Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 31s (2025)
https://www.jneonatalsurg.com

OPEN ACCESS

A Deep Supervision-Enhanced U-Net Model for Laryngeal Cancer Early Detection

Rashed Mohammed Alghamdi[1]

[1]Department of Laboratory Medicine, Faculty of Applied College, Al-Baha University, Saudi Arabia, PhD in infection inflammation immunity, University of Leicester UK

Email ID: rmalghamdi@bu.edu.sa ; rashed053660518@gmail.com

Orchid ID: org/0009-0000-1592-890X

Cite this paper as: Rashed Mohammed Alghamdi, (2025) A Deep Supervision-Enhanced U-Net Model for Laryngeal Cancer Early Detection, *Journal of Neonatal Surgery*, 14 (31s), 905-912

ABSTRACT

In this study, a modified U-Net architecture was used to segment the laryngeal nodule completely automatically while being closely monitored. Diagnosing laryngeal cancer is difficult since the larynx is complicated and the illness only slightly alters it. Deep learning algorithms have shown promise in medical image processing, including the diagnosis of cancer. U-Net is a well-liked deep learning architecture for picture segmentation. The outcomes are compared to those obtained using other contemporary methods and the original U-Net that was published in order to arrive at the original conclusions.

In order to investigate how this affects the segmentation of Laryngeal nodules overall, this work additionally substitutes deconvolution layers for the up-sampling layers in both networks. During training, data augmentation was used right away. Initially, distortion and rotation combinations are used because of the poor quality of the source photos. This resulted in a limited data enrichment technique for combinations. Using the same parameter settings, the network is trained by substituting deconvolutional layers for all of the upsampling layers. There is greater dice overlap when comparing the suggested method for Laryngeal nodule segmentation to the state-of-the-art.

Keywords: *U-Net Architecture, Dice Loss, Laryngeal Cancer, Stochastic Learning, Segmentation, Training rate and Validation*

1. INTRODUCTION

Laryngeal carcinoma has grown to be a serious health concern in recent years. It is believed that the most severe kind of the illness occurs when it develops in the head and neck [1]. The treatment options and prognosis for a patient with laryngeal cancer are determined by the stage of the malignancy [2]. When laryngeal cancer and pre-malignant disorders are detected early, it is possible to provide a higher level of protection for the larynx and to make a clear diagnosis in real time. Patient survival rates are greatly increased by early discovery of laryngeal cancer because prompt interventions can stop the disease from progressing to more advanced stages and lessen the need for drastic treatments that affect speech and swallowing abilities.

On the other hand, advanced laryngeal cancer requires a multimodal diagnostic approach, which may result in a worse quality of life with serious negative effects [3].The confusing benign lesions, subtle morphological changes at early stages, and interobserver heterogeneity among clinicians in interpreting endoscopic images are due to the difficulty in laryngeal cancer detection. It has recently been discovered that there is a higher frequency of a more effective diagnostic method with a survival rate between 34% and 62% [4]. In general, early and accurate diagnosis of the disease is always critical for getting the best medical outcome [5].

In CNN, the diagnoses can be done much faster when compared to the human medical expert's capability for diagnosis [6, 7]. The CNN, in general, is a feed forward NN with a complex topology and can perform the operation of convolution. A technique can be applied on problems that are connected with classification and detection. In particular, as compared to classical methods of image processing, this approach has high potential in features extraction and further analysis [8]. It follows that the performance of the U-Net architecture proposed in [9] has performed very well for medical images. Ever since then, many works have been conducted to deal with different types of U-Net-based segmentation.U-Net models have been modified to incorporate deep supervision, multi-resolution feature extraction, and attention processes in order to further improve segmentation accuracy. This has increased the models' resilience to complicated anatomical variances.

