Improved Bone Cancer Diagnosis: Transfer Learning Integration with U-Net for Segmented Histopathology Image Analysis

Deepak Kumar K
Department of CSE
Rajalakshmi Engineering College
Chennai, India
kdeepak.srmit@gmail.com

Senthil Pandi S
Department of CSE
Rajalakshmi Engineering College
Chennai, India
mailtosenthil.ks@gmail.com

Kumar P
Department of CSE
Rajalakshmi Engineering College
Chennai, India
kumar@rajalakshmi.edu.in

Madhulika G
Department of CSE
Rajalakshmi Engineering College
Chennai, India
210701139@rajalakshmi.edu.in

Mahalakshmi K
Department of CSE
Rajalakshmi Engineering College
Chennai, India
210701143@rajalakshmi.edu.in

***Abstract*—Bone cancer, while rare, poses significant health challenges with often late diagnoses. Though CNN, proposed as a good algorithm for classifying and detecting cancer from images, it faces challenges when it is a histopathological image. This study introduces a novel approach combining machine learning with cutting-edge image processing methods to increase the precision and effectiveness of diagnosis. The model retains several key features. Utilizing U-Net for segmentation and four different CNNs, ResNet50, VGG16, MobileNetV2, and DenseNet121 for classification, along with data augmentation strategies, our method aims to enhance the reliability of bone cancer detection. Preliminary results demonstrate improved performance in accuracy and speed compared to traditional methods.**

Keywords—Bone Cancer, Machine Learning, U-Net, ResNet50, VGG16, MobileNetV2, DenseNet121, Image Processing, Data Augmentation.

I. INTRODUCTION

Bone Cancer is also called Osteosarcoma. The multiplication of osteoblastic cells helps in formation of bone. When this happens rapidly it leads to the onset of malignant disease bone cancer [1]. The symptoms of the disease include pain and fracture might happen when the bone becomes weak. It is detected that the majority of the formation of cancerous cells forms at the knee in the case of this type of cancer. The diagnosis of the disease is usually done with X Rays. In the case of tumour, a surgery called biopsy is done to examine how far the disease has affected the cells and damage in the cell is identified from the small tissue sample extracted from the body. The extracted tissue sample is processed for the staining stage .Natural black 1 and eosin are usually used for this stage of the study or examination of the tissue for cells that are cancerous. After staining, the process of diagnosis requires a equipped pathologists for examining. It also depends on the microscope used. The process is prone to errors due to external factors and it is time consuming [2]. The microscopic examination of tissue is difficult and it provides limited information about the pattern, key patterns might be missed, not appropriate for a very large amount of sample.

The diagnosis process can be improved with help of advanced machine learning algorithms in terms of precision and effectiveness. Deep neural networks are widely employed in the applications of processing and classifying images. The image processing [3] followed by classification is majorly based on the data from which it learns. Image for processing of bone cancer is not much available and it is difficult to get real time data. We need to approach the hospital and get the permission of the patient and hospital, which requires confidentiality and time. The efficiency of the model depends on the data, it needs to be preprocessed and duplicated if the data is less. The technique of data augmentation has been employed to expand the data, which improves the model's performance. It also enables the model to pick up on various aspects of the image thus making sure that key features are not missed. For classification the data should be segregated as a separate folder as classes.

The histopathology image is a high resolution image and the algorithm used should be able to decide the global features of the image. Segmentation of the images [4] in this study is done using the U-Net architecture, which is helpful in determining the feature of interest. The segmented dataset is then classified using four distinct convolutional neural network architectures: These are ResNet50, VGG16, MobileNetV2 and DenseNet121.Implementing the chosen data augmentation techniques the model improves its insensitivity to the specific data distribution and handles the class imbalance problem [5]. The incorporation of these models greatly helps in the enhancement of the detection and identification of the cancerous cells within the images so it will reduce time in terms of classification speed and gains time in diagnosis deeply the steps that are involved in the approach proposed in this paper are as follows:

The steps that are involved in the approach proposed in this paper are as follows:

