**To:** The Journal of Supercomputing Editor

**Re:** Response to reviewers

**Ref:** Submission ID3f13e75d-d8fb-48cc-98f0-b6267908b961

Dear Editor,

Thank you for allowing a resubmission of our following manuscript, with an opportunity to address the reviewers’ comments.

**Original Article Title:** “Artificial Intelligence for Detection of Lung Cancer using Transfer Learning and Morphological Features”

**Authors:**

1. Nafe Muhtasim Hye

Department of Electrical and Electronic Engineering (EEE)

Ahsanullah University of Science and Technology (AUST)

Dhaka, Bangladesh

1. Umma Hany, PhD

Associate Professor

Department of Electrical and Electronic Engineering (EEE)

Ahsanullah University of Science and Technology (AUST)

Dhaka, Bangladesh

1. Tahmina Islam

Department of Electrical and Electronic Engineering (EEE)

Ahsanullah University of Science and Technology (AUST)

Dhaka, Bangladesh

1. Nusrat Nawreen

Department of Electrical and Electronic Engineering (EEE)

Ahsanullah University of Science and Technology (AUST)

Dhaka, Bangladesh

1. Abdullah Al Mamun

Department of Electrical and Electronic Engineering (EEE)

Ahsanullah University of Science and Technology (AUST)

Dhaka, Bangladesh

We are uploading

(a) our point-by-point response to the comments (below)

(b) an updated manuscript highlighted “changes” in “yellow” and “no changes but clarification” of the reviewer’s concern in “green”

(c) a clean updated manuscript (latex files) without highlights (“Main Manuscript”*).*

Best regards,

Umma Hany, PhD

Associate Professor

Department of Electrical and Electronic Engineering (EEE)

Ahsanullah University of Science and Technology (AUST), Dhaka, Bangladesh

**Response to the Reviewer # 1**

**Note: In the revised manuscript, the unchanged contents are highlighted in green for clarification of the reviewer's queries and the changed contents are highlighted in yellow.**

**Reviewer#1, Comments:**

Reviewer 1

* Primary Comments:

The definitions of false positives and false negatives are incorrect, which leads to the wrong results in the performance evaluation.

Authors should rectify the definitions and modify the results accordingly.

There is no cross-validation study performed, serious case of overfitting.

* Secondary:

English needs revision, first sentence of the Abstract “Lung Cancer is an uncontrolled growth … we call lung tumor.” needs to be rewritten.

Authors should confirm the figure reference as “Figure” or “Fig.” and maintain consistency throughout the manuscript.

**Author response:** Thank you very much for your comments.



**Reviewer#1, Concern # 1:**

Primary Comments:

The definitions of false positives and false negatives are incorrect, which leads to the wrong results in

the performance evaluation.

Authors should rectify the definitions and modify the results accordingly.

**Author response and action:** In the present revised manuscript, we have updated the definitions of False Positive and False Negative in section 6.1 as per reference [1] on page 12 as follows (**highlighted in yellow**):

**“6.1 Confusion matrix**

Confusion Matrix is a visualization of ground-truth labels versus model predictions. Each row of the confusion matrix represents the instances in a predicted class and each column represents the instances in an actual class. Each cell in the confusion matrix represents any of the following evaluation factors:

1. True Positive (TP) signifies how many actual positive class samples are predicted correctly by the model.

2. True Negative (TN) signifies how many actual negative class samples are predicted correctly by the model.

3. False Positive (FP) signifies how many actual negative class samples are predicted incorrectly by the model.

4. False Negative (FN) signifies how many actual positive class samples are predicted incorrectly by the model.”

Reference [2] has been added as **Reference [29]** in the “Reference” section (**highlighted in green**) and cited inline on page 12 as below (**highlighted in green**):

**“6 Performance Evaluation**

We apply the following performance metrics to evaluate the performance of the classification models [29].”

Reference

[1] B. Aayush, “Performance Metrics in Machine Learning [Complete Guide],”

<https://neptune.ai/blog/performance-metrics-in-machine-learning-complete-guide>, Online accessed 29th Sep 2023.



**Reviewer#1, Concern # 2:** There is no cross-validation study performed, a serious case of overfitting.

**Author response and action:** Thank you for your thoughtful observation.

In the revised manuscript, 5-fold cross-validation has been performed using LR, KNN, and SVM classifiers on the preprocessed data, VGG16 Transfer learning (TL) features, and morphological features to split the datasets and validate the test accuracy of all the datasets. As we can see from the results in Tables 6 and 7 that LR performs the best with the preprocessed data and TL features, and the best possible accuracy is obtained in the least computational time using the KNN and SVM classifiers on the morphological features. Therefore, we applied 5-fold cross-validation on the morphological features using the LR, KNN, and SVM classifiers where the datasets are shuffled to 42 random states and split into 5 groups each containing 20% of the total datasets. Then one of the groups is taken as the test dataset and the remaining groups are taken as the training dataset. Then the proposed model is fitted on the training datasets and evaluated on the test datasets. Finally, the evaluation score is recorded for each test group and averaged to evaluate the performance of the model.

**The results of the 5-fold cross-validation are included in Table 8 in the revised manuscript**. In the revised manuscript, the k-fold cross-validation method has been discussed in **Section 5.3 k-fold Cross-validation** of the revised manuscriptas highlighted follows:

**“5.3**  **k-fold Cross-validation**

**The k-fold cross-validation is applied to split the datasets and to validate the test accuracy of all the datasets [21]. We apply 5-fold cross-validation where the datasets are shuffled to 42 random states and split into 5 groups each containing 20% of the total datasets. Then one of the groups is taken as the test dataset and the remaining groups are taken as the training dataset. Then the proposed model is fitted on the training datasets and evaluated on the test datasets. Finally, the evaluation score is recorded for each test group and averaged to evaluate the performance of the model.”**

Table 8 has been cited and the results have been also updated in the “Experimentation and Results” section of the revised manuscript on page 19 as follows:

**“Tables 6 and 7 present the comparative results of the seven ML algorithms using the preprocessed raw data, transfer learning (TL) features, and morphological features. It is observed that LR performs the best with preprocessed data and TL features. Whereas, using the low dimensional morphological features, all the ML algorithms perform with reasonably high accuracy among which KNN and SVM perform the best with the least computational time. Therefore, 5-fold cross-validation is applied to mitigate overfitting problems and to verify the best possible accuracy using the LR, KNN, and SVM algorithms. The cross-validation results are presented in Table 8. It is observed in Table 8 that LR performs with 99.36% and 99.27% accuracy with preprocessed data and TL features in 23.71 sec and 16.51 sec computational time, respectively. Whereas, both KNN and SVM perform with 99.76% accuracy in 0.017 sec and 0.008 sec time, respectively. Thus, it may be concluded that using the low dimensional morphological features KNN and SVM classification algorithms perform the best in the least computational time. Table 9 compares the accuracy of the proposed works to other related works in the literature. It is apparent from the comparison that the proposed works outperform the literature with high computational intelligence and accuracy.”**

The Results have been also updated in the “Abstract”, last para of “Introduction” and in “Conclusion” as follows:

**Abstract:**

**“It is observed from the 5-fold cross-validation results that logistic regression (LR) performs with 99.36% accuracy in 23.71 sec time using the preprocessed data. Whereas, using the morphological features, k-Nearest Neighbor (KNN) and the Support Vector Machine (SVM) perform with the highest accuracy of 99.76% with very reduced computational time of 0.017 sec and 0.008 sec, respectively.”**

**Last para of Introduction:**

**“We apply 5-fold cross-validation to mitigate the overfitting problems and to verify the best possible accuracy using the LR, KNN, and SVM algorithms. It is observed from the results that LR performs with 99.36% and 99.27% accuracy with preprocessed data and TL features in 23.71 sec and 16.51 sec computational time, respectively. Whereas, both KNN and SVM perform with 99.76% accuracy in 0.017 sec and 0.008 sec time, respectively. Thus, it may be concluded that using the low dimensional morphological features, KNN and SVM classification algorithms perform the best in the least computational time. The accuracy is also compared to other methods in the literature and it is found that the proposed methods outperform the literature in terms of accuracy and computational time.”**

**Conclusion:**

**“Then, 5-fold cross-validation method is applied to mitigate the overfitting problem. It is observed that LR performed with 99.36% and 99.27% accuracy using preprocessed data and TL features in 23.71 sec and 16.51 sec computational time, respectively. Whereas, both KNN and SVM performed with 99.76% accuracy in 0.017 sec and 0.008 sec time, respectively. Thus, it may be concluded that a significantly high accuracy of lung tumor detection and classification is possible to be achieved with the machine learning algorithms applied on the extracted low dimensional morphological features with low computational complexity and time. Therefore, the proposed methods recommend extracting low-dimensional morphological features and choosing machine learning-based classifiers for real-time lung cancer recognition.”**



**Reviewer#1, Concern # 3:** Secondary:

English needs revision, first sentence of the Abstract “Lung Cancer is an uncontrolled growth … we call lung tumor.” needs to be rewritten.

**Author response and action:** Thank you for the observation. In the revised manuscript, we have updated the first sentence of the abstract as follows:

**“Lung cancer is an uncontrolled growth of tissue causing a lump in the human lung.”**

**Reviewer#1, Concern # 4:** Authors should confirm the figure reference as “Figure” or “Fig.” and maintain consistency throughout the manuscript.

**Author response and action:** Thank you for the observation. We revised the manuscript by replacing “Fig.” as “Figure” throughout the manuscript to maintain consistency.

**Response to the Reviewer # 2**

**Note: In the revised manuscript, the unchanged contents are highlighted in green for clarification of the reviewer's queries and the changed contents are highlighted in yellow.**

**Reviewer#2, Concern # 1:** In the context of lung cancer detection from medical images, could you expound upon the trade-offs between the reported accuracies of various machine learning algorithms such as Artificial Neural Network (ANN), Decision Tree, Support Vector Machine (SVM) with Radial Basis Function kernel, and DenseNet nonnegative sparse and collaborative representation (DenseNet-NSCR)? How do the complexities of the respective feature extraction methods impact the achieved accuracies, and how does the proposed morphological feature extraction method in the current study compare in terms of accuracy, computational complexity, and processing time against these established techniques?

**Author response and action:** Thank you for your observations.

In the revised manuscript, the trade-offs between the reported accuracies of various algorithms reported in the literature have been discussed **in para 2 and 3 of the Introduction on page 2 as highlighted in the revised manuscript as follows:**

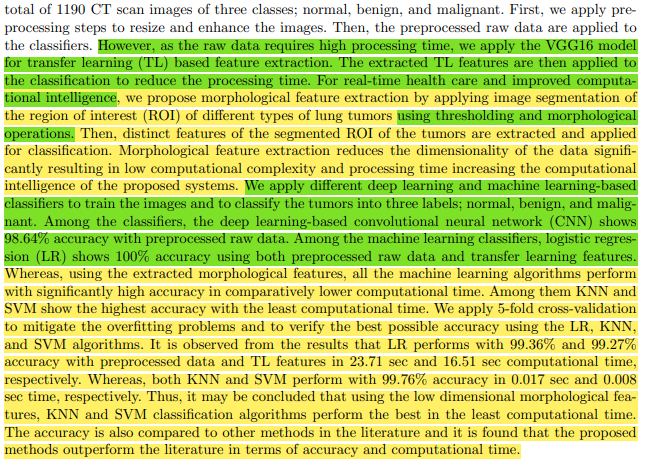
**Para 2 of Introduction:**

****

**Para 3 of Introduction:**

**“Though there are several works in the literature, most of the works are on binary classification of lung tumors or predicting a specific type of cancer without applying cross-validation. Additionally, the computational time has not been reported in most of the above works. Hence, a multi-classification approach with high computational effectiveness is required for the fast detection and classification of lung tumors. In this paper, we propose a multi-classification approach for lung tumor detection and classification as normal, benign, and malignant with high computational effectiveness by applying machine learning and deep learning methods on the CT scan images. We use the IQ-OTHNCCD lung cancer dataset [16] collected from the Kaggle [18]. The dataset contains a”**

The impact of the proposed feature extraction methods on the achieved accuracies has been explained in **Para 2 of the “Introduction” on page 3 as highlighted in the revised manuscript as follows (changes highlighted in yellow):**