Additionally, Dice coefficient [10, 11] has supplanted pixel-wise cross entropy as the industry standard for several optimisation techniques for medical image segmentation [10, 11]

2. LITERATURE REVIEW

Using CT scans and real-world datasets, a more sophisticated deep learning (DL)-based Mask R-CNN model was presented for the detection of laryngeal cancer and its associated symptoms [12]. By accelerating cancer diagnosis and lowering diagnostic uncertainty, this strategy seeks to improve diagnostic accuracy.In principle, the proposed approach would allow for more frequent monitoring of patients by speeding up and improving the identification of small cancers in the larynx. The authors [13] develop a DL approach for in-vivo hyperspectral LC diagnosis. Using a Deep Quantum Neural Network (Deep QNN) tuned by the Adaptive Spotted Hyena Optimizer (ASHO), [14] presented a unique laryngeal cancer classification method. By adjusting to the distinct features of laryngeal cancer images, our approach dynamically improves classification performance by fine-tuning network classifiers. Deep QNN may employ the mechanism that was explained before to change network classifiers in order to classify as a final step. Convolutional neural networks (CNN) and attention mechanisms are combined in the LPCANet network to improve feature extraction and classification accuracy in the diagnosis of laryngeal cancer [15]. This approach efficiently captures both spatial and spectral features by patching the original hyperspectral images and using ResNet50 for local feature extraction, enhancing diagnostic accuracy. First, the original HSI is divided into segments. The patches are fed to ResNet50 to extract local features. Meanwhile, channel attention and location attention modules are built at the same time in order to take channel and spatial interaction into account.

Both techniques produce the fusion feature maps which are helpful for representation of features and classification networks. Authors in [16] presented a DCNN model for the auto-classification of CE-NBI images as benign or malignant with limited human intervention. A DCNN was developed in this domain by adopting an approach of the cut-off layer with pre-trained ResNet50 architecture. The authors in [17] have proposed a CNN model that classifies the input spoken audio with the help of image classification. MFCC approach takes the input from the DNN model. Authors in [18] have discussed CNN model for glottic cancer detection and CNN model performance has been compared with DT ensemble learning, as one of the methods for smaller datasets considering the performance based on the classification accuracy.

The U-Net model, which was originally proposed , is the CNN architecture normally used for segmenting biomedical images. It uses an encoder-decoder architecture together with skip connections to improve segmentation accuracy and keep the spatial information intact. It does this encoding to reduce image size and enable easy feature extraction. The image feature map has to be decoded to increase its size to obtain a segmentation mask. A characteristic feature of U-Nets is the skip link that connects both encoder and decoder. It's this very capability that enables the model to improve segmentation accuracy while still maintaining high resolution.

In the last years, several possible applications have been investigated for the U-Net, including medical image analysis [19-25]. Segmentation of a nodule is performed with a modified U-Net architecture with deep supervision for a fully automated technique in this research work. Deconvolution layers are later ped out for the upsampling layers in both networks to observe how this changes the overall performance in Laryngeal nodule segmentation.

3. METHOD

U-Net-based methods are widely used for laryngeal nodule segmentation largely because of the network's inherent mixing of many scales across it. In this work, Laryngeal Nodule Segmentation (LNS) is achieved using the modified U-Net technique under deep supervision.

3.1 The Original U-Net Model

The original U-Net design is made up of convolutional layers, upsampling layers, and max-pooling layers. With a starting learning rate of $lrate_{initial} = 5 * 10^4$, a learning rate schedule of l^2 with weight decline of 10^4 , and $lrate_{initial} = 0.934$ epoch.

During training, data supplementation was used on the spot. The initial combinations like, deformation, and rotation are used but it was realised that these techniques actually made the outcomes worse because of the poor quality of the input images. As a result, the combinations are chosen as a method of data enrichment. Using the same settings as above, the network is trained by swapping out all of the upsampling layers for deconvolutional layers.

3.2 Deep Supervision based Modified U-Net Model

The encoding and decoding paths are present in this design, just like in the initial U-Net. The input forms are progressively stored along the encoding route as the network depth increases.

The deciphering route, in comparison, combines the acquired models with surface characteristics again, leading to accurate location of the interest structures. For ease of use, the term "encoding module" refers to all of the working units in the encoding paths. The term "decoding module" is also used to describe all of the working units in the decoding paths.

A residual block is represented by each encoding module, which is made up of four convolution layers and a dropout layer ($pdrop = 0.3$). They are joined together by step 2 convolutions. Similar to upscaling, the deciphering route starts with step 2, followed by a convolution operation. The feature maps are then subsequently combined with the encoding route and then

transmitted to the decoding route is cut in half by the up-sampling module.

A 2*2 convolution and a 1*1 convolution make up a deciphering module, which bifurcates the feature map to half. The segmentation layers is incorporated at various network levels in the deciphering route to implement deep control.

