International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Euri – A Deep Ensemble Architecture For Oral Lesion Segmentation And Detection

Sagari S. M.*[1], Vindhya P. Malagi[2], Smita Sasi[3]

Submitted: 22/08/2023 **Revised**: 07/10/2023 **Accepted**: 22/10/2023

Abstract: Oral cancer is a dreadful diseases across the globe and the sixth most cancer types ranked with high rates of mortality and morbidity. The proposed study employs a cost-effective approach using digital images that apply deep learning architectures to classify the images using segmentation techniques. The study proposes a EURI -Ensemble of Resent and Inception as a backbone on the Unet model to classify the images as Cancer. The current work consists of total of 285 Images, where 233 are cancer and 52 arenon-cancer. The EURI model encompasses two variants of Resents - Resnet-34 and Resnet-101 and Inception V3 are ensembled as backbone on Unet. Thus, the classifier models arecontemplated as feature extractors for the Unet. Weighted averaging is carried out on the prediction of each individual model. The model outperformed with an Intersection over Union (IOU) score of 94%.

Keywords: *ResNet, UNet, Inception V3, Weighted Averaging*

1. Introduction

The most lethal cancer form, oral cancer, is responsible for high mortality and morbidity rates over the globe. Neglecting dental health can lead to the sixth most prevalent type of cancer. Oral disorders that can lead to oral cancer are a frequent problem in several fields related to the mouth. If leftuntreated, the condition has a high risk of developing several additional consequences. Uncontrollable cell splitting starts to invade nearby healthy tissues and eventually causes metastasis. In order to stop the disease from progressing further, early detection is essential.

Various methods are working to effectively detect and diagnose. The former technique, histopathological biopsy, an invasive approach, considered as a gold standard used to locate the oral lesion by extracting the cell from the suspicious regions [1]. However,improvements in non-invasive medical imaging techniques have improved the early detection and diagnosis of anomaliesassociated with oral cancer. Several studies are considered to perceive a method [2]. Implementing a cost-effective detection and diagnostic technique assumes a crucial action in the present-day medicine, when taking socioeconomic demographics in low and middle-income nations into consideration. Modern advancements in photographic imaging technologies provide a practical solution in the sh apeof readily available non-invasive technologies. As seen in [3], several researchers are moving in this direction.

For the region to be classified as suspicious or non-suspicious, the clinical appearance of lesions is important [7]. The brain recognizes the patterns of features using neurons, builds layers to represent them, and then encodes and decodes. those using patterns. Deep learning automates pattern recognition by using neurons. The layers are instructed and made to accurately understand the feature patterns. Computer vision and deep learning techniques for Image recognition using classical Image processing techniques, to extract features with elevated accuracy is an effective approach in medical research [8][9].Advancement in deep neural networks [3][4] in both invasive and non-invasive modalities[10] lead to the usage of conventional neural networks extensively in recent years, to localise and categorise the region of Interest in an accurate way [5][12]. A systematic and structured representation of Deep learning models provides an effective approach in classification and detection of oral cancer [6]. For image processing applications, deep neural models like ResNets, VGGs, InceptionVs, UNets, and Dense Nets are frequently used [11]. With an F1 score of 41.18%, Resnets categorizes the oral regions as normal and abnormal regions [6].

The present paper discusses four deep architectures Resnet- 34, Resnet-101, InceptionV3, and Unet to combine classification and segmentation techniques implemented on the current dataset. The datasets include 233 oral cancer Images and 52 oral non-cancer images to perform the task of localization, segmentation, and classification. Our datasetcontains photographic images of several regions of the oral cavity lips, tooth gums, tongue, and hard palate, soft palate irrespective of image size, illumination, and from different sources. To segment, the region of Interest Unet isconsidered. For

[1] Dayananda Sagar College of Engineering Bengaluru,India
[2] *Dayananda Sagar College of Engineering Bengaluru,India*
[3] *Dayananda Sagar College of Engineering Bengaluru,India*
** Corresponding Author Email: sagarimariyappa@gmail.com*

classification, Resnet-34, Resnet-101, and InceptionV3, are considered. The comparative analysis in the form of accuracy of each individual model is noted.

Considering accuracy, Intersection Over Union (IOU) score, and dice coefficient as performance measure metrics, further work focused on U-net considered as a backbone over the classification models to improve the accuracy score on these individual models. As extension work, the four deep nets are ensemble by performing weighted averaging to improve accuracy.

