu, O., Leavey, P., Rakheja, D. and Sengupta, A., 2018. Convolutional neural network for histopathological analysis of osteosarcoma. Journal of Computational Biology, 25(3), pp.313-325.

[6] Mishra, R., Daescu, O., Leavey, P., Rakheja, D. and Sengupta, A., 2017, May. Histopathological diagnosis for viable and non-viable tumor prediction for osteosarcoma using convolutional neural network. In International Symposium on Bioinformatics Research and Applications (pp. 12-23). Springer, Cham.

[7] Huang, L., Xia, W., Zhang, B., Qiu, B. and Gao, X., 2017. MSFCN-multiple supervised fully convolutional networks for the osteosarcoma segmentation of CT images. Computer methods and programs in biomedicine, 143, pp.67-74.

[8] D'Acunto, M., Martinelli, M. and Moroni, D., 2018, September. Deep learning approach to human osteosarcoma cell detection and classification. In International Conference on Multimedia and Network Information System (pp. 353-361). Springer, Cham.

[9] Fu, Y., Xue, P., Ji, H., Cui, W. and Dong, E., 2020. Deep model with Siamese network for viable and necrotic tumor regions assessment in osteosarcoma. Medical Physics, 47(10), pp.4895-4905.

[10] Arunachalam, H.B., Mishra, R., Daescu, O., Cederberg, K., Rakheja, D., Sengupta, A., Leonard, D., Hallac, R. and Leavey, P., 2019. Viable and necrotic tumor assessment from whole slide images of osteosarcoma using machine-learning and deep-learning models. PloS one, 14(4), p.e0210706.

[11] Rebecca Shen, Zhi Li, Linglin Zhang, Yingqi Hua, Min Mao, Zhicong Li, Zhengdong Cai, Yunping Qiu, Jonathan Gryak, and Kayvan Najarian. Osteosarcoma Patients Classification Using Plain X-Rays and Metabolomic Data. In 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pages 690–693, Honolulu, HI, July 2018. IEEE.

[12] Alge, O., Lu, L., Li, Z., Hua, Y., Gryak, J. and Najarian, K., 2020, July. Automated Classification of Osteosarcoma and Benign Tumors using RNA-seq and Plain X-ray. In 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 1165-1168). IEEE.

[13] Leavey, P., Sengupta, A., Rakheja, D., Daescu, O., Arunachalam, H. B., & Mishra, R. (2019). Osteosarcoma data from UT Southwestern/UT Dallas for Viable and Necrotic Tumor Assessment [Data set].

[14] Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M. and Tarbox, L., 2013. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. Journal of digital imaging, 26(6), pp.1045-1057.

[15] McQuin, C., Goodman, A., Chernyshev, V., Kamentsky, L., Cimini, B.A., Karhohs, K.W., Doan, M., Ding, L., Rafelski, S.M., Thirstrup, D. and Wiegraebe, W., 2018. CellProfiler 3.0: Next-generation image processing for biology. PLoS biology, 16(7), p.e2005970.

[16] Zhu, Y., Ting, K.M. and Angelova, M., 2018, June. A distance scaling method to improve density-based clustering. In Pacific-Asia Conference on Knowledge Discovery and Data Mining (pp. 389-400). Springer, Cham.

[17] Mol, P.R. and Mary, C.I., 2021. Classification of Network Intrusion Attacks Using Machine Learning and Deep Learning. Annals of the Romanian Society for Cell Biology, pp.1927-1943.

[18] Breiman, L., 2001. Random forests. Machine learning, 45(1), pp.5-32.

[19] Abdulkareem, N.M. and Abdulazeez, A.M., 2021. Machine learning classification based on Radom Forest Algorithm: A review. International Journal of Science and Business, 5(2), pp.128-142.

[20] Bradley, A.P., 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern recognition, 30(7), pp.1145-1159.

[21] Krupinski, E.A., 2021. Evaluating AI Clinically—It's Not Just ROC AUC!.