979-8-3315-3496-7/24/$31.00 ©2024 IEEE

- Data Collection
- Data Augmentation
- Feature Extraction
- Model Training
- Performance measure

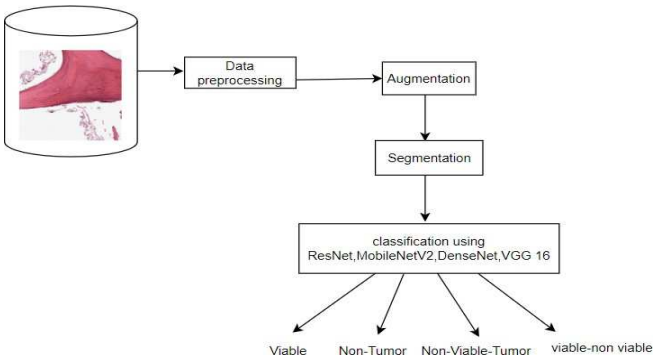

Figure. 1. Flow Diagram

II. Literature Survey

Jasper Gnana Chandran, J et al. [6] focused on identifying cancer type that occurs within the marrow of bone by following deep learning techniques which include data synthesis using augmentation by taking out the edges, features, rotating the picture and finally classified using three layers. They proposed dense convolutional neural network for determining cancer.

Walid et al. [7] conducted in order to build a model for the hard-voting, unbiased detection of bone cancer. The highest kappa and accuracy scores are 87.55% and 92.14% for MobileNetV2, respectively. Performing better, with an accuracy percentage of 93.88%, is the Max voting classifier.

Wu et al. [8] discussed about the accurate detection of nuclei in osteosarcoma using pathological images by considering the differences in styles of the stain and improving the diagnosis in underdeveloped regions with the help of conv-transformer architecture called ENMVit.

Loraksa et al [9]. proposed a module for detecting lung nodules in osteosarcoma using SSD-VGG16 which detects the nodules by constructing the bounding box around them and calculating the confidence scores with accuracy of 75.97%. The proposed model could not detect the lung nodules when the size of the nodules in the image is very tiny or blur.

Saber et al. [10] proposed a model for breast cancer classification using a pre-trained network combined with CNN by performing comparison with other four models, VGG-16 achieved high accuracy. The data is enhanced with the help of augmentation, fine tuning and cross validation is done to overcome overfitting.

Sun et al. [11] proposed a model where the histopathological images are segmented into patches using transfer learning before image level features are extracted using MIL and classified the liver cancer images into normal and abnormal. They used the approach of selecting and sorting the k top and bottom features for classification.

Weng et al. [12] employed NAS in medical image segmentation and in the NAS-Unet model, two supporting cells which are downsampling (DownSC) and the upsampling (UpSC) cells improve the model. To address issues related to small batch sizes, the architecture applies group normalization besides using a DAG-based mixed operation strategy for architecture search. During training, binary gating occurs by changing two paths at once in order to cut the GPU memory usage in half.

Zeng et al. [13] improved the actual efficiency in segmenting diseased cells, residual blocks, multi-scale inception modules, and channel attention modules were incorporated into the planned RIC-Unet architecture. This structure handles overlapping and various cell topologies by using both local and global context. This model uses focal loss to assist it focus on hard samples during the training phase and achieve joint nuclei and contour segmentation.

Fu et al. [14] created a subnetwork of the multifunctional spatial focused attention module, with an encoder-decoder core serving as its foundation. This model utilizes PET-spatial attention maps that guides tumor localization in CT feature maps, enhancing segmentation by integrating spatial information from both PET and CT modalities.

Ashwath et al. [15] combined three branches which were targeted for determining the type of medical images with scattered and irregular regions, a three-tier self-interpretable DCNN model was presented. Three branches are used: a global subdivision that uses the entire image to extract patterns, an attention subdivision that highlights important regions, along with a fusion subdivision that combines information from both to improve categorization.