In this work, six-fold cross-validation algorithm has been used to train and test the network on the training dataset. The leaking ReLu, selected patches of size $256 * 256$, batch sizes of 6, validation batch sizes of 12, $padding = same$, instance normalisation rather than batch normalization, and no validation patch overlap are used to train the network design. The Adam optimizer has been used for training, with the initial learning rate set to $5 * 10^4$, the learning rate schedule set to 10^4 with l^2 weight decay, and the initial learning rate has been set to 0.934 epochs.

The total loss is summed across all instances in the group of size M because the optimisation is stochastic. Formally, the N-voxel image's expected and real-world voxels j in a group of size M should be $p_{j,i} \in P_j$ and $g_{j,i} \in G_j$, where $\forall p, g;\ p, g \in [0, 1]$. The definition of the stochastic dice coefficient D is as follows:

$$D = \frac{1}{M}\sum_j^M \frac{2\sum_i^N p_{j,i}g_{j,i}}{\sum_i^N p_{j,i}+\sum_i^N g_{j,i}} \quad (1)$$

$$D_{multiclass} = \frac{1}{M}\sum_j^M \frac{1}{|K|}\sum_{k\in K} \frac{2\sum_i^N p_{j,i}g_{j,i}}{\sum_i^N p_{j,i}+\sum_i^N g_{j,i}} \quad (2)$$

The $D_{multiclass}$ denotes dice coefficient for the situation of segmentation considering multiclasses (more than one labels separate from the information of background). Here, $k \in K$ are the classes. When the target-label amount is plentiful, the dice coefficient's shape remains continuous; however, the target-label quantity gets closer to zero, it becomes pointed and flat. Zero loss results in no slopes for the model to change its parameters and traps it in local optimum, making it challenging to learn the data. Additionally, the problem is not resolved by the use of uniform dice loss. To lessen the effects of the issue, a random combined die variable is suggested.

3.3 Stochastic Aggregated Dice Loss (SADL)

This work used Stochastic Aggregated Dice Loss (SADL), which deviates from the conventional approach by calculating per-image dice values and aggregating them for each group. In this method, the dice loss is calculated using the combined image after merging all segmentation results from a group into a single big image. Consider $M, P,$ and G stand for the group size, projected nodule area, and ground truth nodule area, respectively. Let $p, g \in [0, 1] \forall p, g$ be the anticipated and actual image voxels j in batch M, respectively, and let $g_{i,j} \in G_j$be the ground truth voxels. Equation (4) defines the suggested loss function D_{SA} as follows:

$$D_{SA} = \frac{2\sum_j^M \sum_i^N p_{j,i}g_{j,i}}{\sum_i^M \sum_i^N p_{j,i}+\sum_i^M \sum_i^N g_{j,i}} \quad (3)$$

$$D_{SA_{multiclass}} = \frac{1}{|K|}\sum_j^M \frac{2\sum_j^M \sum_i^N p_{j,i}g_{j,i}}{\sum_j^M \sum_i^N p_{j,i}+\sum_j^M \sum_i^N g_{j,i}} \quad (4)$$

The SADL dice coefficient is specified in equation (4) for the situation of multiclass segmentation, with $k \in K$ denoting the classes.

This strategy is justified by the likelihood that other examples with labels will fit the no-target examples during training. As a result of the labels from such images, the combined die coefficient shape will become softer and curved.

3.4 Stochastic Multi-resolution Loss Component (SMLC)

In current U-net systems if we consider high-resolution ground facts, there is a training loss at the top-level layer's segmentation output. The SMLC which calculates loss components at various resolutions. The segmentation result of the individual level and its corresponding downsampled ground-truths are used to compute the per-resolution component. Assume, for instance, that the result at level 2 is $64 * 64$, the resolution of the ground truth is down-sampled $64 * 64$ and the level 2 computes the die coefficient component. The weighted aggregate of these loss components represents the model's overall loss.

Let $P_k, G_k, and\ k$ represent the expected and actual image as well as the at level per-resolution loss coefficient (resolution) k, respectively. The chosen dice coefficient loss function is $Loss(P, Q)$. This can be another loss function or one of formulae (1) or (3). Equation (5) defines the overall multi-resolution loss as follows:

$$D_{multi-res} = \sum_k \alpha_k\ Loss(P_k, G_k) \quad (5)$$

The variable k stays a hyperparameter or trainable parameter in this study, but α_1 is maintained fixed at unity. Empirically, it was advantageous to use trainable parameters or decreasing values for lesser levels. The first, second, and third values of α are specified and performance gains are also produced by using trainable α_k.