The accuracy and computational effectiveness of the proposed morphological feature extraction method to the reported works in the literature have been presented and compared **in Table 9 of the revised manuscript. Three new columns named “Binary/Multi-class” and “Cross-validation” and “Training time” have been also added in Table 9 to compare the computational complexity.**

**Table 9 on page 20 has been updated in the revised manuscript (changes highlighted in yellow) and a recent work Pandian et al [17] of 2022 has been compared in Table 9 on page 20 and cited in the literature review in the “Introduction” on the page 2 (as highlighted in yellow) as follows:**

**“In [17], the authors use CNN and Google Net deep learning algorithms for lung cancer detection and binary classification and achieve a precision of 98% in detection and classification.”**

The impact of the proposed morphological feature extraction method on the accuracy and computational time **has been presented and compared in Tables 5, 6, 7 and 8 (as highlighted in the revised manuscript**). **In the revised manuscript, 5-fold cross-validation has been applied using LR, KNN, and SVM classifiers to mitigate the overfitting problem and to find the best possible accuracy.** **The results of 5-fold cross-validation are presented in Table 8.** As we can see in Tables 5, 6 and 7, using the proposed morphological feature extraction method, KNN and SVM can obtain significantly high accuracy using very low dimensional features of 4x1 dimension in a significantly reduced computational time which indicates the high computational effectiveness of the proposed methods. It is also observed in Table 8 that using the morphological features, both KNN and SVM classification algorithms perform the best in the least computational time. **Table 8 has been cited and the impact and results using the feature extraction have been compared** in terms of accuracy, computational complexity, and processing time against the established techniques **on page 19 in the revised manuscript as follows (highlighted in yellow):**

**“Tables 6 and 7 present the comparative results of the seven ML algorithms using the preprocessed raw data, transfer learning (TL) features, and morphological features. It is observed that LR performs the best with preprocessed data and TL features. Whereas, using the low dimensional morphological features, all the ML algorithms perform with reasonably high accuracy among which KNN and SVM perform the best with the least computational time. Therefore, 5-fold cross-validation is applied to mitigate overfitting problems and to verify the best possible accuracy using the LR, KNN, and SVM algorithms. The cross-validation results are presented in Table 8. It is observed in Table 8 that LR performs with 99.36% and 99.27% accuracy with preprocessed data and TL features in 23.71 sec and 16.51 sec computational time, respectively. Whereas, both KNN and SVM perform with 99.76% accuracy in 0.017 sec and 0.008 sec time, respectively. Thus, it may be concluded that using the low dimensional morphological features KNN and SVM classification algorithms perform the best in the least computational time. Table 9 compares the accuracy of the proposed works to other related works in the literature. It is apparent from the comparison that the proposed works outperform the literature with high computational intelligence and accuracy.”**



**Reviewer#2, Concern # 2:** Could you elucidate the comparative assessment of the proposed hybrid attention mechanism and DenseNet-based parallel deep learning model for lung cancer image segmentation, as outlined in paper [8], in terms of their achieved accuracy and computational efficiency? Furthermore, in the light of the current study's findings, how do these results contribute to the field's understanding of effective image segmentation techniques for lung cancer, particularly in relation to the examined literature methods and the concurrently proposed morphological feature extraction-based classification approach?

**Author response:** Thank you for your comments. Paper [8] **(Paper [9] in the revised manuscript)** presents a lung cancer image segmentation method using a hybrid attention mechanism and a DenseNet-based parallel deep learning model. In our paper, it has been reported **as one of** **the recent works on automatic segmentation algorithms for lung tumors**. However, the method proposed in [9] has not been used in our proposed method of morphological feature extraction.



**Reviewer#2, Concern # 3:** Within the framework of lung tumor detection and classification utilizing CT scan images, could you elaborate on the intricacies of the proposed tri-fold classification strategy as illustrated in Figure 1? Specifically, could you elucidate the interplay between the initial preprocessing steps involving scaling, normalization, data cropping, and enhancement, and their subsequent impact on the diverse feature extraction mechanisms utilized in conjunction with machine learning and deep learning classifiers? In light of the three distinct classification approaches – raw data-based classification, transfer learning-based classification, and morphological feature-based classification – how does the proposed morphological feature extraction strategy, aimed at both dimensionality reduction and computational efficiency, diverge from and converge with the alternative classification methodologies detailed in the study, and how do these methodologies collectively contribute to a comprehensive understanding of lung tumor classification accuracy and computational complexity?

**Author response:** Thank you for your observations.

Two different preprocessing steps have been applied before applying to the two different types of feature extraction methods. Before VGG16 transfer learning (TL) feature extraction, scaling and normalization method is applied on the RGB images where scaling is applied to transform data so that it fits within a specific scale and the normalization is applied to change the observations so that they can be described as a normal distribution. We randomly shuffled the train pictures into a state of 25. Then we scale each pixel using a factor of 255. The majority of picture data has integer pixel values between 0 and 255. Small weight values are processed by neural networks, while high integer values might interfere with or slow down learning. Since, each pixel value of the image should range from 0 to 1, normalizing the pixel values is a good option. We use min-max scaling as follows to normalize the data such that the feature value remains within a specific range of 0 and 1.

Before morphological feature extraction, image cropping and enhancement is applied on the grayscale images as the preprocessing steps as shown in Figure 3. First, the images are resized and cropped to get the required dimensions. Finally, the contrast level of the images is adjusted to enhance the image features such as boundaries and edges. The original image, the cropped image, and the enhanced image are shown in Figure 3. After applying the image cropping and enhancement process, the raw RGB image of 512×512×3 uint8 dimension is converted to a binary image of 136×151 logical dimension.

We have developed three different approaches of lung cancer classification to compare the accuracy. In the first approach, the classifier receives the preprocessed raw data for classification using machine learning and deep learning methods. In the second approach, the features of the preprocessed data are extracted using VGG16-based transfer learning (TL) and classified using machine learning or deep learning methods. In the third approach, low-dimensional morphological features are extracted to reduce the processing time, and then the extracted features are applied to the ML classifiers.

In the above three approaches, we applied two feature extraction methods called VGG16 transfer learning (TL) and morphological feature extraction.

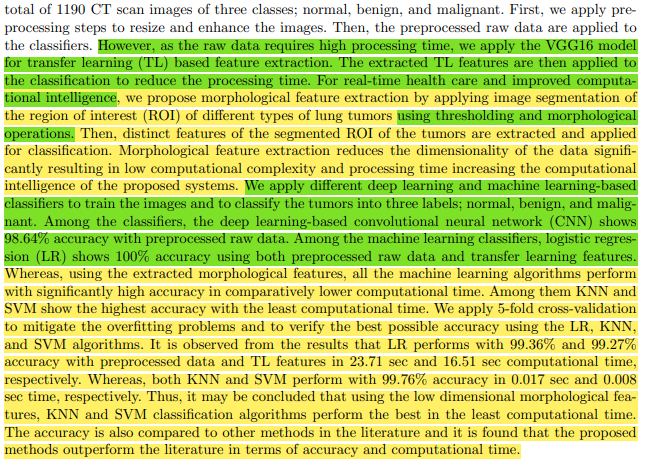
TL models work better with small datasets. The VGG16 TL feature extraction has been chosen as it is deeper, less complex, and faster than other transfer learning models. After applying the TL feature extraction, a new dataset is created from the input image dataset of lung tumors using the pre-trained model, and a three-dimensional feature stack containing the recognized visual features is provided. The extracted features are then fed into the machine learning (ML) and deep learning (DL) classifier for lung tumor classification.

For morphological feature extraction, 7 types of morphological operations are applied on the thresholded binary images of lung tumors for ROI segmentation of the malignant, benign, and normal type tumors as shown in **Figures 5, 6, and 7**. The ROI segmentation processes for malignant, benign, and normal-type tumors are sequentially applied as shown in the flow chart in Figure 8. Finally, the four morphological features (area, eccentricity, perimeter, and compactness) of the segmented ROIs are extracted before applying them to the classifiers.

By applying morphological feature extraction, the dimension of the extracted features is significantly reduced as compared to the dimension of the original raw data. The dimension of the extracted features is 4×1 double containing the four extracted features of each segmented region. The extracted morphological features are classified using machine learning and deep learning classifiers. As morphological feature extraction compresses the size of the training and test data significantly, it results in less computational complexity and reduced training and test time of classification.

**The preprocessing, ROI segmentation, feature extraction, and classification methods are explained in detail (pages 4 to 13 of the revised manuscript) in “Section 3 Preprocessing”, “Section 4 Feature Extraction” and “Section 5 Classification”**.The headings of **Sections 3, 4, and 5 and** the captions of **Figures 3, 5, 6, and 7 are highlighted in green.**

The impact of the proposed feature extraction methods on the achieved accuracies has been explained in **Para 2 of the “Introduction” on page 3 (as highlighted in yellow and green) in the revised manuscript as follows:**



The impact of the proposed morphological feature extraction method on the accuracy and computational time **has been presented and compared in Tables 5, 6, 7 and 8 (as highlighted in the revised manuscript**). **In the revised manuscript, 5-fold cross-validation has been applied using LR, KNN, and SVM classifiers to mitigate the overfitting problem and to find the best possible accuracy.** **The results of 5-fold cross-validation are presented in Table 8.** As we can see in Tables 5, 6 and 7, using the proposed morphological feature extraction method, KNN and SVM can obtain significantly high accuracy using very low dimensional features of 4x1 dimension in a significantly reduced computational time which indicates the high computational effectiveness of the proposed methods. It is also observed in Table 8 that using the morphological features, both KNN and SVM classification algorithms perform the best in the least computational time. **Table 8 has been cited and the impact and results using the feature extraction have been compared** in terms of accuracy, computational complexity, and processing time against the established techniques **on page 19 in the revised manuscript as follows (highlighted in yellow):**

**“Tables 6 and 7 present the comparative results of the seven ML algorithms using the preprocessed raw data, transfer learning (TL) features, and morphological features. It is observed that LR performs the best with preprocessed data and TL features. Whereas, using the low dimensional morphological features, all the ML algorithms perform with reasonably high accuracy among which KNN and SVM perform the best with the least computational time. Therefore, 5-fold cross-validation is applied to mitigate overfitting problems and to verify the best possible accuracy using the LR, KNN, and SVM algorithms. The cross-validation results are presented in Table 8. It is observed in Table 8 that LR performs with 99.36% and 99.27% accuracy with preprocessed data and TL features in 23.71 sec and 16.51 sec computational time, respectively. Whereas, both KNN and SVM perform with 99.76% accuracy in 0.017 sec and 0.008 sec time, respectively. Thus, it may be concluded that using the low dimensional morphological features KNN and SVM classification algorithms perform the best in the least computational time. Table 9 compares the accuracy of the proposed works to other related works in the literature. It is apparent from the comparison that the proposed works outperform the literature with high computational intelligence and accuracy.”**



**Reviewer#2, Concern # 4:** The process of data preprocessing is fundamental to the efficacy of medical image analysis, as demonstrated in the employed strategies of scaling, normalization, image cropping, and enhancement. Could you delve into the intricacies of the multifaceted image preprocessing methods, such as the conversion between color spaces, the rationale for scaling and normalization, and the role of grayscale conversion in image resizing, cropping, and enhancement? Additionally, in the context of enhancing image features like boundaries and edges, how do these preprocessing procedures synergize with the subsequent morphological feature extraction, and what are the potential implications of these preprocessing strategies on the downstream performance of machine learning and deep learning classifiers, thereby contributing to the broader understanding of efficient and accurate lung tumor classification methodologies?

**Author response:** Thank you for your comments. Data preprocessing steps are applied to make the data suitable for feature extraction or classification. We apply two types of preprocessing called 1) Scaling and normalization, and 2) Image cropping and enhancement.

Before the scaling and normalization, the CT scan data in jpg format in BGR color space are loaded and converted to RGB format. Next, scaling is applied to transform data so that it fits within a specific scale and the normalization is applied to change the observations so that they can be described as a normal distribution. We randomly shuffled the train pictures into a state of 25. Then we scale each pixel using a factor of 255. The majority of picture data has integer pixel values between 0 and 255. Small weight values are processed by neural networks, while high integer values might interfere with or slow down learning. Since, each pixel value of the image should range from 0 to 1, normalizing the pixel values is a good option. We use min-max scaling as follows to normalize the data such that the feature value remains within a specific range of 0 and 1.