III. Methodology

The methodology entails data collecting and preprocessing to create histopathology images for training and assessment. Datasets for the purpose of detecting bone cancer were obtained from publicly accessible sources that contained tagged images. For consistency in input size and format, the photos were preprocessed using techniques including scaling and normalization. In order to reduce overfitting and broad the amount of training data, additional techniques for augmenting data were applied, including scaling, flipping, and random rotations. By mimicking common variances found in real-world medical imaging, these strategies improve themodel's generalization.

The collected data is supported in four CNN models, namely ResNet50, VGG16, MobileNetV2, DenseNet121 and U-Net which are used to enhance feature extraction on medical images while enhancing the sample segmentation and classification. For each of the four models, the segmented images from the U-Net network were used as input in order to extract optimal features and to classify. This approach guarantees that different strengths of each model are used in identification of fine patterns and the accuracy of the diagnosis of bone cancer is enhanced. To assess the degree of success of the model to differentiate between bone cancer and results derived from histopathological findings measures including precision, recall, f1 scores and accuracy were used.

Data Collection: The collection of precise as well as valid data is essential for model's training in order to get the highest level of classification accuracy that makes the model more credible. This study uses the publicly accessible data set in which bone cancer images can be found as histopathological photos. After collecting data, preliminary processing is carried out to transform the images to a form which the model

may accept as input.Data is first divided into four folders illustrating four classes: non-tumor, non-viable-tumor, viable and viable_non-viable respectively. Afterwards, data is separated into train and test data with a proportion of 80:20.

Exploratory Data Analysis: After collecting the data it is further analyzed for better performance of the system.There are four classes of bone cancer they are non-tumor, non-viable-tumor, viable tumor and viable_non-viable tumor.From the observation of analysis it was found that the other classes except non-tumor was undersampled.The performance of the model depends on the data used for training.Minimizing the maximum data and maximizing the minimum data may lead to loss of important features.

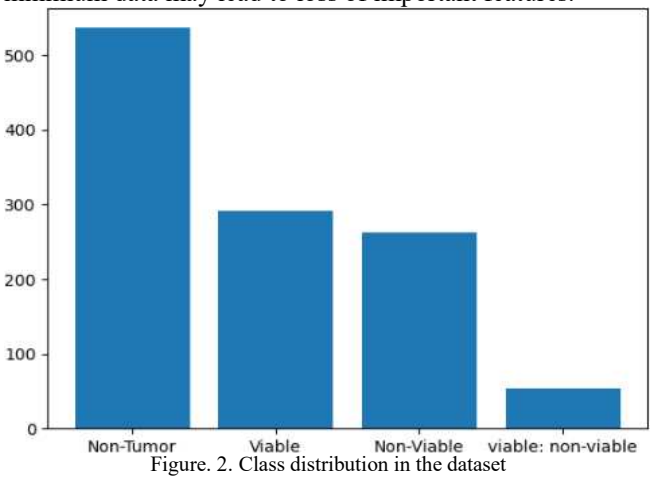

Figure. 2. Class distribution in the dataset

Data Augmentation: A range of data augmentation procedures were utilized to decrease overfitting by increasing the training dataset's effective size and improve our model's generalizability. Since medical imaging data is not readily available for the diagnosis of bone cancer, augmentation is essential in recreating real-world deviations commonly found in medical scans. Among the augmentation techniques were random rotations to mimic different scan orientations, flipping (horizontally and vertically) to account for mirrored or inverted scans, and scaling to introduce variation in the size of regions of interest.