Results and Discussion

The dataset has been taken from Kaggle online database for histopathological images [https://www.kaggle.com/ashenafifasilkebede/dataset?select=val] and this work used 1224 images. The images are scaled down to 256*256 resolution. They are divided into a training set, testing set and a validation set using ratios of 80% (979 images), 10% (122 images) and 10% (123 images), respectively. This split of dataset has been shown in Table 1. The numerical results have been presented in Table 2. The averaged cross-validation performance is calculated across all trials after doing numerous rounds of cross-validation. Figure 1 displays the best validation performance. The average dice coefficient for the segmentation job throughout the whole dataset for the proposed deep supervision approach for Laryngeal nodule segmentation was 0.912.

Additionally, the model was able to identify between several malignant Laryngeal nodule subtypes with an AUC of 0.92. Figure 2 displays a Receiver Operating Characteristic (ROC) curve for the training, testing, and validation sets, which is a graphical depiction of a classification model's performance at all threshold levels for Class 1 (Normal) and Class 2 (Cancerous). AUC measures the capability of the model to distinguish between classes using the area under the ROC curve.

The high AUC value, 0.92, achieved by the suggested model, ensures the ability of this model to effectively distinguish between benign and malignant laryngeal nodules, which is of prime importance for early-stage treatments. Conventional radiological reviews typically reach an AUC range of only 0.80–0.85, emphasizing the advantages of automated deep-learning-based segmentation methods. Moreover, integration of such models into the clinical workflow can lead to better consistency in decision-making by radiologists and oncologists and can reduce diagnostic variability .

The segmentation algorithm's ability to match the ground truth data accurately is gauged by the mean squared error. The histogram displays the proportion of photos where a certain mean squared error was created by the segmentation method. This might be used to compare various algorithms and would provide a broad indicator of how accurate the algorithm is.

With accuracy values ranging from 85% to 90%, prior research employing CNN-based segmentation models for laryngeal cancer diagnosis, such as Mask R-CNN and Deep Quantum Neural Networks, has shown encouraging results. However, because these techniques rely on specified feature hierarchies, they frequently have higher false-positive rates. By dynamically adjusting segmentation parameters according to contextual picture features, SADL's incorporation into our model helps to reduce these problems .

The Laryngeal nodule segmentation of experimental findings is shown in Table 3. As can be observed, the UNet model with the SADL gets a dice score of 0.9, which is 0.025 die points higher than the dice score of 0.875 from the original U-Net architecture.

The quantity of false-positives is decreased because of SADL enhancement of the gradient signatures of samples with SMLC. Due to the way it dynamically organises various combinations of training images, it also functions as an augmentation approach. Additionally, SMLC provides enhancements over the standard method. Combining this strategy with SADL in a dice score of up to 0.915, which is 0.04 points greater than the starting position. The findings imply that employing decreasing is essential for the improvement in performance.

Table 4 presents the experimental findings for the chosen dataset, employing alone SADL which does not result in any appreciable performance increase, however adding SMLC leads in modest gains. The multi-resolution losses with declining factors results in a performance of 0.935 dice for the overall tumor, 0.769 dice for the core nodule, and 0.723 dice for the improved nodule segmentation benchmark. This demonstrates 0.015, 0.02, and 0.022 performance improvements, respectively, above the original U-Net architecture .

4. CONCLUSION

The proposed method for Laryngeal nodule segmentation shows more dice overlap as compared to the existing methods. There are a few issues that must be resolved for LNS. The fundamental drawback of the proposed method is that most cutting-edge techniques employ whole Laryngeal histopathological images, but trained on previously chopped images inside the Laryngeal nodule ROI due to a lack of diverse datasets. The remainder of the approaches also had to deal with Laryngeal nodule ROI identification in addition to segmentation, thus comparing them to the state-of-the-art may not be entirely accurate. As a result, training with the dataset was substantially simpler. A new U-Net model with deep supervision is developed for early laryngeal cancer detection due to classic U-Net design flaws. The model uses deep supervision, which adds supervision signals to training. By making the model more durable and discriminative, performance improves. Research is needed to evaluate the method's therapeutic efficacy.