The image cropping and enhancement preprocessing steps include data loading, image resizing, cropping, and enhancement processes. First, all the collected CT scan images are loaded and converted to grayscale images. Then, the images are resized and cropped to get the required dimensions. Finally, the contrast level of the images is adjusted to enhance the image features such as boundaries and edges. The original image, the cropped image, and the enhanced image are shown in Figure 3. After applying the image cropping and enhancement process, the raw RGB image of 512×512×3 uint8 dimension is converted to a binary image of 136×151 logical dimension.

**The two types of preprocessing have been explained in “Section 3 Preprocessing” of the manuscript.**



**Reviewer#2, Concern # 5:** In the context of lung tumor detection and classification, could you elaborate on the intricacies of transfer learning as applied to the VGG16 model for feature extraction? How does the architectural depth and complexity of the VGG16 model contribute to its effectiveness in feature extraction, particularly in comparison to other transfer learning models such as VGG19, InceptionNet, XCeption, and ResNet-50? Furthermore, what considerations and methodologies are employed to mitigate overfitting in the transfer learning process, and how does the absence of data augmentation impact the efficiency and accuracy of feature extraction using the VGG16 model?

**Author response:** Thank you for your observations. We have used the VGG16 transfer learning model as shown in Figure 4 for feature extraction. We develop the TL feature extraction method using the VGG16 model of 16 convolutional layers including the Maxpooling layers, 3 dense layers (2 fully connected layers and 1 SoftMax classifier), and an output layer of 1,000 nodes.

A deep Convolutional neural network (CNN) is difficult and expensive to train with small datasets and complex models. When a pre-trained model is repurposed for a different related task is known as transfer learning. Since the transfer learning model can perform with improved accuracy with small datasets, we apply VGG16 [18] pre-trained model to implement transfer learning (TL) for feature extraction. VGG16 is deeper, less complex, and faster than other transfer learning models as VGG19, Inception V3, XCeption, and ResNet-50. VGG16 has an exceptional feature extraction capability as it has a greater capacity to learn new features because it is deeper than certain transfer learning models, such as AlexNet. VGG16 uses just 3×3 convolution layers and 2×2 pooling layers repeatedly, which makes it significantly less complex than other transfer learning models like InceptionNet and enables it to generalize and adapt more effectively to a larger variety of data sets. VGG19 is a similar type of CNN model with 19 layers. However, due to the increased number of CNN layers of VGG19, the ability to fit complex functions also increases which, in turn, increases the training time of VGG19 significantly.

VGG16 may be applied with data augmentation to prevent overfitting and improve accuracy. However, we apply VGG16 without data augmentation as it is faster than the method with data augmentation.

To mitigate overfitting in the transfer learning process, we applied 5-fold cross-validation on the VGG16 features of all the datasets using Logistic regression (LR) machine learning classifier. In the 5-fold cross-validation, the datasets are shuffled to 42 random states and split into 5 groups each containing 20% of the total datasets.

The reason of choosing VGG16 Transfer learning model has been explained **in “Section 4.1 Transfer learning for Feature extraction” of the manuscript.**

**In the revised manuscript, the results of the 5-fold cross-validation method are presented in Table 8.** In the revised manuscript, the k-fold cross-validation method has been discussed **on page 12 in “Section 5.3 k-fold Cross-validation” of the revised manuscript (highlighted in yellow) as follows:**

**“5.3**  **k-fold Cross-validation**

**The k-fold cross-validation is applied to split the datasets and to validate the test accuracy of all the datasets [21]. We apply 5-fold cross-validation where the datasets are shuffled to 42 random states and split into 5 groups each containing 20% of the total datasets. Then one of the groups is taken as the test dataset and the remaining groups are taken as the training dataset. Then the proposed model is fitted on the training datasets and evaluated on the test datasets. Finally, the evaluation score is recorded for each test group and averaged to evaluate the performance of the model.”**

**Table 8 has been cited and the results have been also updated in the revised manuscript on page 19 (highlighted in yellow) as follows:**

**“Tables 6 and 7 present the comparative results of the seven ML algorithms using the preprocessed raw data, transfer learning (TL) features, and morphological features. It is observed that LR performs the best with preprocessed data and TL features. Whereas, using the low dimensional morphological features, all the ML algorithms perform with reasonably high accuracy among which KNN and SVM perform the best with the least computational time. Therefore, 5-fold cross-validation is applied to mitigate overfitting problems and to verify the best possible accuracy using the LR, KNN, and SVM algorithms. The cross-validation results are presented in Table 8. It is observed in Table 8 that LR performs with 99.36% and 99.27% accuracy with preprocessed data and TL features in 23.71 sec and 16.51 sec computational time, respectively. Whereas, both KNN and SVM perform with 99.76% accuracy in 0.017 sec and 0.008 sec time, respectively. Thus, it may be concluded that using the low dimensional morphological features KNN and SVM classification algorithms perform the best in the least computational time. Table 9 compares the accuracy of the proposed works to other related works in the literature. It is apparent from the comparison that the proposed works outperform the literature with high computational intelligence and accuracy.”**



**Reviewer#2, Concern # 6:** The morphological feature extraction process, pivotal to the classification of lung tumors, involves a sequence of steps including thresholding and morphological operations. Could you elucidate the specific methodologies employed in the thresholding process, with a focus on the significance of Otsu's method for global threshold computation? Additionally, delve into the intricacies of the morphological operations, encompassing processes like creating structuring elements, opening operations, background subtraction, image fill operations, logical XOR operations, and their cumulative roles in segmenting and extracting regions of interest (ROIs) for malignant, benign, and normal tumors. How do these operations contribute to the subsequent extraction of morphological features, and what is the underlying rationale for each step in achieving a streamlined and accurate tumor classification?

**Author response:** Thank you for your queries.

Thresholding is a way to convert the gray-scale image to a binary image based on the computed threshold value where the threshold level indicates the intensity value of the image. It is the simplest method of image segmentation. We apply Otsu’s method [19] to compute the global threshold of the contrasted grayscale image. In the binary conversion, the output binary image replaces all pixels in the input grayscale with luminance greater than level with the value 1 (white) and replaces all other pixels with the value 0 (black) to a binary image based on the computed threshold value. The threshold level indicates the intensity value of the image.

The seven (07) morphological operations as explained in the manuscript as below (on pages 7 and 8 **in the revised manuscript**) are applied for ROI segmentation for malignant, benign, and normal tumors:

1. Creating structuring element: First, we extract the background of the binary images by creating a morphological structuring element (SE) of disk shape.

2. Opening operation: Then we apply the morphological opening operation to open the SE from the binary image and to get the background of the image.

3. Background subtraction: The extracted background is then subtracted from the binary image to get the background subtracted image.

4. Complement image: The background-subtracted images are complemented. The complemented background subtracted binary image is considered as the ROI of the malignant type tumors.

5. Image fill operation: The image fill operation is applied to fill the holes in the complemented background subtracted binary images.

6. Logical XOR operation and complement: The logical exclusive OR operation is applied on the complemented images (step 4) and the holes-filled images (step 5) to obtain the ROI of the benign type tumors.

7. Complement the XOR-operated image: The XOR-operated image is complemented to get the ROI of the normal type tumors.

As we can see in **Figures 5, 6 and 7** that the ROI segmentation of Malignant tumors includes steps 1 to 4 of the morphological operations, ROI segmentation of Benign type tumors includes steps 1 to 6 of the morphological operations and ROI segmentation of Normal type tumors includes all the proposed steps 1 to 7 of the morphological operations.

**Figure 8** presents the process flow of ROI segmentation for Malignant, benign and normal type tumors. In the training phase, all the training images of the three types of tumors are gone through the proposed morphological operations for ROI segmentation. However, in the test phase, morphological operations for malignancy are applied first and if malignancy is detected by the classifier, then no further morphological operations are applied for ROI segmentation of Benign or Normal type tumors. Similarly, if benign tumors are detected by the classifier, then no further operations are applied for ROI segmentation of the Normal type tumors.

The morphological feature extraction and the process flow as shown in Figure 8 have been explained in detail in Section **“4.2 Morphological Feature Extraction” (from pages 7 to 10 of the revised manuscript)**.



**Reviewer#2, Concern # 7:** Could you provide a detailed exposition of the classification strategies applied in the study, encompassing both deep learning and machine learning-based approaches? Specifically, elaborate on the architecture and configuration of the Convolutional Neural Network (CNN) employed for deep learning-based classification, highlighting the utilization of convolutional layers, ReLU activation functions, and dense layers. Subsequently, explore the application of Decision Tree (DT), k-Nearest Neighbors (KNN), Random Forest (RF), and Extra Trees (ET) algorithms for machine learning-based classification using preprocessed raw data, transfer learning features, and morphological features. How do these algorithms harness patterns within the data to make accurate predictions, and how does their performance compare across the different feature types and tumor classes?

**Author response:** Thank you. The Convolutional Neural Network (CNN) architecture of the deep learning model is shown in Figure 9. We add two convolutional layers with an input layer of (224, 224, 3) and an element-wise activation function on each of those, commonly known as a Rectified-Linear Unit (ReLu). The ability to activate the input nodes is decided by the ReLu layer. If the filters in the convolution layer pick up a visual characteristic, the activation is indicated. ReLu function operates by applying a max (0, x) function thresholded at 0. Followed by two Maxpooling layers of (2, 2), a down-sampling strategy is applied to reduce the width and height of the output volume. After adding a flattened layer, two dense layers with 128 neurons and 3 neurons output, respectively are added. The model is then trained and tested using Google Colaboratory GPU. We chose the optimizer Adam and the sparse categorical cross-entropy loss function with batch size = 64, epoch = 20 for the compilation stage to optimize the model during training and to minimize the loss function.

We have also applied seven machine learning-based classification algorithms named Decision Tree (DT), k-Nearest Neighbors (KNN), Random Forest (RF), and Extra Trees (ET), Extreme Gradient Boosting (XGB), Support vector machine (SVM) and Logistic regression (LR) for lung tumor classification. The algorithms have been introduced in page 11 as highlighted. First, the dataset of CT images is split into training and test images. 80% of the data are used for training and 20% are used for testing. For 5-fold cross-validation, the datasets are shuffled to 42 random states and split into 5 groups each containing 20% of the total datasets. Then one of the groups is taken as the test dataset and the remaining groups are taken as the training dataset. Then the proposed model is fitted on the training datasets and evaluated on the test datasets. Finally, the evaluation score is recorded for each test group and averaged to evaluate the performance of the model.

**In the revised manuscript,** 5-fold cross-validation method has been explained in **“Section 5.3 k-fold Cross-validation” (on page 12 as highlighted in yellow).**

Tables 6 and 7 present the comparative results of the seven algorithms using the preprocessed raw data, transfer learning (TL) features, and morphological features. It is observed that LR performs the best with preprocessed data and TL features, whereas, KNN and SVM perform the best with morphological features. **In the revised manuscript,** a 5-fold cross-validation method is applied using **LR, KNN, and SVM** algorithms and **the cross-validation results have been presented in Table 8.** It is observed in Table 8 that the LR performs with 99.36% and 99.27% accuracy with preprocessed data and TL features in 23.71 sec and 16.51 sec computational time, respectively. Whereas, KNN and SVM perform with 99.76% accuracy in 0.017 sec and 0.008 sec time, respectively. Thus, it may be concluded using the low dimensional morphological features KNN and SVM classification algorithms perform the best in the least computational time.



**Reviewer#2, Concern # 8:** In the morphological feature extraction and classification flow, the distinction between the training and test phases is evident, with considerations for optimal processing. Could you elucidate the specific strategies employed to ensure accurate and efficient classification in both phases, particularly in terms of the application of morphological operations for malignancy detection and their potential influence on subsequent benign and normal tumor detection? How does this dynamic adaptation contribute to the overall robustness of the classification process, and how is this tailored approach aligned with the practical goal of accurate lung tumor classification in real-world scenarios?