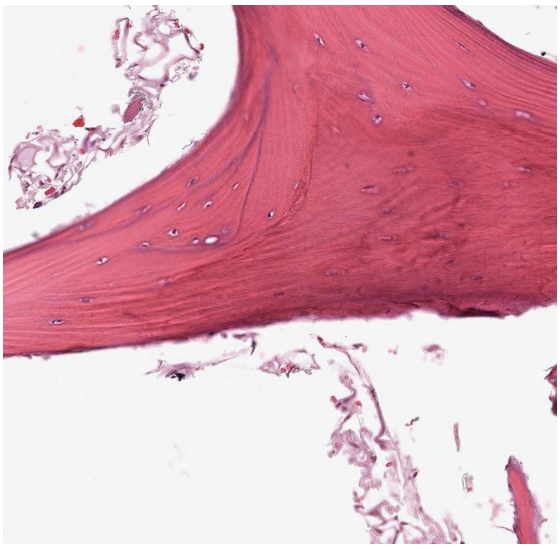

Figure. 3. Before augmentation

Model Architecture:

U-Net: An encoder-decoder structure is used which will not only generate the class labels but generates masks that are of same dimensions as input images, by eliminating need for a fully connected network making it useful for the identification of what is in the image.The encoder framework consists of four encoder blocks,each including an activation function of ReLU after two layers of convolution with a valid padding and 3x3 kernel size.The result is sent into a 2x2 kernel-sized max pooling layer, which splits learnt spatial dimensions into half and lowers the training cost of the model. The bottom layer, known as the bottleneck layer, is composed of two convolutional layers, subsequent to ReLU, and is situated between the encoding and decoding networks. The final aspect of feature map illustration is created by this layer. U-Net's skip connections along with decoding frameworks are its key strengths in segmenting images.

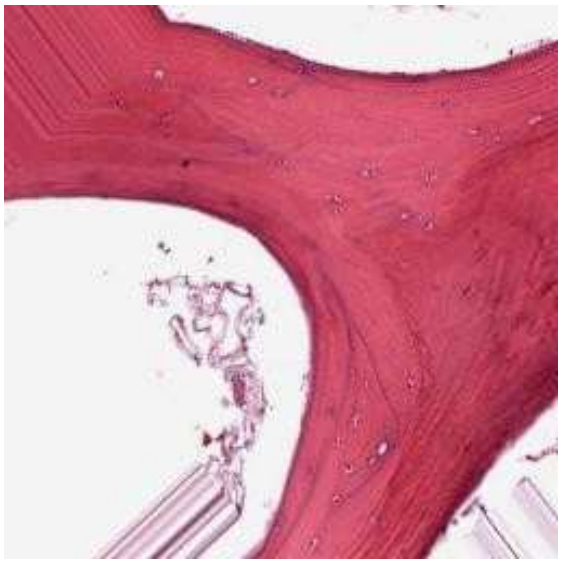

Figure. 4. After augmentation

Decoder network up samples the provided feature maps to imitate input image’s size. This network uses the Feature Map associated with bottleneck layer to produce segmentation masks utilizing skip connections. It assists in determining where the object is located in the image. Each of the four decoder blocks in it has a transpose convolution layer with a 2x2 kernel size. Following this output's concatenation with the encoder block's associated skip layer connection, the network applies a Relu activation function and a pair of convolutional layers with kernel sizes of 3 by 3.

By using skip connections, we can more effectively create our segmentation map using the contextual feature data that was gathered in the encoder blocks. The idea is to project the bottleneck layer's output, or feature map using high-resolution features that have been learnt from the encoder blocks (via skip connections).

The last decoder block is immediately followed by a 1x1 convolution with sigmoid activation, resulting in the output of a segmentation mask with pixel-wise classification. Thus, by employing a U-Net, it is possible to obtain both the feature information and the localization.

DenseNet121 is a densely connected CNN. It has a dense block ,transition layer which reduces the number of channels without losing the feature. In this model every layer gets input

from the preceding layer.Each layer shares the knowledge gained which reduces the duplication of the calculation and makes it easy for the model to learn. The reuse of features makes it pass information efficiently through the dense layers. Residual Network is a deep learning model that was aimed to help train deeper neural networks. It does so by exploiting some pathways called “skip connections” which enable information to leap across some layers.One of the issues, which can occur in deep networks, can be fixed, for instance, one can lose information as it passes through layers.

VGG 16 is a deep learning model, used for image classification and it consists of a systematic design structure with multiple convolution and pooling layers consecutively. It accepts only images of a fixed size but then extracts hierarchical features by utilizing small convolutional filters and using pooling layers that keep reducing the image spatial size. Following the multiple convolutional blocks the feature maps are flattened and fed to fully connected layers, the ones which do the reasoning and classification.