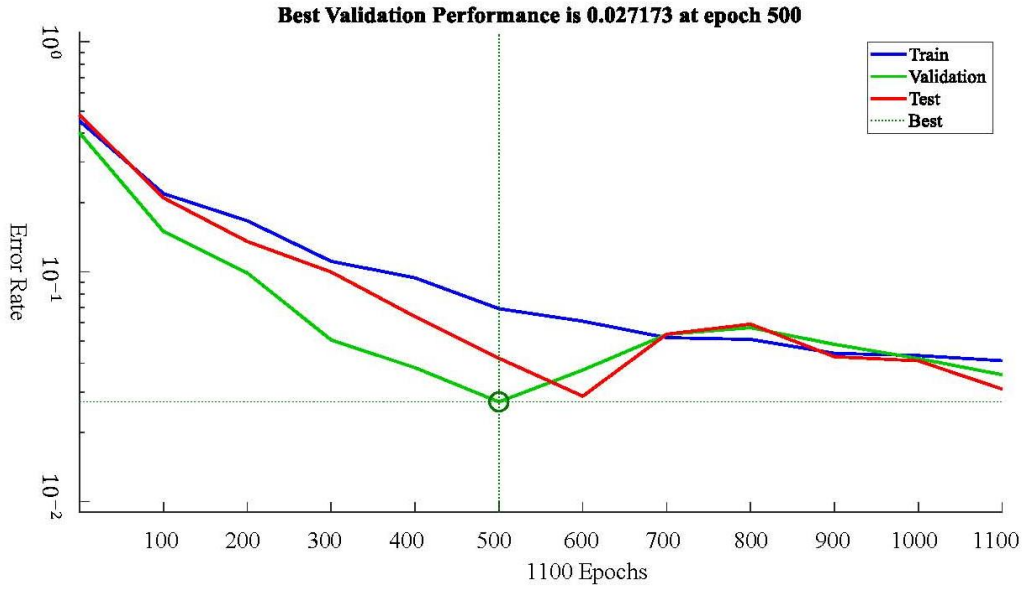

Figure1: Performance graph for training, testing, and validation with respect to the suggested method's error rate