**Author response:** Thank you for your queries.

Figure 8 presents the process flow of ROI segmentation for Malignant, benign and normal type tumors. From the morphological feature extraction and classification flow as shown in Figure 8, the distinction between the training and test phases is evident.

As explained in section **4.2.2 Morphological operations**, seven (07) morphological operations are applied for ROI segmentation for malignant, benign, and normal tumors. Figures 5, 6 and 7 presents the ROI segmentation steps for Malignant, benign and normal type tumors. As we can see that the ROI segmentation of malignant tumors includes steps 1 to 4 of the morphological operations, ROI segmentation of Benign type tumors includes steps 1 to 6 of the morphological operations and ROI segmentation of Normal type tumors includes all the steps 1 to 7 of the morphological operations.

In the training phase, all the training images of the three types of tumors are gone through the proposed morphological operations for ROI segmentation. However, in the test phase, morphological operations for malignancy are applied first and if malignancy is detected by the classifier, then no further morphological operations are applied for ROI segmentation of Benign or Normal type tumors. Similarly, if benign tumors are detected by the classifier, then no further operations are applied for ROI segmentation of the Normal type tumors. The morphological operation applied for ROI segmentation and the process flow have been explained in detail (**on pages 7 and 8 in the revised manuscript**) in section **4.2.2 Morphological operations.**

The above dynamic adaptation process contributes to the overall robustness of the classification process by extracting distinct features for malignant, benign, and normal tumors by applying different morphological operations for ROI segmentation ofmalignant, benign, and normal tumors**.**



**Reviewer#2, Concern # 9:** As the proposed methodologies culminate in classification results, could you delve into the implications of the achieved accuracies across the various classification techniques, both deep learning and machine learning-based? How does the accuracy of the Convolutional Neural Network (CNN) compare when using preprocessed raw data, and how does it fare against the machine learning classifiers like Decision Tree (DT), k-Nearest Neighbors (KNN), Random Forest (RF), and Extra Trees (ET) when employing morphological features? Moreover, how do these accuracy metrics reflect the potential clinical applicability of the proposed methodologies in real-world lung tumor diagnosis and classification scenarios, considering the complexities of data preprocessing, feature extraction, and classification algorithms?

**Author response:** Thank you for the comments. As we can see in Table 2, CNN performs with high accuracy of 98.64% with only preprocessed raw data requiring very high computational time due to the high dimensionality of preprocessed raw data. **Whereas it is observed in the cross-validation results in Table 8 that KNN and SVM perform with a very high accuracy of 99.76% accuracy in the least possible computational time of 0.017 sec and 0.008 sec, respectively** using a very low dimensional (4x1) morphological features which is suitable for real-time applications for lung cancer diagnosis and classification.



**Response to the Reviewer # 3**

**Note: In the revised manuscript, the unchanged contents are highlighted in green for clarification of the reviewer's queries and the changed contents are highlighted in yellow.**

**Reviewer 3 Comments:**

**Suggestions to the authors**

**The paper looks well organized and exclusive experimentation has been conducted.**

**Author response:** Thank you very much for your comments.



**However few things in the paper needs correction such as:**

**Reviewer#3, Concern # 1:** **Avoid using ‘we’. In abstract ‘we’ has been used more often. Sentences should be re-framed avoiding ‘we’.**

**Author response:** Thank you for your suggestion. As per your suggestion, **the sentences in “Abstract” have been reframed avoiding ‘we’ in the revised manuscript.**



**Reviewer#3, Concern # 2:** **ii. In the abstract three classes are not specified.**

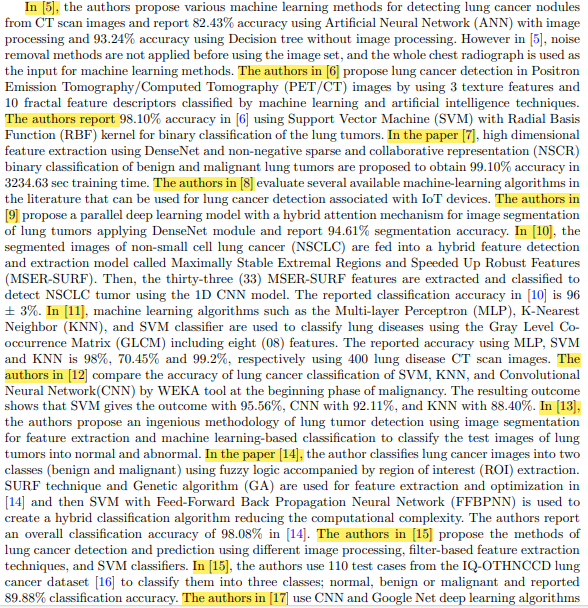
**Author response:** Thank you for your suggestion. As per your suggestion, the **three classes are specified in the revised manuscript as highlighted in yellow as follows:**

**“Lung Cancer is an uncontrolled growth of tissue causing a lump in the human lung. If lung cancer can be detected early, it can increase the survival rate. Therefore, a multi-classification approach of lung tumor detection with high computational effectiveness is required. In this paper, a multi-classification approach of Lung tumor detection and classification is proposed using artificial intelligence on Computed Tomography (CT) scan images. Different pre-processing steps are applied for resizing, smoothing, and enhancement of the CT images. Then, two different approaches for feature extraction using VGG16 transfer learning (TL) and morphological segmentation are proposed. Morphological segmentation and feature extraction are applied for the segmentation of the region of interest and to extract the distinct features. Finally, the proposed deep learning architecture and seven different machine learning algorithms are applied on the preprocessed data and the extracted features for the classification of lung tumors into three classes; malignant, benign and normal. It is observed that all the ML algorithms perform with reasonably high accuracy using the low dimensional morphological features. It is observed from the 5-fold cross-validation results that logistic regression (LR) performs with 99.36% accuracy in 23.71 sec time using the preprocessed data. Whereas, using the morphological features, k-Nearest Neighbor (KNN) and the Support Vector Machine (SVM) perform with the highest accuracy of 99.76% with very reduced computational time of 0.017 sec and 0.008 sec, respectively..”**



**Reviewer#3, Concern # 3:** **In the literature survey all works are explained in the same format i.e. “in the paper” instead the explanation can be re-written in other forms. The explanations of the works are incomplete.**

**Author response:** Thank you for your observation. As per your suggestion, the sentence in the literature survey has been reframed **by avoiding the same words as “In the paper” in the revised manuscript**. The explanations of the works have been completed by adding the limitations and trade-offs in the “Introduction” **as highlighted on page 2 in the revised manuscript as follows:**





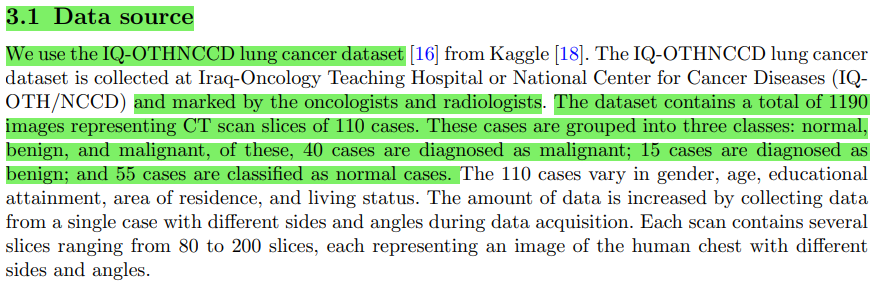
**Reviewer#3, Concern # 4:**  **Were lung segmentation results verified using ground truth?**

**Author response:** Thank you for your query. We have used **labeled datasets** of IQ-OTHNCCD lung cancer dataset marked by the oncologists and radiologists. The **datasets** are split into training and test datasets for classification. Thus, the results have been verified using ground truth original labels of the classes.

Further, in the revised manuscript, the results have been verified using 5-fold cross-validation where the datasets are shuffled to 42 random states and split into 5 groups each containing 20% of the total datasets. Then one of the groups is taken as the test dataset and the remaining groups are taken as the training dataset. Then the proposed model is fitted on the training datasets and evaluated on the test datasets. Finally, the evaluation score is recorded for each test group and averaged to evaluate the performance of the model.

5-fold cross-validation has been applied using LR, KNN and SVM classification algorithms to validate the test accuracy of all the ground truth labeled datasets.

**The data source** of the labeled datasetshas been mentioned **in Section “3.1 Data source” on page 4 highlighted as follows:**



The 5-fold cross-validation method has been explained **on page 12 in “Section 5.3 k-fold Cross-validation” of the revised manuscript (highlighted in yellow) as follows:**

**“5.3**  **k-fold Cross-validation**

**The k-fold cross-validation is applied to split the datasets and to validate the test accuracy of all the datasets [21]. We apply 5-fold cross-validation where the datasets are shuffled to 42 random states and split into 5 groups each containing 20% of the total datasets. Then one of the groups is taken as the test dataset and the remaining groups are taken as the training dataset. Then the proposed model is fitted on the training datasets and evaluated on the test datasets. Finally, the evaluation score is recorded for each test group and averaged to evaluate the performance of the model.”**

The results of the 5-fold cross-validation are included in **Table 8 in the revised manuscript as highlighted in yellow on page 19.**



**Reviewer#3, Concern # 5:** **It can be observed from Table 1 that benign cases are significantly less compared to the other two. Survey sample imbalance and its impact on classification.**

Author Response: Thank you. Yes, the data of benign cases are significantly less compared to the other two. Therefore, we have applied data augmentation to the available datasets to address the sample imbalance issue. However, the classification accuracy does not improve by applying the data augmentation.

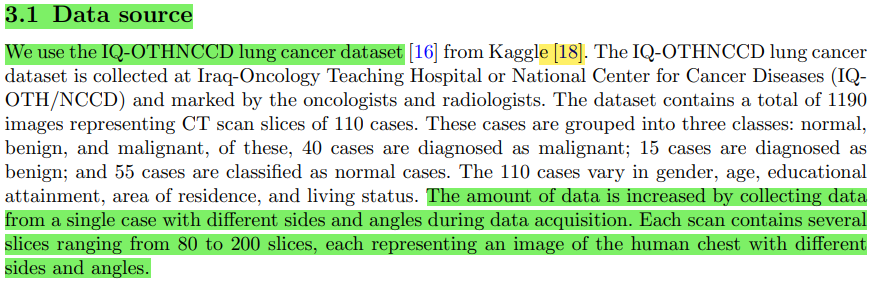
It is worth mentioning that **during the data acquisition, the amount of data is increased by collecting the data of a single case with different sides and angles** (**Ref 16 and 18 in the revised version**). The IQ-OTHNCCD lung cancer dataset [16, 18] contains a total of 1190 images representing CT scan slices of 110 cases. These cases are grouped into three classes: normal, benign, and malignant, of these, 40 cases are diagnosed as malignant; 15 cases diagnosed with benign; and 55 cases classified as normal cases. The 110 cases vary in gender, age, educational attainment, area of residence and living status. Each scan contains several slices. The number of these slices range from 80 to 200 slices, each of them represents an image of the human chest with different sides and angles **[16, 18]**.

**References:**

[16] H. Alyasriy and A. Muayed, “The IQ-OTHNCCD Lung Cancer Dataset,” Mendeley Data, vol. 1, p.2020, 2021.

[18] “The IQ-OTHNCCD Lung Cancer Dataset,” https://www.kaggle.com/datasets/antonixx/theiqothnccd-lung-cancer-dataset, [Online accessed 2022-04-19]

**The data source information and data collection process has been explained in Section “3.1 Data source” on page 4 as follows (highlighted):**



Additionally, in the revised manuscript, we have applied 5-fold cross-validation method using the LR, KNN and SVM machine learning classification algorithms to validate the test accuracy of all the datasets. We applied 5-fold cross-validation where the datasets are shuffled to 42 random states and split into 5 groups each containing 20% of the total datasets. Then one of the groups is taken as the test dataset and the remaining groups are taken as the training dataset. Then the proposed model is fitted on the training datasets and evaluated on the test datasets. Finally, the evaluation score is recorded for each test group and averaged to evaluate the performance of the model.