MobileNetV2 is a deep neural network architecture suitable for use in low power platforms Depthwise separable convolutions are applied which decrease load while increasing accuracy. Its main idea is an inverted residual building block that allows carefully preserving the features of the data during the transformation.

In this research, we present the result of a comparative analysis of four state-of-art deep learning architectures namely DenseNet121, MobileNetV2, VGG16, and ResNet50 on a dataset. Each model uses transfer learning where their first layers are set to be non-trainable and then gradually trained on the dataset. The DenseNet121 architecture allowed efficient parameterization, facilitated by dense connections improving feature reuse, and a high accuracy was reached. Another effective model is MobileNetV2 which is narrow and deep both speeding up and delivering results at a needed performance level for mobile use. Although VGG16 is larger and slower, depth makes it a good feature extractor. Last but not the least ResNet50 comes with residual connections which seems to have proved effective in mitigating vanishing gradient problems which makes training to be stable and fairly high accurate especially after fine-tuning. All models have different advantages and disadvantages concerning accuracy, computational time, and inference time where DenseNet121 was the most efficient model with respect to Accuracy, Precision, Recall and F1-score. MobileNetV2 capacitated all parameters very closely while VGG16 and ResNet50 stood lower

Implementation:

TensorFlow and Keras were used for data augmentation after EDA was used to check for disparities in classes in the data cleaning.

IV. Results & Discussions

In this work, we tested four architectures: Among the pre-trained models MobileNetV2, DenseNet121, VGG16, and ResNet50 were included in the framework for bone cancer detection and segmentation. The analysis of the results showed a higher accuracy of DenseNet121 and MobileNetV2, where DenseNet121 had a slightly better accuracy of 79.88% compared to the accuracy achieved by MobileNetV2 of x75.93 %. To assess the performance of models, Precision, Recall, and F1-score were chosen as benchmark parameters based on prior studies; models have been found to outperform the benchmark, and DenseNet121 has been proven to be the most promising model in the current experiment. In the same vein, VGG16 and ResNet50 failed to perform exceptionally as VGG16 had an accuracy of 73.96 while ResNet50 a low rating of 57.50% accuracy.

The best performers in this investigation, DenseNet121 and MobileNetV2, showed their ability to run time for applying in clinical practice. DenseNet121 provided precise and accurate feature learning from the histopathology images with maximum phase. Image segmentation was done through the U-Net model from the TensorFlow and Keras library and classification through models such as ResNet, VGG16, MobileNet V2, DenseNet121 built using TensorFlow. In order to have quick training and cut down the time taken during learning, a GPU was employed which helps process the large dataset.

Recall values along with high accuracy whereas MobileNetV2 has less computational overhead compared to others and hence provides fast completion of procedure. These models used in the study achieved image processing speeds of less than 0.5 seconds and can therefore be useful in diagnostic applications. Although DenseNet121 performed well in feature extraction, MobileNetV2 had an efficient architecture with appropriate performance and speed.

In order to improve bone cancer diagnosis, MobileNetV2's efficiency would provide the ideal balance, while DenseNet121 appeared to perform better than other models on the dense feature extraction front. By highlighting these aspects, these models can improve diagnosis speed and accuracy.

The model's real-time processing capability suggests its potential for clinical applications, enhancing efficiency in diagnostic workflows. These enhanced detection methods discussed in this paper could greatly improve the clinical practice by offering quicker and more reliable diagnosis of bone cancer. With increased accuracy and precision, the model can aid medical professionals in identifying early-stage tumors, which could lead to better patient outcomes and fewer diagnostic errors.

However, the study faced limitations, including the dataset size, as the availability of labeled histopathology images for bone cancer remains limited. Although augmentation techniques were employed to expand the effective training size, this constraint poses a challenge. Additionally, the model required substantial computational resources, which can be a drawback in implementing the model in conditions of limited computational resources.