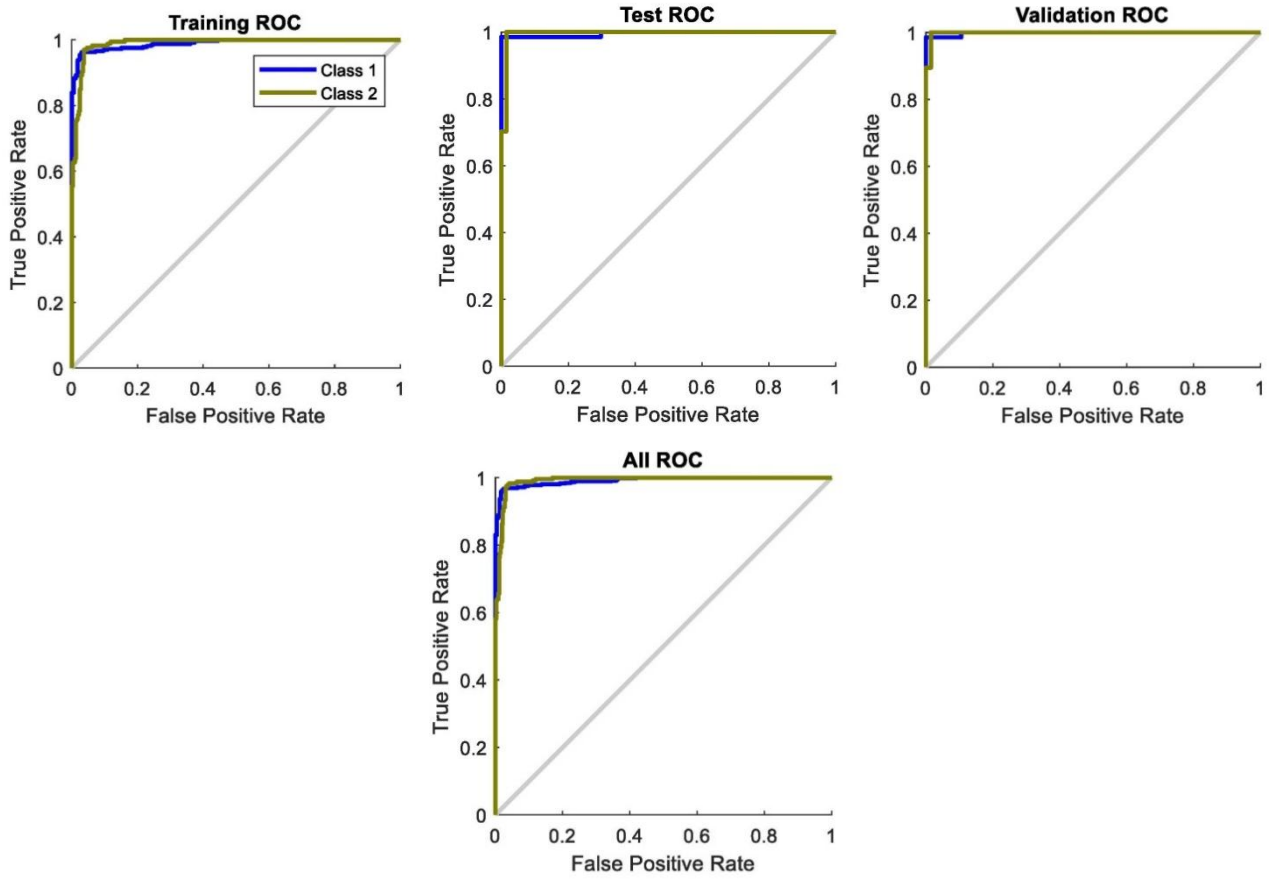

Figure 2: ROC curves

Table 1: Number of images for training, test and validation

Method	Training Set		Test set		Validation set		
	%	Image count	%	Image count	%	Image count	
$SADL$	80	979	10	122	122	123	
$SADL + SMLC$	80	979	10	122	10	123	

Table 2: Numerical results of proposed method

	Accuracy		**Precision**		**Sensitivity**		**Specificity**	
	Test	Val.	Test	Val.	Test	Val.	Test	Val.
$SADL$	97.24	97.44	97.22	96.88	98.74	97.65	97.25	97.54
$SADL + SMLC$	98.06	98.66	98.55	95.85	99.51	98.35	99.21	98.34

Table 3: Cross Validated results of Laryngeal nodule Segmentation task

Model	**SMLC (α)**	**Dice**
U-net (Original) [4]	-	0.873
SADL	-	0.910
SADL +SMLC	$\alpha = (1,1,1)$	0.918
SADL +SMLC	$\alpha = (1,0.1,0.01)$	**0.916**
SADL +SMLC	Trainable α	0.927

Table 4: Cross validated results for Dataset segmentation task

		Dice Loss	
Model	**SMLC (α)**	**Whole area**	**Enhanced**
U-net (Original) [4]	-	0.892	0.746
SADL	-	0.854	0.776
SADL	$\alpha = (1,1,1)$	0.923	0.723
SADL +SMLC	$\alpha = (1,0.1,0.01)$	0.918	0.734
SADL +SMLC	Trainable α	0.915	0.756

Funding Statement

This work was supported and funded by the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU) (grant number IMSIU-DDRSP2501).

Acknowledgement

This work was supported and funded by the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU) (grant number IMSIU-DDRSP2501)

REFERENCES

[1] Johnson DE, Burtness B, Leemans CR, Lui VW, Bauman JE, Grandis JR. Head and neck squamous cell carcinoma. Nature reviews Disease primers. 2020 Nov 26;6(1):92.

[2] Hoffman HT, Porter K, Karnell LH, Cooper JS, Weber RS, Langer CJ, Ang KK, Gay G, Stewart A, Robinson RA. Laryngeal cancer in the United States: changes in demographics, patterns of care, and survival. The Laryngoscope. 2006 Sep;116(S111):1-3.

[3] Forastiere AA, Ismaila N, Lewin JS, Nathan CA, Adelstein DJ, Eisbruch A, Fass G, Fisher SG, Laurie SA, Le QT, O'Malley B. Use of larynx-preservation strategies in the treatment of laryngeal cancer: American Society of Clinical Oncology clinical practice guideline update. Journal of Clinical Oncology. 2018 Apr 10;36(11):1143-69.

[4] Miller KD, Siegel RL, Lin CC, Mariotto AB, Kramer JL, Rowland JH, Stein KD, Alteri R, Jemal A. Cancer treatment and survivorship statistics, 2016. CA: a cancer journal for clinicians. 2016 Jul;66(4):271-89.