**The results of the 5-fold cross-validation are included in Table 8 in the revised manuscript on page 19** of the revised manuscript. It is observed in Table 8 that the LR performs with 99.36% and 99.27% accuracy with preprocessed data and TL features in 23.71 sec and 16.51 sec computational time, respectively. Whereas, KNN and SVM perform with 99.76% accuracy in 0.017 sec and 0.008 sec time, respectively. Figure 12 shows the confusion matrix for the best possible accuracy using CNN, LR and SVM. It is observed that benign tumors are detected 100% correctly using the LR and SVM.

In the revised manuscript, the k-fold cross-validation method has been discussed in **Section 5.3 k-fold Cross-validation** as highlighted.



**Reviewer#3, Concern # 6:** **Section 7 should be renamed as experimentation and results instead of simulation.**

**Author Response:** Thank you for your advice. As per your suggestion, **Section 7 has been renamed as** **“Experimentation and Results” in the revised manuscript as highlighted**.



**Reviewer#3, Concern # 7:** **It would be fair to compare the proposed work with other works performed on the same dataset.**

**Author Response:** Thank you for your suggestion.

We have used the IQ-OTHNCCD lung cancer dataset [16], [17] available in the Kaggle. **The work of Kareem et al [15] 2021** has also used the IQ-OTHNCCD lung cancer dataset. In [15], the authors propose the methods of lung cancer detection and prediction using different image processing, filter-based feature extraction techniques, and SVM classifiers. The authors in [15] use 110 test cases from the IQ-OTHNCCD lung cancer dataset [16] to classify them into three classes; normal, benign or malignant and reported 89.88% accuracy.

**The work of Kareem et al [15] 2021 has been discussed in the literature review in the “Introduction” and also compared in Table 9.**

References:

[15] H. F. Kareem, M. S. Al-Huseiny, F. Y. Mohsen, and K. Al-Yasriy, “Evaluation of svm performance in the detection of lung cancer in marked ct scan dataset,” Indonesian Journal of Electrical Engineering and Computer Science, vol. 21, no. 3, p. 1731, 2021.

[16] H. Alyasriy and A. Muayed, “The IQ-OTHNCCD Lung Cancer Dataset,” Mendeley Data, vol. 1, p.2020, 2021.

[17] “The IQ-OTHNCCD Lung Cancer Dataset,” https://www.kaggle.com/datasets/antonixx/the-iqothnccd-lung-cancer-dataset, [Online accessed 2022-04-19].



**Reviewer#3, Concern # 8:** **A technical explanation needs to be provided as to why logistic regression outperformed other classifiers.**

**Author Response:** As per the report in Ref [A], Logistic regression analysis is a powerful tool for assessing the relative importance of factors that determine outcomes. It is increasingly used in clinical medicine to develop diagnostic algorithms and evaluate prognosis. Multivariate Logistic Regression is the statistical technique used when we wish to estimate the probability of a dichotomous outcome, such as the presence or absence of disease.

The logistic regression model is used to predict response based on one or more predictor variables (e.g. features). Logistic regression measures the relationship between the categorical dependent variable and one or more independent variables by estimating probabilities using a logistic function, which is the cumulative logistic distribution. The term regression comes from the fact that we are fitting a linear model to the feature space. Logistic Regression involves a more probabilistic view of classification. Logistic regression models are usually computationally less complicated to build and require less computation time to train as compared with CNNs.

References:

[A] Pal, Ankita, “Logistic regression: A simple primer”, Cancer Research, Statistics, and Treatment 4(3):p 551-554, Jul–Sep 2021, Wolters Kluwer – Medknow.



**Reviewer#3, Concern # 9:** **Precision, recall, and F-score of each class need to be presented.**

**Author Response:** Thank you.Precision, recall, and F-score of all the classes have been presented in Tables 3, 4, and 5 using preprocessed raw data, TL features and morphological features. The confusion matrix of all the classes have also been presented for the best possible accuracies in Figure 16(a), (b), (c) and (d).



**Reviewer#3, Concern # 10:** **In conclusion ‘can’ word usage is improper. Ex: machine learning-based logistic regression (LR) can acquire 100% accuracy. This sentence can be re-framed as machine learning-based logistic regression (LR) acquires 100% accuracy.**

**Author response:** Thank you for your suggestion. As per your suggestion, the sentences in the Conclusion have been reframed avoiding ‘can’ word in the revised manuscript as follows:

**“Early detection and classification of lung tumors are required to increase the survival rate of the patients. Thus, a system is required with high computational intelligence and low complexity. In this paper, we have proposed lung tumor detection and classification using different preprocessing steps, ROI segmentation, feature extraction, and classification methods. First, we have applied different deep learning and machine learning-based classification algorithms to the preprocessed raw CT scan data. It is observed that using the deep learning-based CNN model, 98.64% accuracy is acquired and using machine learning-based logistic regression (LR), 100% accuracy is acquired with the preprocessed raw dataset. However, since preprocessed raw data requires a high processing time, we have applied two feature extraction methods to extract low-dimensional features to reduce the processing time. First, we have applied the VGG16 model to extract transfer learning (TL) features of the preprocessed data. Hence, we have obtained 95.928% accuracy using deep learning CNN and 100% accuracy using the LR method with the extracted TL features with a comparatively low processing time of training. Then, to further improve the computational intelligence, we have proposed morphological segmentation and feature extraction to extract low-dimensional features. The extracted morphological features are applied for classification using machine learning algorithms. It is observed that all the machine learning algorithms show significantly high classification accuracy with reduced training time using the low dimensional morphological features. Then, 5-fold cross-validation methods is applied to mitigate the overfitting problem. It is observed that LR performed with 99.36% and 99.27% accuracy using preprocessed data and TL features in 23.71 sec and 16.51 sec computational time, respectively. Whereas, both KNN and SVM performed with 99.76% accuracy in 0.017 sec and 0.008 sec time, respectively. Thus, it may be concluded that a significantly high accuracy of lung tumor detection and classification is possible to be achieved with the machine learning algorithms applied on the extracted low dimensional morphological features with low computational complexity and time. Therefore, the proposed methods recommend extracting low-dimensional morphological features and choosing machine learning-based classifiers for real-time lung cancer recognition.”**



**Reviewer#3, Concern # 11:** **The authors are suggested to refer paper after 2021 and update the literature survey.**

**Author Response:** Thank you for your suggestion. As per your suggestion, a recent work of **Pandian et al [17] of 2022** has been added to the literature review in the “Introduction” and also compared in Table 9. In [17], the authors use CNN and Google Net deep learning algorithms for lung cancer detection and binary classification and achieve a precision of 98% in detection and classification.

In the revised version of the manuscript, the work [17] has been discussed and highlighted in the “Introduction” as follows and also compared in Table 9 as highlighted.

**“In [17], the authors use CNN and Google Net deep learning algorithms for lung cancer detection and binary classification and achieve a precision of 98% in detection and classification.”**



**Reviewer#3, Concern # 12:** **The authors need to clearly specify why and how higher accuracy was achieved using transfer learning features and not from morphological features.**

**Author Response:** Thank you for your observation. Tables 6 and 7 present the comparative results of the seven ML algorithms using the preprocessed raw data, transfer learning (TL) features, and morphological features. It is observed that LR performs the best with greater dimensional preprocessed data and TL features. Whereas, SVM performs the best with low dimensional morphological features as it is not suitable for large datasets. **In the revised manuscript, a 5-fold cross-validation method is applied using LR, KNN, and SVM algorithms and the cross-validation results have been presented in Table 8.** It is observed from the cross-validation results in Table 8 that the LR performs with 99.36% and 99.27% accuracy with preprocessed data and TL features in 23.71 sec and 16.51 sec computational time, respectively. Whereas, KNN and SVM perform with 99.76% accuracy in 0.017 sec and 0.008 sec time, respectively. Thus, it may be concluded using the low dimensional morphological features KNN and SVM classification algorithms perform the best in the least computational time. Thus, for real-time applications, KNN and SVM classifiers are better choice to use with low dimensional morphological features or small datasets.



**Reviewer#3, Concern # 13:** **However not much difference is observed between the results of TL and morphological features results. Then justify the need for TL over feature extraction.**

**Author Response:** Thank you for your observation. As per the results in Tables 6 and 7 and the cross-validation results in Table 8, it is observed that the LR performs with significantly high accuracy using preprocessed data and TL features. However, LR requires high computational time 23.71 sec and 16.51 sec to process preprocessed data and TL features, respectively. Whereas, SVM performs with 99.76% accuracy in a very low computational time of 0.008 sec only using the low dimensional morphological features. Thus, SVM classifier **with low-dimensional morphological features** would be a better choice for real-time applications with very low computational time.

In the revised manuscript, the above observations have been added in the last para of the “Experimentation and Results” section as follows:

**“Tables 6 and 7 present the comparative results of the seven ML algorithms using the preprocessed raw data, transfer learning (TL) features, and morphological features. It is observed that LR performs the best with preprocessed data and TL features. Whereas, using the low dimensional morphological features, all the ML algorithms perform with reasonably high accuracy among which KNN and SVM perform the best with the least computational time. Therefore, 5-fold cross-validation is applied to mitigate overfitting problems and to verify the best possible accuracy using the LR, KNN, and SVM algorithms. The cross-validation results are presented in Table 8. It is observed in Table 8 that LR performs with 99.36% and 99.27% accuracy with preprocessed data and TL features in 23.71 sec and 16.51 sec computational time, respectively. Whereas, both KNN and SVM perform with 99.76% accuracy in 0.017 sec and 0.008 sec time, respectively. Thus, it may be concluded that using the low dimensional morphological features KNN and SVM classification algorithms perform the best in the least computational time. Table 9 compares the accuracy of the proposed works to other related works in the literature. It is apparent from the comparison that the proposed works outperform the literature with high computational intelligence and accuracy.”**



**Response to the Reviewer # 4**

**Note: In the revised manuscript, the unchanged contents are highlighted in green for clarification of the reviewer's queries and the changed contents are highlighted in yellow.**

**Reviewer # 4 Concern 1: How it is having less complexity and high computational intelligence. Justify**

**Author Response:** Thank you for your observation. We have applied two feature extraction methods using VGG16 transfer learning model and morphological feature extraction that reduces the complexity and increases the computational intelligence.

A deep Convolutional neural network (CNN) is difficult and expensive to train with small datasets and complex models. The transfer learning model can perform with improved accuracy with small datasets. When a pre-trained model is repurposed for a different related task is known as transfer learning. We applied VGG16 [18] pre-trained model to implement transfer learning (TL) for feature extraction using 16 convolutional layers including the Maxpooling layers, 3 dense layers (2 fully connected layers and 1 SoftMax classifier), and an output layer of 1,000 nodes as shown in Figure 4. VGG16 is deeper, less complex, and faster than other transfer learning models as VGG19, Inception V3, XCeption, and ResNet-50. VGG16 has an exceptional feature extraction capability as it has a greater capacity to learn new features because it is deeper than certain transfer learning models, such as AlexNet. VGG16 uses just 3×3 convolution layers and 2×2 pooling layers repeatedly, which makes it significantly less complex than other transfer learning models like InceptionNet and enables it to generalize and adapt more effectively to a larger variety of data sets. VGG19 is a similar type of CNN model with 19 layers. However, due to the increased number of CNN layers of VGG19, the ability to fit complex functions also increases which, in turn, increases the training time of VGG19 significantly.

For morphological feature extraction, 7 types of morphological operations are applied on the thresholded binary images of lung tumors for ROI segmentation of the malignant, benign, and normal type tumors as shown in Figures 5, 6, and 7. Then, four morphological features (area, eccentricity, perimeter, and compactness) of the segmented ROIs are extracted before applying the classifiers. By applying morphological feature extraction, the dimension of the extracted features is significantly reduced as compared to the dimension of the original raw data. The dimension of the extracted features is 4×1 double containing the four extracted features of each segmented region. The extracted morphological features are classified using machine learning and deep learning classifiers. **As the morphological feature extraction method compresses the size of the training and test data significantly by extracting low dimensional features, it results in less computational complexity by reducing the training time of classification.**

As per the results in Tables 6 and 7 and the cross-validation results in Table 8, it is observed that the LR performs with significantly high accuracy using preprocessed data and TL features in high computational time 23.71 sec and 16.51 sec, respectively. Whereas, SVM performs with 99.76% accuracy in a very low computational time of 0.008 sec only using the low dimensional morphological features. Thus, SVM classifier with low-dimensional morphological features is suitable for real-time applications time having less complexity and high computational intelligence with very low computational time.