The following sections discuss the proposed methodology,which includes a description of the datasets, deep architectures of the individual classification models as a backbone, Unet and ensemble models (EURI models).

2. Proposed Methods

The proposed models include pretrained models ResNet andInception as a backbone to the Unet. The models are pre trained on ImageNet dataset to classify the region of interest in an accurate way. The pretrained models are fine-tuned by adjusting the hyperparameters including learning rates and the number of epochs.

2.1. Datasets

The study includes 285 oral cavity images of both suspicious and non-suspicious regions. The present datasets include suspicious lesions that are well blended with the non- suspicious region, however, the localization of ROI in this stage becomes challenging.The collection of the datasets is from web search engines, and repositories like Kaggle and from a medical practitioner in specialization of Oral Oncology. The Image datasets consist of various kinds of lesions in the oral cavity includes buccal mucosa, tongue, lips,and hard palate, soft palates, and tooth gums. Images are unstructured based on their varied size, varied resolution, andimage capture methods. The collection is annotated by clinical experts as cancer (malignant) and non-cancer (non- malignant) based on their visual clinical appearance. In Table 1

The datasets are divided into 165 as training sets, 60as validation, and 60 as test sets. Table 2 depicts the clinical names of the lesions with their respective Images.

Table 1. Datasets used for the proposed method

# Datasets	# Cancer	# Non-Cancer	# Training	# Testing	# Validation
285	233	52	165	60	60

Table 2. Clinical names of the lesions with their respective Images

Images	Name of the lesion	Category of the lesion
	Carcinoma left buccal mucosa	Malignant
	Ulcer	Non- Malignant
	Carcinoma tongue right lateral border	Malignant
	Carcinoma right lower alveolus	Malignant

2.2. Deep Architectures

Convolution Neural Networks in neural networks, analyze the images as a matrices and a core for various deep models, perform the action by activating the neuron and learning Image pixels efficiently. During back propagation, however, the model leads to overfitting problems. Nowadays, various state-of- the-art deep net architectures are in stride to classification and segmentation tasks, where each differs through its layer structure, architecture, and depth.

The proposed work focuses on the various pre-trained deep models for good precision for the current image datasets, which include ResNet, Inception, UNet, DeepNet, and EfficientNet Models, as each of them with differentfunctionalities within, among which few are rich in extractinglow level and high-level features in predicting the region of Interest. However, Resnet and Inception models outperformed with good accuracy compared to plain U-Net, Deep Net, and Efficient net. Considering these models as bestsuited to predict the present dataset accurately as malignant and benign or cancer and non-cancer.

After extensive testing Resnet 101 and Resnet 34, the twovariations are contemplated in the study of Resnets, are thought of as learning models in deep neural networks that categorize theregions as malignant or benign. The ResNets include residual blocks which are called so by their important characteristic known as skip connection [17]. It regularizes the effect of thenetwork, so improves the efficacy of network. ResNets of the current study include 3x3 max poolingand a pair of dense layers with activation functions Relu and sigmoid implemented. Inception architectures evolved byimplementing the reiterations of the kernel process, to progress in the concepts of traditional CNNetworks. Considering the Inception architecture on the Image dataset of oral lesion plays a vital role, as Images in current study are varied in size and resolution [18] to Fpredict and classify theregions as normal and abnormal.

The DeepNetworks are wiider, includes overlapping max pooling with various convolution filters, pair of dense layers and one flatten layer are implemented to classify the image as depicted in ROC cure. Thus, reduces computation costs withan increase in accuracy of about 85% for 233 cancer imagesand 52 non-cancer images.

2.3. Deep models as a backbone of U-net

The proposed work discusses Unet implemented over presentdatasets. Based on the best-proven classifier model ResNet and Inception are considered as a backbone to the UNet as a feature extractor and Unet to segment the oral cavity lesions into suspicious or non-suspicious regions in an effective way.Resnet outperformed with good accuracy Inception with moderate accuracy and U-net with lower accuracy on the present datasets, however, these are pre-trained models with pre-trained weights considering them as encoders in U-net architecture that improves training accuracy.