[5] Swets JA. Measuring the accuracy of diagnostic systems. Science. 1988 Jun 3;240(4857):1285-93.

[6] Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nature medicine. 2019 Jan;25(1):44-56.

[7] Kermany DS, Goldbaum M, Cai W, Valentim CC, Liang H, Baxter SL, McKeown A, Yang G, Wu X, Yan F, Dong J. Identifying medical diagnoses and treatable diseases by image-based deep learning. cell. 2018 Feb 22;172(5):1122-31.

[8] Lu D, Weng Q. A survey of image classification methods and techniques for improving classification performance. International journal of Remote sensing. 2007 Mar 1;28(5):823-70.

[9] Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J. Unet++: A nested u-net architecture for medical image segmentation. InDeep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4 2018 (pp. 3-11). Springer International Publishing.

[10] Hasan MM, Islam MU, Sadeq MJ, Fung WK, Uddin J. Review on the evaluation and development of artificial intelligence for COVID-19 containment. Sensors. 2023 Jan 3;23(1):527.

[11] Abdusalomov AB, Mukhiddinov M, Whangbo TK. Brain tumor detection based on deep learning approaches and magnetic resonance imaging. Cancers. 2023 Aug 18;15(16):4172.

[12] Dou Q, So TY, Jiang M, Liu Q, Vardhanabhuti V, Kaissis G, Li Z, Si W, Lee HH, Yu K, Feng Z. Federated deep learning for detecting COVID-19 lung abnormalities in CT: a privacy-preserving multinational validation study. NPJ digital medicine. 2021 Mar 29;4(1):60.

[13] Akbari H, Halig LV, Schuster DM, Osunkoya A, Master V, Nieh PT, Chen GZ, Fei B. Hyperspectral imaging and quantitative analysis for prostate cancer detection. Journal of biomedical optics. 2012 Jul 1;17(7):076005-.

[14] Li C, Chen H, Li X, Xu N, Hu Z, Xue D, Qi S, Ma H, Zhang L, Sun H. A review for cervical histopathology image analysis using machine vision approaches. Artificial Intelligence Review. 2020 Oct;53:4821-62.

[15] Sahoo PK, Mishra S, Panigrahi R, Bhoi AK, Barsocchi P. An improvised deep-learning-based mask R-CNN model for laryngeal cancer detection using CT images. Sensors. 2022 Nov 15;22(22):8834.

[16] Esmaeili N. *Computer-Aided-Diagnosis for Laryngeal Lesion Assessment: A Feature Extraction and Machine Learning Approach Applied on Enhanced Contact Endoscopy Images* (Doctoral dissertation, Technische Universität München).

[17] Hershey S, Chaudhuri S, Ellis DP, Gemmeke JF, Jansen A, Moore RC, Plakal M, Platt D, Saurous RA, Seybold B, Slaney M. CNN architectures for large-scale audio classification. In2017 ieee international conference on acoustics, speech and signal processing (icassp) 2017 Mar 5 (pp. 131-135). IEEE.

[18] Han Y, Ma Y, Wu Z, Zhang F, Zheng D, Liu X, Tao L, Liang Z, Yang Z, Li X, Huang J. Histologic subtype classification of non-small cell lung cancer using PET/CT images. European journal of nuclear medicine and

molecular imaging. 2021 Feb;48:350-60.

[19] Ibtehaz N, Rahman MS. MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation. Neural networks. 2020 Jan 1;121:74-87.

[20] Kang DH, Cha YJ. Efficient attention-based deep encoder and decoder for automatic crack segmentation. Structural Health Monitoring. 2022 Sep;21(5):2190-205.

[21] Nixon M, Aguado A. Feature extraction and image processing for computer vision. Academic press; 2019 Nov 17.

[22] Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence. 2017 Jan 2;39(12):2481-95.

[23] Zhou Z, Siddiquee MM, Tajbakhsh N, Liang J. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE transactions on medical imaging. 2019 Dec 13;39(6):1856-67.

[24] Litjens G, Kooi T, Bejnordi BE, Setio AA, Ciompi F, Ghafoorian M, Van Der Laak JA, Van Ginneken B, Sánchez CI. A survey on deep learning in medical image analysis. Medical image analysis. 2017 Dec 1;42:60-88.

[25] Ker J, Wang L, Rao J, Lim T. Deep learning applications in medical image analysis. Ieee Access. 2017 Dec 29;6:9375-89