Table.1. Proposed Model Metrics Comparison

Model / Metrics	Accuracy(%)	Precision (%)	Recall (%)	F1-Score (%)
Mobile NetV2	75.93	76.10	75.93	75.96
Dense Net121	80.01	79.96	79.88	79.85
VGG16	73.96	73.79	73.96	73.44

ResNet50	57.50	60.59	57.50	58.25

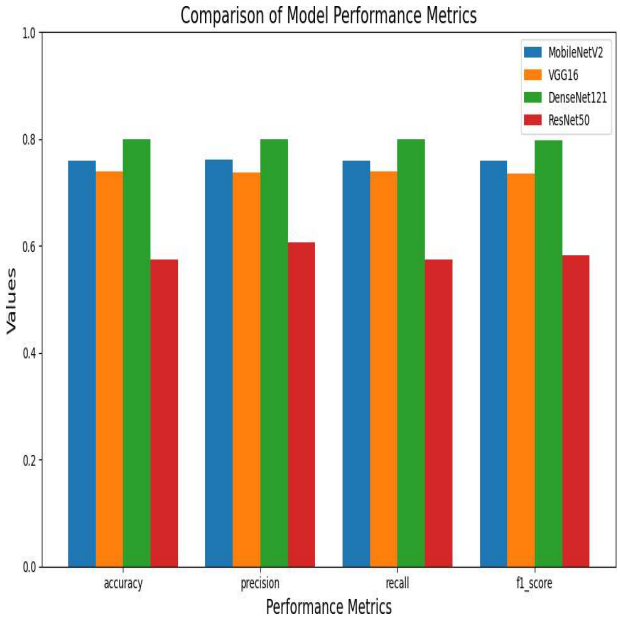

Figure. 5. Performance Metrics

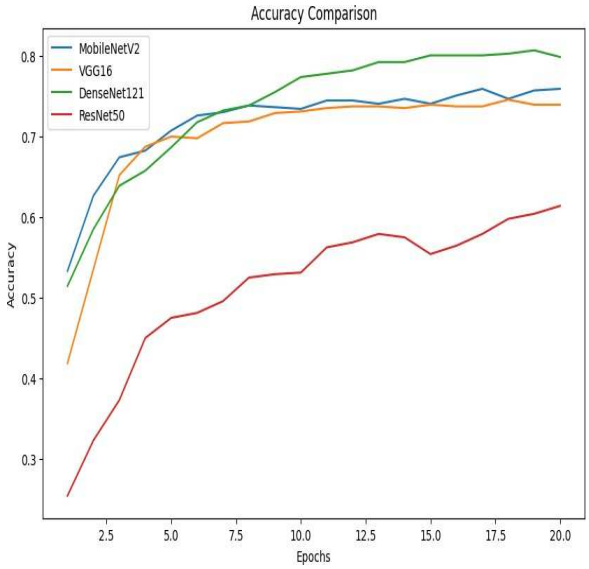

Figure. 6. Accuracy Comparison

V. Conclusion

When performing EDA in this research, we realized that we have a huge class imbalance problem in the provided dataset. To this end, we used some data augmentation strategies as rotation, flipping, and zooming in order to augment the data set and reduce overfitting of the model. First, the histopathology images were selected and segmented using the U-Net to obtain features of interest and classify them using other models including ResNet50, VGG16, MobileNetV2, and DenseNet121. By comparing the results obtained by distinct models, it was possible to measure the accuracy, precision, recall, and F1 score of each one while examining this dataset of medical images in detail. To increase the performance even more, we also looked into transformers and the use of techniques such as fine-tuning, with attention mechanisms to select certain segments within the images. To this end, through the utilization of these sophisticated methods and employing learning through transfer, we were able to compare the improved precision, recall mean F1 score, subsequently strengthening the accuracy of classifications in histopathology images analysis.