Image classifications are important in predicting and classifying the regions by extracting the features of interest into two states of the regions as normal and abnormal, the semantic segmentation, the variation of classification in U- net becomes important to localize the group of pixels into the same category of feature and segment into one single mask. However, U-net functions by discarding the dense layer, which contains only convolution layers. Thus, decreasing the length of architecture.The architecture of the U-net is a U-shaped encoder and decoders [16] include only two layers: convolution and max pooling layer, whereas the latter includes deconvolution level, by condensing the image size and as well information and hence increasing the channel size. Skip connections are included, where concatenations of outputs of encoders and decoders are performed between the same levels to maintainthe same input and output image size. U-net also locates a site.

The study includes a total of 285 images belonging to both suspicious and non-suspicious regions, splitted into 165 for training sets, 60 for validation, and 60 for test sets. The datasets are annotated by a clinician in accordance to clinical features. Fig1, Fig 2, and Fig 3 show the layer visualization of the deep models implemented on the currentdatasets. Fig 4 shows layer visualization of non-cancer in deep architecture.

The present datasets include suspicious lesions effectively blended with the non-suspicious region, however, the localization and classification, of current study, are challenging. The architecture pixel-wise segmentation is best suited for thepresent work.Input Images are resized to 224 x 224 and fed into a convolutional layer consisting of 64 filters, max-pooling layers, and batch normalization. The images are down sampled with max pooling layers by enhancing the number of filters, thus reducing image size.

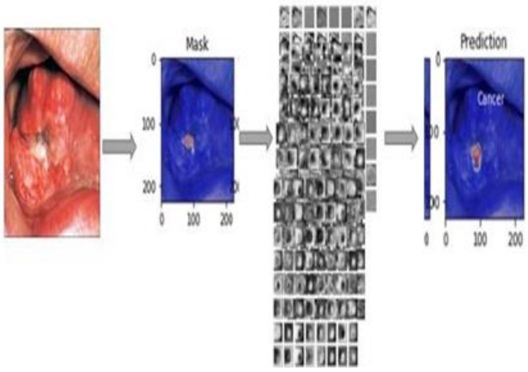

Fig 1. Layer Visualization of Resnet-34 deep architecture

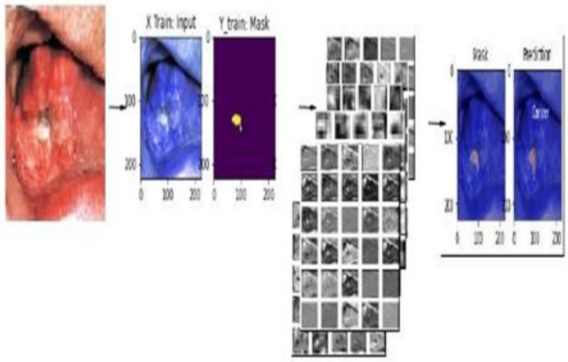

Fig 2. Layer Visualization of InceptionV3 deep architecture

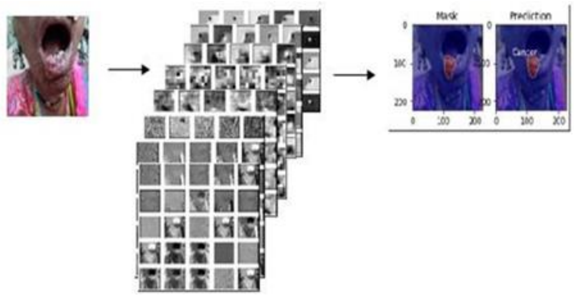

Fig 3. Layer Visualization of Resnet101 deep architecture

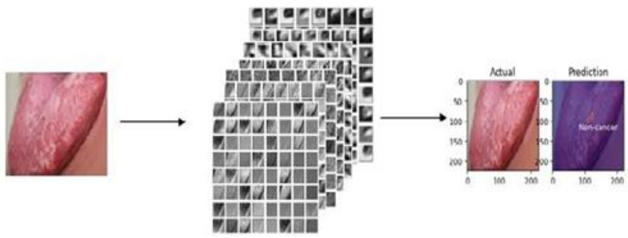

Fig 4. Layer Visualization of Non-cancer in deep architecture.

2.3. Ensemble models

Perception of the features through their layer visualization iscarried out in all deep architectures that are considered in thecurrent work. Visualization of feature extraction in the layers increases the visibility and interpretability of convolution filters in all layers of the deep architectures. Feature representation varies from layer to layer, identifying patternswith the different deep nets shown in Fig 5 signifies the flowchart of the EURI model.