**Reviewer # 4 Concern 2: Sub figures should be named (Figure 6 and 7)**

**Author Response:** Thank you for your suggestion. The figures have been named as Figure 6 or Figure 7, etc. In the revised manuscript, the sub-figures of Figure 16 have been named **Figure 16(a), 16(b),… and so on (highlighted on page 18 of the revised manuscript).**



**Reviewer # 4 Concern 3: In the flow chart why ROI segmentation is shown separately for each type. Any specific reason (Figure 8)**

**Author Response:** Thank you for your query. Figure 8 presents the process flow of ROI segmentation for Malignant, benign, and normal type tumors. From the morphological feature extraction and classification flow as shown in Figure 8, the distinction between the training and test phases is evident.

As explained in section **4.2.2 Morphological operations**, seven (07) morphological operations are applied for ROI segmentation for malignant, benign, and normal tumors. Figures 5, 6 and 7 present the ROI segmentation steps for Malignant, benign, and normal type tumors. As we can see that the ROI segmentation of malignant tumors includes steps 1 to 4 of the morphological operations, ROI segmentation of Benign type tumors includes steps 1 to 6 of the morphological operations, and ROI segmentation of Normal type tumors includes all the steps 1 to 7 of the morphological operations.

In the training phase, all the training images of the three types of tumors are gone through the proposed morphological operations for ROI segmentation. However, in the test phase, morphological operations for malignancy are applied first and if malignancy is detected by the classifier, then no further morphological operations are applied for ROI segmentation of Benign or Normal type tumors. Similarly, if benign tumors are detected by the classifier, then no further operations are applied for ROI segmentation of the Normal type tumors. The morphological operation applied for ROI segmentation and the process flow have been explained in detail (as highlighted in the revised manuscript) in section **4.2.2 Morphological operations.**

The above dynamic adaptation process contributes to the overall robustness of the classification process by extracting distinct features for malignant, benign, and normal tumors by applying different morphological operations for ROI segmentation ofmalignant, benign, and normal tumors**.**



**Reviewer # 4 Concern 4: Latest references were not included.**

**Author Response:** Thank you for your observation. In the manuscript, the latest references [5]-[16] were included and discussed in the literature review in the Introduction and have been compared in Table 9.

As per your suggestion, in the revised manuscript. a recent work of **Pandian et al [17] 2022** has been added to the literature review in the “Introduction” and also compared in Table 9. In [17], the authors use CNN and Google Net deep learning algorithms for lung cancer detection and binary classification and achieve a precision of 98% in detection and classification.

**In the revised version of the manuscript,** the work [17] has been discussed and **highlighted in the “Introduction” on page 2 as follows and also compared in Table 9 as highlighted**.

**“In [17], the authors use CNN and Google Net deep learning algorithms for lung cancer detection and binary classification and achieve a precision of 98% in detection and classification.”**

**In Table 9, the year of the reported works has been added and highlighted.**



**Reviewer # 4 Concern 5: add more discussion and future directions to the research.**

**Author Response:** Thank you for your suggestion. More discussion has been added in “Experimentation and Results” as follows:

**“Tables 6 and 7 present the comparative results of the seven ML algorithms using the preprocessed raw data, transfer learning (TL) features, and morphological features. It is observed that LR performs the best with preprocessed data and TL features. Whereas, using the low dimensional morphological features, all the ML algorithms perform with reasonably high accuracy among which KNN and SVM perform the best with the least computational time. Therefore, 5-fold cross-validation is applied to mitigate overfitting problems and to verify the best possible accuracy using the LR, KNN, and SVM algorithms. The cross-validation results are presented in Table 8. It is observed in Table 8 that LR performs with 99.36% and 99.27% accuracy with preprocessed data and TL features in 23.71 sec and 16.51 sec computational time, respectively. Whereas, both KNN and SVM perform with 99.76% accuracy in 0.017 sec and 0.008 sec time, respectively. Thus, it may be concluded that using the low dimensional morphological features KNN and SVM classification algorithms perform the best in the least computational time. Table 9 compares the accuracy of the proposed works to other related works in the literature. It is apparent from the comparison that the proposed works outperform the literature with high computational intelligence and accuracy.”**

More discussion and future direction have been added in the “Conclusion” as highlighted below:

**“Early detection and classification of lung tumors are required to increase the survival rate of the patients. Thus, a system is required with high computational intelligence and low complexity. In this paper, we have proposed lung tumor detection and classification using different preprocessing steps, ROI segmentation, feature extraction, and classification methods. First, we have applied different deep learning and machine learning-based classification algorithms to the preprocessed raw CT scan data. It is observed that using the deep learning-based CNN model, 98.64% accuracy is acquired and using machine learning-based logistic regression (LR), 100% accuracy is acquired with the preprocessed raw dataset. However, since preprocessed raw data requires a high processing time, we have applied two feature extraction methods to extract low-dimensional features to reduce the processing time. First, we have applied the VGG16 model to extract transfer learning (TL) features of the preprocessed data. Hence, we have obtained 95.928% accuracy using deep learning CNN and 100% accuracy using the LR method with the extracted TL features with a comparatively low processing time of training. Then, to further improve the computational intelligence, we have proposed morphological segmentation and feature extraction to extract low-dimensional features. The extracted morphological features are applied for classification using machine learning algorithms. It is observed that all the machine learning algorithms show significantly high classification accuracy with reduced training time using the low dimensional morphological features. Then, 5-fold cross-validation methods is applied to mitigate the overfitting problem. It is observed that LR performed with 99.36% and 99.27% accuracy using preprocessed data and TL features in 23.71 sec and 16.51 sec computational time, respectively. Whereas, both KNN and SVM performed with 99.76% accuracy in 0.017 sec and 0.008 sec time, respectively. Thus, it may be concluded that a significantly high accuracy of lung tumor detection and classification is possible to be achieved with the machine learning algorithms applied on the extracted low dimensional morphological features with low computational complexity and time. Therefore, the proposed methods recommend extracting low-dimensional morphological features and choosing machine learning-based classifiers for real-time lung cancer recognition.”**



**Reviewer # 4 Concern 6: Data can be splitted into different proportions and can be verified**

**Author Response:** Thank you for your thoughtful observation.

In the revised manuscript, we have applied 5-fold cross-validation to split the data (preprocessed raw data, extracted VGG16 transfer learning (TL) features and the morphological features) into 5 groups each containing 20% of the total datasets and verified the classification accuracy.

As we can see in Tables 6 and 7, the best accuracy of 100% is obtained using a specific proportion of preprocessed raw data and the extracted VGG16 TL features using the LR classifier. However, the best accuracy in the least possible computational time has been obtained using the KNN and SVM classifiers on the extracted morphological features. Therefore, 5-fold cross-validation has been applied to verify the accuracy of all the datasets using LR, KNN and SVM classifiers only. During 5-fold cross-validation, the datasets are shuffled to 42 random states and split into 5 groups each containing 20% of the total datasets. Then one of the groups is taken as the test dataset and the remaining groups are taken as the training dataset. Then the proposed model is fitted on the training datasets and evaluated on the test datasets. Finally, the evaluation score is recorded for each test group and averaged to evaluate the performance of the model.

**The 5-fold cross-validation results are presented in Table 8 as highlighted on page 19 of the revised manuscript.**

As we can see from the 5-fold cross-validation results in Table 8 that LR performs with significantly high accuracy of 99.36% and 99.27% in 23.71 sec and 16.51 sec computational time using the preprocessed raw data and the TL features, respectively. It is also observed in the 5-fold cross-validation results in Table 8 that KNN and SVM perform better with low dimensional morphological features with 99.76% accuracy in a very low computational time of 0.017 sec and 0.008 sec, respectively. Thus, for real-time applications, KNN and SVM classifiers would be the better choice to classify the lung tumors with high accuracy using the low-dimensional morphological features in a very low computational time.

In the revised manuscript, the k-fold cross-validation method has been discussed **in Section 5.3 k-fold Cross-validation of the revised manuscript as highlighted follows:**

**“5.3**  **k-fold Cross-validation**

**The k-fold cross-validation is applied to split the datasets and to validate the test accuracy of all the datasets [21]. We apply 5-fold cross-validation where the datasets are shuffled to 42 random states and split into 5 groups each containing 20% of the total datasets. Then one of the groups is taken as the test dataset and the remaining groups are taken as the training dataset. Then the proposed model is fitted on the training datasets and evaluated on the test datasets. Finally, the evaluation score is recorded for each test group and averaged to evaluate the performance of the model.”**

Table 8 has been cited and the results have been updated **in the revised manuscript on page 19** as follows:

**“Tables 6 and 7 present the comparative results of the seven ML algorithms using the preprocessed raw data, transfer learning (TL) features, and morphological features. It is observed that LR performs the best with preprocessed data and TL features. Whereas, using the low dimensional morphological features, all the ML algorithms perform with reasonably high accuracy among which KNN and SVM perform the best with the least computational time. Therefore, 5-fold cross-validation is applied to mitigate overfitting problems and to verify the best possible accuracy using the LR, KNN, and SVM algorithms. The cross-validation results are presented in Table 8. It is observed in Table 8 that LR performs with 99.36% and 99.27% accuracy with preprocessed data and TL features in 23.71 sec and 16.51 sec computational time, respectively. Whereas, both KNN and SVM perform with 99.76% accuracy in 0.017 sec and 0.008 sec time, respectively. Thus, it may be concluded that using the low dimensional morphological features KNN and SVM classification algorithms perform the best in the least computational time. Table 9 compares the accuracy of the proposed works to other related works in the literature. It is apparent from the comparison that the proposed works outperform the literature with high computational intelligence and accuracy.”**



**Reviewer # 4 Concern 7: no novelity in the paper. why only specific features are extracted**

**Author Response:** Thank you for your observation.

In the manuscript, we have proposed the method of early detection of lung cancer by classifying the lung tumor into three classes; Malignant, benign and normal. In this paper, we have proposed novel method of feature extraction using morphological segmentation and feature extraction. We have also used VGG16 transfer learning for feature extraction which is deeper, less complex, and faster than other transfer learning models and can perform with high accuracy with small datasets. We have compared the accuracy and computational time using both feature extraction methods. As per the results in Tables 6 and 7 and the cross-validation results in Table 8, It has been observed that LR performs with significantly high accuracy of 99.36% and 99.27% using the preprocessed raw data and the TL features, respectively. However, LR requires comparatively high computational time. It has been also observed in Table 8 that KNN and SVM performs better with low dimensional morphological features with 99.76% accuracy in a very low computational time of 0.017 sec and 0.008 sec, respectively. Thus, for real-time applications, KNN and SVM classifiers would be the better choice to classify the lung tumors with high accuracy using the low-dimensional morphological features in a very low computational time. The proposed methods include several preprocessing steps, ROI segmentation, feature extraction, and classification methods. The results have been compared to the existing methods in the literature in Table 9. It was observed that the proposed method outperforms the literature with in terms of accuracy and computational time and suitable for real time applications for early detection of lung cancer.