References

[1] K. T, B. S. Jeevena, J. A, S. P. S and R. C. V, "Bone Age Assessment (BAA) using Convolutional Transformer and MultiLayer Perceptron," 2024 ICOSEC, Trichy, India, 2024, pp. 1772-1777, doi: 10.1109/ICOSEC61587.2024.10722729.

[2] S. P. S, S. J, K. T and K. P, "Enhanced Classification of Oral Cancer Using Deep Learning Techniques," 2024 ICAIT, Chikkamagaluru, Karnataka, India, 2024, pp. 1-5, doi: 10.1109/ICAIT61638.2024.10690583.

[3] Kumar, P., Senthilselvi, A., Manju, I. et al. HUMRC-PS: Revolutionizing plant phenotyping through Regional Convolutional Neural Networks and Pelican Search Optimization. Evolving Systems 15, 2211–2230 (2024). https://doi.org/10.1007/s12530-024-09612-6

[4] Nagarani, N., et al. "Self-attention based progressive generative adversarial network optimized with momentum search optimization algorithm for classification of brain tumor on MRI image." Biomedical Signal Processing and Control 88 (2024): 105597.

[5] Rajagopal RK, Karthick R, Meenalochini P, Kalaichelvi T. Deep Convolutional Spiking Neural Network optimized with Arithmetic optimization algorithm for lung disease detection using chest X-ray images. Biomedical Signal Processing and Control. 2023 Jan 1;79:104197.

[6] Jasper Gnana Chandran, J., et al. "Dual-channel capsule generative adversarial network optimized with golden eagle optimization for pediatric bone age assessment from hand X-ray image." International Journal of Pattern Recognition and Artificial Intelligence 37.02 (2023): 2354001.

[7] M. A. A. Walid and P. C. Shill, "A Transfer-Learning Based Unbiased Voting Bone Cancer Detection Framework from Histological Osteosarcoma Images," 2023 ICCCNT, Delhi, India, 2023, pp. 1-7.

[8] J. Wu, T. Yuan, J. Zeng and F. Gou, "A Medically Assisted Model for Precise Segmentation of Osteosarcoma Nuclei on Pathological Images," in IEEE Journal of Biomedical and Health Informatics, vol. 27, no. 8, pp. 3982-3993, Aug. 2023.

[9] C. Loraksa, S. Mongkolsomlit, N. Nimsuk, M. Uscharapong and P. Kiatisevi, "Development of the Osteosarcoma Lung Nodules Detection Model Based on SSD-VGG16 and Competency Comparing With Traditional Method," in IEEE Access, vol. 10, pp. 65496-65506, 2022.

[10] A. Saber, M. Sakr, O. M. Abo-Seida, A. Keshk and H. Chen, "A Novel Deep-Learning Model for Automatic Detection and Classification of Breast Cancer Using the Transfer-Learning Technique," in IEEE Access, vol. 9, pp. 71194-71209, 2021.

[11] C. Sun, A. Xu, D. Liu, Z. Xiong, F. Zhao and W. Ding, "Deep Learning-Based Classification of Liver Cancer Histopathology Images Using Only Global Labels," in IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 6, pp. 1643-1651, June 2020.

[12] Y. Weng, T. Zhou, Y. Li and X. Qiu, "NAS-Unet: Neural Architecture Search for Medical Image Segmentation," in IEEE Access, vol. 7, pp. 44247-44257, 2019.

[13] Z. Zeng, W. Xie, Y. Zhang and Y. Lu, "RIC-Unet: An Improved Neural Network Based on Unet for Nuclei Segmentation in Histology Images," in IEEE Access, vol. 7, pp. 21420-21428, 2019.

[14] X. Fu, L. Bi, A. Kumar, M. Fulham and J. Kim, "Multimodal Spatial Attention Module for Targeting Multimodal PET-CT Lung Tumor Segmentation," in IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 9, pp. 3507-3516, Sept. 2021.

[15] V. A. Ashwath, O. K. Sikha and R. Benitez, "TS-CNN: A Three-Tier Self-Interpretable CNN for Multi-Region Medical Image Classification," in IEEE Access, vol. 11, pp. 78402-78418, 2023.