The current work focuses on the extraction of complex lesion features, where suspicious lesions are submerged with the normal region, hence classifying them as pixel wise play a significant task for better classification, and for better detection of the area of interest. The EURI Model signifies the Ensemble of Pretrained Models ResNets, Inception on Unet. The predictions of every single model used in the current study are added together and performed a averaging on their corresponding weights on each prediction. The weights are fine-tuned to elevate the precision of the current model. The weights are selected using grid search algorithm, the common approach for optimizing hyperparameters in deep learning models. The pretrained models are ensembled based on the classification rate of each model depicted in Table 4. The EURI algorithm outperformed with accuracy of 94%

The current study proposes the EURI (Ensemble of Resnet and Inception on Unet) model, which includes three deep learning approaches Inception v3 and Resnet 34, and Resnet 101 are ensembled on the Unet as a backbone classifier models to segment and classify the suspicious regions with improved accuracy. The current datasets includelow-level and high features; therefore, the improved model extracts low-level and high-level features in an efficient way. The Images are preprocessed based on pretrained models used as a backbone by resizing the images to 224x224.

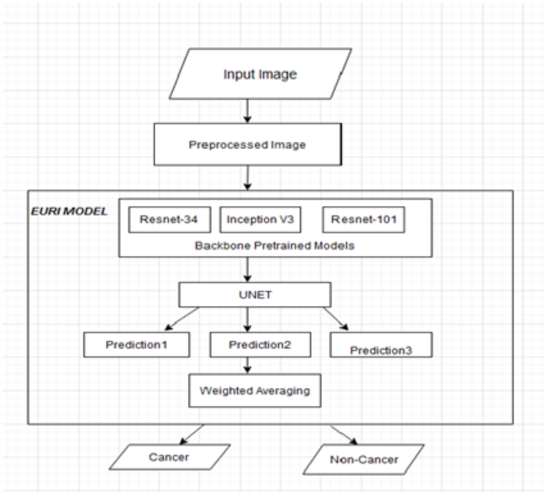

Fig 5. Flow diagram of the EURI model

3. Experimental Results

Performance evaluation of the model into cancer or non-cancer measures carried by the performance metric IOU (Intersection over Union) Score, F1 score validation loss, andloss are the metrics measured for the proposed model performance as in Table 3. Training loss and IOU comparison of Resnet-34, Resnet-101, and Inception V3 are shown in Fig 5 and Fig 6 to compare individual models based on the segmentation Dice coefficients are used to measure the performance of each individual model shown in Fig 7. Table 4 depicts the IOU score of the different ensemble deep models tested on the current dataset. The study implements the grid search algorithm for the weighted averaging values[19] for the weights for better results.

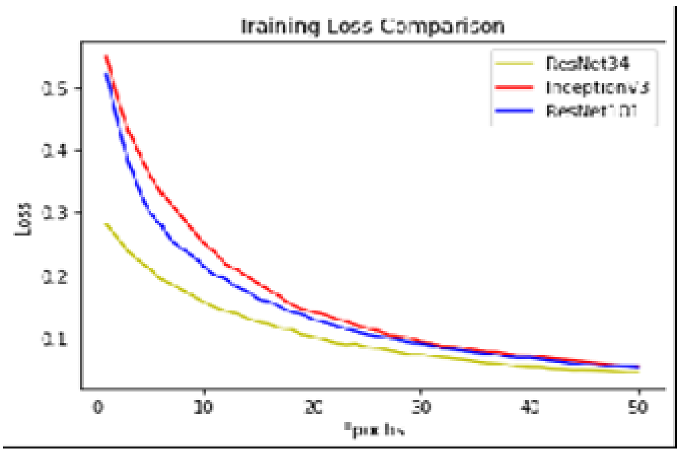

Fig 6. Training Loss Comparison of each individual model

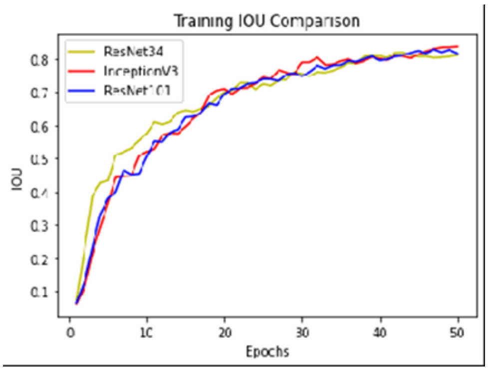

Fig 7. Intersection Over Union (IOU) Comparison of each individual model.