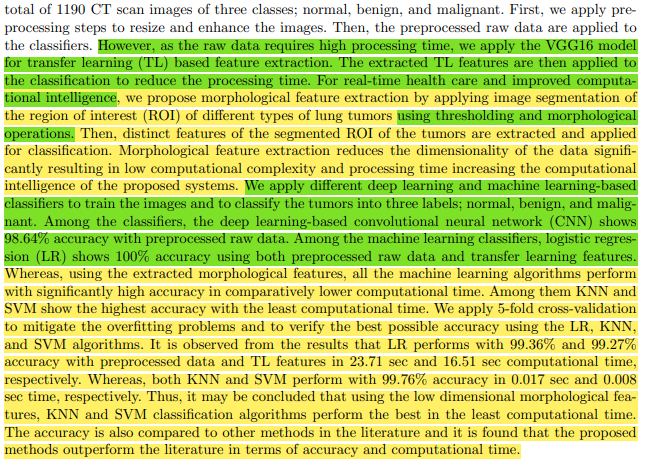
The research gap of the existing methods and the major contribution of the present manuscript has been updated in the “**Introduction” as follows:**

**Para 3 of Introduction:**

**“Though there are several works in the literature, most of the works are on binary classification of lung tumors or predicting a specific type of cancer without applying cross-validation. Additionally, the computational time has not been reported in most of the above works. Hence, a multi-classification approach with high computational effectiveness is required for the fast detection and classification of lung tumors. In this paper, we propose a multi-classification approach for lung tumor detection and classification as normal, benign, and malignant with high computational effectiveness by applying machine learning and deep learning methods on the CT scan images. We use the IQ-OTHNCCD lung cancer dataset [16] collected from the Kaggle [18]. The dataset contains a”**

VGG16 TL features have been extracted as it is deeper, less complex, and faster than other transfer learning models and can perform with high accuracy with small datasets. Morphological features have been extracted to reduce the dimensionality of the features and to remove the outliers for real-time application and early detection of lung cancer. The effectiveness of the proposed feature extraction methods are explained detail in the “**4. Feature Extraction”** section **from pages 5 to 10**.

The impact of the proposed feature extraction methods on the achieved accuracies has been explained in **Para 2 of the “Introduction” on page 3 as highlighted in the revised manuscript as follows (changes highlighted in yellow):**



The impact of the proposed morphological feature extraction method on the accuracy and computational time **has been presented and compared in Tables 5, 6, 7 and 8 (as highlighted in the revised manuscript**). **In the revised manuscript, 5-fold cross-validation has been applied using LR, KNN, and SVM classifiers to mitigate the overfitting problem and to find the best possible accuracy.** **The results of 5-fold cross-validation are presented in Table 8.** As we can see in Tables 5, 6 and 7, using the proposed morphological feature extraction method, KNN and SVM can obtain significantly high accuracy using very low dimensional features of 4x1 dimension in a significantly reduced computational time which indicates the high computational effectiveness of the proposed methods. It is also observed in Table 8 that using the morphological features, both KNN and SVM classification algorithms perform the best in the least computational time. **Table 8 has been cited and the impact and results using the feature extraction have been compared** in terms of accuracy, computational complexity, and processing time against the established techniques **on page 19 in the revised manuscript as follows (highlighted in yellow):**

**“Tables 6 and 7 present the comparative results of the seven ML algorithms using the preprocessed raw data, transfer learning (TL) features, and morphological features. It is observed that LR performs the best with preprocessed data and TL features. Whereas, using the low dimensional morphological features, all the ML algorithms perform with reasonably high accuracy among which KNN and SVM perform the best with the least computational time. Therefore, 5-fold cross-validation is applied to mitigate overfitting problems and to verify the best possible accuracy using the LR, KNN, and SVM algorithms. The cross-validation results are presented in Table 8. It is observed in Table 8 that LR performs with 99.36% and 99.27% accuracy with preprocessed data and TL features in 23.71 sec and 16.51 sec computational time, respectively. Whereas, both KNN and SVM perform with 99.76% accuracy in 0.017 sec and 0.008 sec time, respectively. Thus, it may be concluded that using the low dimensional morphological features KNN and SVM classification algorithms perform the best in the least computational time. Table 9 compares the accuracy of the proposed works to other related works in the literature. It is apparent from the comparison that the proposed works outperform the literature with high computational intelligence and accuracy.”**

The accuracy and computational effectiveness of the proposed morphological feature extraction method to the reported existing works in the literature have been presented and compared **in Table 9 of the revised manuscript. Three new columns named “Binary/Multi-class” and “Cross-validation” and “Training time” have been also added in Table 9 to compare the computational complexity. Table 9 on page 20 has been updated in the revised manuscript (changes highlighted in yellow).**



**Reviewer # 4 Concern 8: Results show that accuracy is similar to the state of art techniques.so how it is different?**

**Author Response:** Thank you for your observation. The results of the proposed methods have been compared to the existing methods in the literature in Table 9. It was observed that the proposed method outperforms the literature in terms of accuracy and computational time. Using the proposed morphological feature extraction, the proposed method requires very low computational time which is suitable for real-time applications and early detection of lung cancer.

The accuracy and computational effectiveness of the reported works to the proposed methods have also been presented and compared **in Table 9 of the revised manuscript.** Three new columns named **“Binary/Multi-class”, “Cross-validation” and “Training time”** have been added in **Table 9 to compare the accuracy and computational effectiveness**.

The impact of the proposed morphological feature extraction method on the accuracy and computational time **has been presented and compared in Tables 5 and 6** as highlighted. As we can see in Tables 5 and 6, using the proposed morphological feature extraction method, KNN and SVM can obtain significantly high accuracy using very low dimensional features of 4x1 dimension in a significantly reduced computational time which indicates the high computational effectiveness of the proposed methods. I**n the revised manuscript, 5-fold cross-validation has been applied using LR, KNN, and SVM classifiers to mitigate the overfitting problem and computational complexity.** 5-fold cross-validation has been applied on the preprocessed data, VGG16 TL features and morphological features of all the datasets where the datasets are shuffled to 42 random states and split into 5 groups each containing 20% of the total datasets. **The results of 5-fold cross-validation are presented in Table 8.** It is observed that using the low dimensional morphological features both KNN and SVM classification algorithms perform the best in the least computational time. Table 8 has been cited and the impact and results using the feature extraction have been compared in terms of accuracy, computational complexity, and processing time against the established techniques on page 19 in the revised manuscript as follows:

**“Tables 6 and 7 present the comparative results of the seven ML algorithms using the preprocessed raw data, transfer learning (TL) features, and morphological features. It is observed that LR performs the best with preprocessed data and TL features. Whereas, using the low dimensional morphological features, all the ML algorithms perform with reasonably high accuracy among which KNN and SVM perform the best with the least computational time. Therefore, 5-fold cross-validation is also applied to mitigate overfitting problems and to verify the best possible accuracy using the LR, KNN, and SVM algorithms. The cross-validation results are presented in Table 8. It is observed in Table 8 that LR performs with 99.36% and 99.27% accuracy with preprocessed data and TL features in 23.71 sec and 16.51 sec computational time, respectively. Whereas, both KNN and SVM perform with 99.76% accuracy in 0.017 sec and 0.008 sec time, respectively. Thus, it may be concluded that using the low dimensional morphological features KNN and SVM classification algorithms perform the best in the least computational time. Table 9 compares the accuracy of the proposed works to other related works in the literature. It is apparent from the comparison that the proposed works outperform the literature with high computational intelligence and accuracy.”**



**Response to the Reviewer # 5**

**Note: In the revised manuscript, the unchanged contents are highlighted in green for clarification of the reviewer's queries and the changed contents are highlighted in yellow.**

**Good work and new approach to feature extraction**

**Author Response:** Thank you very much.

**Reviewer # 5 Concern 1. Why is VGG NET specifically used for the work of feature extraction?**

**Author Response:** Thank you for your observations. We have used the VGG16 transfer learning model as shown in Figure 4 for feature extraction. We develop the TL feature extraction method using the VGG16 model of 16 convolutional layers including the Maxpooling layers, 3 dense layers (2 fully connected layers and 1 SoftMax classifier), and an output layer of 1,000 nodes.

A deep Convolutional neural network (CNN) is difficult and expensive to train with small datasets and complex models. When a pre-trained model is repurposed for a different related task is known as transfer learning. Since the transfer learning model can perform with improved accuracy with small datasets, we apply VGG16 [18] pre-trained model to implement transfer learning (TL) for feature extraction. VGG16 is deeper, less complex, and faster than other transfer learning models as VGG19, Inception V3, XCeption, and ResNet-50. VGG16 has an exceptional feature extraction capability as it has a greater capacity to learn new features because it is deeper than certain transfer learning models, such as AlexNet. VGG16 uses just 3×3 convolution layers and 2×2 pooling layers repeatedly, which makes it significantly less complex than other transfer learning models like InceptionNet and enables it to generalize and adapt more effectively to a larger variety of data sets. VGG19 is a similar type of CNN model with 19 layers. However, due to the increased number of CNN layers of VGG19, the ability to fit complex functions also increases which, in turn, increases the training time of VGG19 significantly.

VGG16 may be applied with data augmentation to prevent overfitting and improve accuracy. However, we apply VGG16 without data augmentation as it is faster than the method with data augmentation. To mitigate overfitting in the transfer learning process, we applied 5-fold cross-validation on the VGG16 features of all the datasets using Logistic regression (LR) machine learning classifier. In the 5-fold cross-validation, the datasets are shuffled to 42 random states and split into 5 groups each containing 20% of the total datasets.

The reason of choosing VGG16 Transfer learning model has been explained **in “Section 4.1 Transfer learning for Feature extraction” of the manuscript.**

**In the revised manuscript, the results of the 5-fold cross-validation method are presented in Table 8.** In the revised manuscript, the k-fold cross-validation method has been discussed **in Section 5.3 k-fold Cross-validation of the revised manuscript as highlighted follows:**

**“5.3**  **k-fold Cross-validation**

**The k-fold cross-validation is applied to split the datasets and to validate the test accuracy of all the datasets [21]. We apply 5-fold cross-validation where the datasets are shuffled to 42 random states and split into 5 groups each containing 20% of the total datasets. Then one of the groups is taken as the test dataset and the remaining groups are taken as the training dataset. Then the proposed model is fitted on the training datasets and evaluated on the test datasets. Finally, the evaluation score is recorded for each test group and averaged to evaluate the performance of the model.”**

Table 8 has been cited and the results has been also updated in the revised manuscript on page 19 as follows:

**“Tables 6 and 7 present the comparative results of the seven ML algorithms using the preprocessed raw data, transfer learning (TL) features, and morphological features. It is observed that LR performs the best with preprocessed data and TL features. Whereas, using the low dimensional morphological features, all the ML algorithms perform with reasonably high accuracy among which KNN and SVM perform the best with the least computational time. Therefore, 5-fold cross-validation is applied to mitigate overfitting problems and to verify the best possible accuracy using the LR, KNN, and SVM algorithms. The cross-validation results are presented in Table 8. It is observed in Table 8 that LR performs with 99.36% and 99.27% accuracy with preprocessed data and TL features in 23.71 sec and 16.51 sec computational time, respectively. Whereas, both KNN and SVM perform with 99.76% accuracy in 0.017 sec and 0.008 sec time, respectively. Thus, it may be concluded that using the low dimensional morphological features KNN and SVM classification algorithms perform the best in the least computational time. Table 9 compares the accuracy of the proposed works to other related works in the literature. It is apparent from the comparison that the proposed works outperform the literature with high computational intelligence and accuracy.”**



**Reviewer # 5 Concern 2. After feature extraction with VGG NET, why are morphological features extracted? What is the purpose?**

**Author Response:** Thank you for your query. After the VGG16 feature extraction, we have not applied the Morphological feature extraction. Rather, we have proposed two methods of feature extraction using VGG16 feature extraction and Morphological feature extraction to compare the accuracy of classification using Machine learning and deep learning methods. The feature extraction methods are explained as below:

Since the transfer learning model can perform with improved accuracy with small datasets, we apply VGG16 [18] pre-trained model has been chosen to implement transfer learning (TL) for feature extraction as it is deeper, less complex, and faster than other transfer learning models. VGG16 has an exceptional feature extraction capability as it has a greater capacity to learn new features because it is deeper than certain transfer learning models which makes it significantly less complex than other transfer learning models and enables it to generalize and adapt more effectively to a larger variety of data sets.

Before VGG16 TL feature extraction, the datasets are preprocessed using scaling and normalization. After applying the TL feature extraction, a new dataset is created from the input image dataset of lung tumors using the pre-trained model, and a three-dimensional feature stack containing the recognized visual features. The extracted features are then fed into the machine learning (ML) and deep learning (DL) classifier for lung tumor classification.