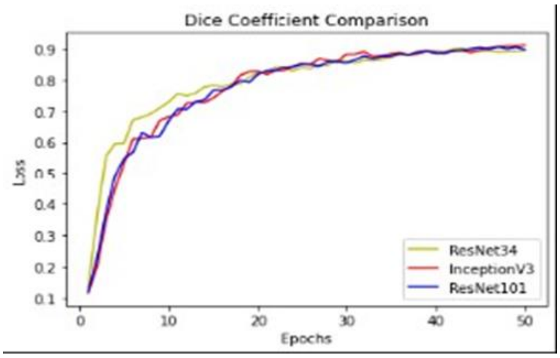

Figure 8. Dice coefficient of each individual model

Table 3. Performance evaluation of deep nets on the U-net backbone without varying weights model

Deep Models	IOU Score	Val IOU Score	F1 Score	Val F1 Score	Loss	Val Loss
RESNET-34	0.62	0.40	0.76	0.56	0.16	0.24
Inception V3	0.77	0.39	0.87	0.57	0.06	0.16
RESNET-101	0.74	0.39	0.85	0.55	0.06	0.17

Table 4. Ensemble of various deep models on the current dataset with different values of weights to perform weighted averaging for better accuracy

ENSEMBLE MODELS	WEIGHTS	IOU SCORE
➢ ResNet 34	0.3	0.910
➢ InceptionV3	0.6	
➢ ResNet 101	0.1	
➢ ResNet 34	0.4	0.9485
➢ InceptionV3	0.6	
➢ ResNet 101	0	
➢ ResNet 34	0.3	0.911
➢ InceptionV3	0.6	
➢ DenseNet 101	0.1	
➢ DenseNet 101	0.7	0.8963
➢ InceptionV3	0.6	
➢ ResNet 101	0.1	
➢ ResNet 34	0.4	0.900
➢ InceptionV3	0.5	
➢ DenseNet 101	0.5	

4. Discussion

The Present study discusses oral cancer detection through photographic Images, an easily accessible and cost-effective device. Considering low- and mid-income countries, the current modalities to detect the abnormalities are at high cost and are unreachable for socioeconomic background people. The challenge lies in such true color images during feature extraction. The images in the current study are preprocessed,as the images are from different sources with different illumination conditions and irrespective of the regions of theoral cavity like tongue, lips, hard palate, soft palate, gingiva, and tooth gums. Deep learning as an advanced technique is included in the current study [9]. In the present work the task to classify and segment the region of interest using deep architectures are considered [11]. In the present work, Images are fed to the deep architectures to extract features and segment and classify them as cancer or non- cancer. Training and testing of various pre-trained deep network models on the current dataset and the results of each individual network is noted and depicted as conveyed in the tables above. Resnets, InceptionV3, Deepnets, and Efficientnet are used as a classifier to classify the entire image as cancer or non-cancer on the current datasets. As a first step of findings, individual model performances are measured through accuracy, F1 score, loss function, and IOU Scores. On the basis of scores of each individual model depicted in Table 3 and is evaluated as the good classifier model for the current datasets. All the results in the current paper arevalidated by Oncologist.

5. Acknowledgement

We thank Dr. Hamsa Nandini, Consultant Head and Neck surgeon, Kerudi Cancer Hospital, Bagalkot for validating our results and for her constant support and Inputs.

6. Conclusion

The current paper concludes that ensembles of pre-trained models including Resnet's and Inception on UNet (EURI) provide better precision than individual plain networks. The proposed model works on oral cancer images by segmenting and classifying the region as cancer, resulting in improved performances on the existing metrics. The future work of the present study will be to robust the model by elevating the number of images.

References

[1] Guanghuan Yang,[1,†] Luqi We “A Systematic Review of Oral Biopsies, Sample Types, and Detection Techniques Applied in Relation to Oral Cancer Detection” BioTech (Basel). 2022 Mar; 11(1): 5.Published online 2022 Mar 2. doi: 10.3390/biotech11010005

[2] María GraciaSarrión Pérez[a]José V.Bagán,”Utility of imaging techniques in the diagnosis of oral cancer”,Journal of Cranio-Maxillofacial Surgery Volume 43, Issue 9,November 2015, Pages 1880-1894

[3] Huiping Lin, Hanshen Chen,”Automatic detection of oral cancer in smartphone-based images using deep learning for early diagnosis”,J Biomed Opt. 2021 Aug; 26(8): 086007. Published online 2021 Aug 28. doi: 10.1117/1.JBO.26.8.086007

[4] Kritsasith Warin , Wasit Limprasert “Automatic classification and detection of oral cancer in photographic images using deep learning algorithms”J Oral Pathol Med 2021 Oct;50(9):911-918. doi: 10.1111/jop.13227. Epub 2021 Aug 16.