For morphological feature extraction, image cropping and enhancement is applied as the preprocessing steps as shown in Figure 3. Then, 7 types of morphological operations are applied on the thresholded binary images of lung tumors for ROI segmentation of the malignant, benign, and normal type tumors as shown in Figures 5, 6, and 7. The ROI segmentation processes are shown in the flow chart in Figure 8. Finally, the four morphological features (area, eccentricity, perimeter, and compactness) of the segmented ROIs are extracted before applying them to the classifiers. By applying morphological feature extraction, the dimension of the extracted features is significantly reduced as compared to the dimension of the original raw data. The dimension of the extracted features is 4×1 double containing the four extracted features of each segmented region. The extracted morphological features are classified using machine learning and deep learning classifiers. As morphological feature extraction compresses the size of the training and test data significantly, it results in less computational complexity and reduced training and test time of classification.

**In the revised manuscript**, **the preprocessing, feature extraction, and classification methods are explained in detail in “Section 3 Preprocessing”, “Section 4 Feature Extraction” and “Section 5 Classification”**. **The captions of sections 3, 4, and 5 and the concerned figures are highlighted.**

The results of the feature extraction in terms of accuracy and computational time are presented in Tables 6 ,7 and 8. It has been compared to other methods in the literature in Table 9. As we can that by applying the TL feature and morphological feature extraction, we can obtain significantly high recognition accuracy in lower computational time. The morphological feature extraction can obtain the highest accuracy in the least computational time which is suitable for real-time application and for early detection of lung cancer.



**Reviewer # 5 Concern 3. Classification using deep learning or machine learning must be defined clearly. Both techniques are explained, but which one is implemented is not appropriate.**

**Author Response:** Thank you for your advice. The classification using deep learning or machine learning has been defined clearly and explained in the Section **“Classification” on pages 10, 11 and 12.**

Both deep learning and machine learning methods have been applied to compare the accuracy and computational time. The results of deep learning CNN has been presented in Table 2, Figures 11 and 12. The evaluation scores using 5 Machine learning (ML) algorithms have been presented in Tables 3, 4 and 5. The accuracy and computational time of the ML methods using raw data and extracted features have been compared in Tables 6 and 7. Then the 5-fold cross-validation results of the best algorithms have been presented and compared in Table 8.



**Reviewer # 5 Concern 4. Whether the features are derived from binary or gray-scale images? Can 3D images be used? If so why not in this process of work?**

**Author Response:** Thank you for your observation. For morphological feature extraction in MATLAB, the features are derived from binary-scale images. For Transfer learning (TL) feature extraction in Python, Scaling and Normalization are applied normalize the data such that the feature value remains within a specific range of 0 and 1.

The dataset used in this research, the “IQ-OTH/NCCD lung cancer dataset” was collected over three months in fall 2019 at the Iraq-Oncology Teaching Hospital/National Center for Cancer Diseases (IQ- OTH/NCCD). It includes CT scans of patients diagnosed with lung cancer in different stages, as well as healthy subjects. The dataset is diverse, covering various demographics and backgrounds, including gender, age, educational attainment, area of residence, and occupation. The CT scans were originally collected in DICOM format using a Siemens SOMATOM scanner with specific imaging parameters.

In our research, we used 2D cross-sectional slices extracted from CT scans for the purpose of lung nodule detection and classification.

To extract the Transfer learning (TL) features in Python, Scaling and Normalization are applied to convert the data into binary data. During the preprocessing using Scaling and normalization, BGR color space images are loaded and converted to RGB format. Next, we applied scaling to transform data so that it fits within a specific scale and to change the range of data. The normalization is applied to change the observations so that they can be described as a normal distribution. Then we used min-max scaling to normalize the data such that the feature value remains within a specific range of 0 and 1. Then the TL features are extracted.

For morphological feature extraction in MATLAB, the morphological features are extracted using binary images. During the preprocessing using Image cropping and enhancement, the RGB images are loaded and converted to grayscale images. After applying a few preprocessing steps, the grayscale images are converted to binary images. Then the morphological features are extracted.

The choice to use 2D images was driven by the consideration of computational complexity. Analyzing 3D volumes would require significantly greater computational resources and time, as it involves processing a larger dataset and often necessitates more complex algorithms. Furthermore, 2D images are commonly used in machine learning, particularly in the context of deep learning with Convolutional Neural Networks (CNNs). Traditional machine learning models like logistic regression, decision trees, random forests, SVM, and KNN are generally more appropriate for structured data and may require image flattening to 1D for compatibility.

References:

Alyasri, Hamdalla; AL-Huseiny, Muayed (2021), “The IQ-OTHNCCD lung cancer dataset,” Mendeley Data, Version 2, doi: 10.17632/bhmdr45bh2.2



**Response to the Reviewer # 6**

**Note: In the revised manuscript, the unchanged contents are highlighted in green for clarification of the reviewer's queries and the changed contents are highlighted in yellow.**

In predictive mining it is very essential to prepare the datasets very carefully and each observation is worth full but some of advantages and disadvantage have been seen for this articles which may assist the authors to improve the quality of paper.

**Reviewer # 6 Concern 1: In abstract classification accuracy may mentioned and problem statement is missing.**

**Author Response:** Thank you for your observation. In the abstract of the revised manuscript, the problem has been stated and the classification accuracy along with the computational time have been mentioned as highlighted below:

“Lung Cancer is an uncontrolled growth of tissue causing a lump in the human lung. **If lung cancer can be detected early, it can increase the survival rate. Therefore, a multi-classification approach of lung tumor detection with high computational effectiveness is required. In this paper, a multi-classification approach of Lung tumor detection and classification is proposed using artificial intelligence on Computed Tomography (CT) scan images.** Different pre-processing steps are applied for resizing, smoothing, and enhancement of the CT images. Then, two different approaches for feature extraction using VGG16 transfer learning (TL) and morphological segmentation are proposed. Morphological segmentation and feature extraction are applied for the segmentation of the region of interest and to extract the distinct features. Finally, the proposed deep learning architecture and seven different machine learning algorithms are applied on the preprocessed data and the extracted features for the classification of lung tumors into three classes; malignant, benign and normal. **It is observed that all the ML algorithms perform with reasonably high accuracy using the low dimensional morphological features. It is also observed from the 5-fold cross-validation results that logistic regression (LR) performs with 99.36% accuracy in 23.71 sec time using the preprocessed data. Whereas, using the morphological features, both k-Nearest Neighbor (KNN) and the Support Vector Machine (SVM) show the highest accuracy of 99.76% with the lowest computational time of 0.017 sec and 0.008 sec, respectively.”**



**Reviewer # 6 Concern 2: In introduction there are heterogeneous datasets, how they are preprocessed and how the image sizes have been resized to construct the classification model because sigmoid function needs homogenous image data sizes. For example 512×512×3 uint8 dimension is converted to a binary image of 136×151 is quit dissimilar and why the tensor flow is used during the classification model construction because it is not needs in state of art methods.**

**Author Response:** Thank you for your observation.

To extract the Transfer learning (TL) features, Scaling and Normalization are applied in Python. During the preprocessing using Scaling and normalization, BGR color space images are loaded and converted to RGB format. Next, we applied scaling to transform data so that it fits within a specific scale and to change the range of data. The normalization is applied to change the observations so that they can be described as a normal distribution. We randomly shuffled the train pictures into a state of 25. Then we scale each pixel using a factor of 255. The majority of picture data has integer pixel values between 0 and 255. Small weight values are processed by neural networks, while high integer values might interfere with or slow down learning. Since, each pixel value of the image should range from 0 to 1, normalizing the pixel values is a good option. Then we used min-max scaling to normalize the data such that the feature value remains within a specific range of 0 and 1.

For morphological feature extraction in MATLAB, the morphological features are extracted using binary images. During the preprocessing using Image cropping and enhancement, the RGB images are loaded and converted to grayscale images. Then, the images are resized and cropped to get the required dimensions. Finally, the contrast level of the images is adjusted to enhance the image features such as boundaries and edges. the grayscale images are converted to binary images. That means that after completion of the Image cropping and enhancement preprocessing, the raw RGB image of 512×512×3 uint8 dimension is converted to a binary image of 136×151 logical dimension.

The data preprocessing before the feature extraction has been explained in **“3.2.1 Scaling and normalization”** on **page 4** and in **“3.2.2 Image cropping and enhancement”** on **page 5** of Section “**3 Data Preprocessing”**

TensorFlow is used in classification model construction for its versatility, strong community, and ecosystem, support for TPUs, model deployment tools, transfer learning capabilities, and customization options. Despite other frameworks, TensorFlow remains a relevant choice in research and state-of-the-art classification methods.

Reference: <https://doi.org/10.48550/arXiv.1605.08695>



**Reviewer # 6 Concern 3: In methods, Enforcement learning and other methods could be use full to construct the classification model but in most of cases the purity of data could provide more precise results and extracts the deepest knowledge of datasets in shape of prediction.**

**Author Response:** Thank you for your suggestion. We have a plan to apply enforcement learning in our future projects.



**Reviewer # 6 Concern 4: VGG16 and VGG19 rotations are used to build the variety of dataset feature why not ROIs are used to prepare the dataset because only tumor related feature are needed but why the binary threshold method is used because in deep learning these conversions are to set by default by defining the filters in architecture. Unnecessary thresholding may be omitted in deep learning.**

**Author Response:** Thank you for your observation. We have applied two types of feature extraction using VGG16-based transfer learning (TL) and morphological feature extraction. Among those VGG16 TL features are applied to the machine learning and deep learning classifiers where the binary threshold method has not been applied. Instead, the binary threshold method has been applied for ROI-based morphological feature extraction and the extracted features are applied to the machine learning classifiers only. Thresholding has not been applied to the data before applying to the deep learning classifier.

The TL and morphological feature extraction methods are explained in Section **“4 Feature Extraction”** from **pages 5 to 10**.



**Reviewer # 6 Concern 5: In results, what is difference between benign and normal because these both categories are almost same?**

**Author Response:** Thank you for your comments.

A tumor is an abnormal lump or growth of cells. If the cells in the tumor are normal, it’s benign. If they’re abnormal and grow uncontrollably, they’re cancerous cells and the tumor is malignant. If there are no tumors or cells are not considered as tumors, then it is considered as “Normal”.

The data source of the paper is the IQ-OTHNCCD lung cancer dataset [16] from Kaggle [18]. The IQ-OTHNCCD lung cancer dataset is collected at the Iraq-Oncology Teaching Hospital or National Center for Cancer Diseases (IQ-OTH/NCCD) and marked by the oncologists and radiologists. The dataset contains a total of 1190 images representing CT scan slices of 110 cases. These cases are grouped into three classes: normal, benign, and malignant, of these, 40 cases are diagnosed as malignant; 15 cases are diagnosed as benign; and 55 cases are classified as normal cases.

The data source and the detail of the data has been mentioned in **“3.1 Data source”** on **page 3 to 4** in the revised manuscript.

Reference: https://data.mendeley.com/datasets/bhmdr45bh2/1



**Reviewer # 6 Concern 6: The comparison of accuracy is missing a table of literature may be included.**

**Author Response:** Thank you for your observation. The results of the proposed methods have been compared to the existing methods in the literature in Table 9. It was observed that the proposed method outperforms the literature in terms of accuracy and computational time. Using the proposed morphological feature extraction, the proposed method requires very low computational time which is suitable for real-time applications and early detection of lung cancer.

The accuracy and computational effectiveness of the reported works to the proposed methods have also been presented and compared **in Table 9 of the revised manuscript.** Three new columns named **“Binary/Multi-class”, “Cross-validation” and “Training time”** have been added in **Table 9 to compare the accuracy and computational effectiveness**.

In the revised manuscript. a recent work of **Pandian et al [17] 2022** has been added to the literature review in the “Introduction” and also compared in Table 9. In [17], the authors use CNN and Google Net deep learning algorithms for lung cancer detection and binary classification and achieve a precision of 98% in detection and classification.

In the revised version of the manuscript, the work [17] has been discussed and highlighted in the “Introduction” as follows and also compared in Table 9 as highlighted.

**“In [17], the authors use CNN and Google Net deep learning algorithms for lung cancer detection and binary classification and achieve a precision of 98% in detection and classification.”**

In Table 9, we have updated the year of the reported works as well (as highlighted).



**Reviewer # 6 Concern 7: The paper clearly restructured and arranged by headings, introduction, literature review, methods, results and conclusion.**

**Author Response:** Thank you very much for your nice comments.