[5] Fahed Jubair, Omar Al-karadsheh,”A novel lightweight deep convolutional neural network for early detection of oral cancer”,26 February 2021https://doi.org/10.1111/odi.13825.

[6] Roshan Alex Welikala , Paolo Remagnino,”Automated

[7] Andre W Van Zyl, Belinda Bunn “Clinical features of oral cancer Sefako Makgatho Health Sciences UniversTandheelkundige VerenigingNovember 2012 SADJ:journal of the South African Dental Association = tydskrif van die Suid-Afrikaanse Tandheelkundige Vereniging 67(10):566-9.

[8] Mohammed Shamim, Sadatullah Syed 2 , Mohammad Shible “Automated detection of oral precancerous tongue lesions using deep learning for early diagnosis of oral cavitycancer”.

[9] DEEPORCD: Detection of Oral Cancer using Deep Learning R. Dharani and S. Revathy Journal of Physics: Conference Series 1911 (2021) 012006 IOP Publishing doi:10.1088/1742-6596/1911/1/012006.

[10] Fati, S.M.; Senan, E.M.;Javed, Y. Early Diagnosis of Oral Squamous Cell Carcinoma Based on Histopathological Images Using Deep and Hybrid Learning Approaches. Diagnostics 2022, 12,1899. https://doi.org/10.3390/diagnostics12081899.

[11] Jiazhi Liang ”Image classification based on RESNET “Computer technology, Northwest MuZu University, LanZhou, GanShu, China 2020 J. Phys.: Conf. Ser. 1634 012110.

[12] Toqa A. Sadoon and Mohammad H. Ali “An Overview of Medical Images Classification based on CNN” International Journal of Current Engineering and Technology E-ISSN 2277 – 4106, P-ISSN 2347 –

5161.

[13] A. Ramalingam et al 2020 IOP Conf. Ser.: Mater. Sci. Eng. 993 01205 Inceptionv3 Analysis of Oral Squamous CellCarcinoma into Various Stages using Pre-Trained Convolutional Neural Networks.

[14] Mohanad A. Deif,Hani Attar "Diagnosis of Oral Squamous Cell Carcinoma Using Deep Neural Networks andBinary Particle Swarm Optimization on Histopathological Images: An AIoMT Approach" Volume 2022 | Article ID 6364102 | https://doi.org/10.1155/2022/6364102.

[15] Bofan Song 1 , Shaobai Li "Exploring uncertainty measures in convolutional neural network for semantic segmentation of oral cancer images" 2022 Nov;27(11):115001. doi: 10.1117/1.JBO.27.11.115001.

[16] Olaf Ronneberger,Philipp Fische "U-Net: Convolutional Networks for Biomedical Image Segmentation" International Conference on Medical Image Computing and Computer-Assisted Intervention.

[17] MICCAI 2015: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015 pp 234–241.

[18] Kaiming He, Xiangyu Zhang,"Deep Residual Learning for Image Recognition"Computer Vision and PatternRecognition, Submitted on 10 Dec 2015 https://doi.org/10.48550/arXiv.1512.03385

[19] Christian Szegedy,Vincent Vanhoucke "Rethinking theInception Architecture for Computer Vision"CVPR 2016.

[20] Kavitha Shaga Devan, Hans Kestler," Weighted average ensemble-based semantic segmentation in biological electron microscopy images", DOI:10.1007/s00418-022-02148-3,Springer.

[21] Qureshi, D. I. ., & Patil, M. S. S. . (2022). Secure Sensor Node-Based Fusion by Authentication Protocol Using Internet of Things and Rfid. Research Journal of Computer Systems and Engineering, 3(1), 48–55. Retrieved from https://technicaljournals.org/RJCSE/index.php/journal/article/view/41

[22] Fu, J. ., & Saad, N. H. M. . (2023). Cross Border E-Commerce Uses Blockchain Technology to Solve Payment Risks. International Journal on Recent and Innovation Trends in Computing and Communication, 11(3s), 205–215. https://doi.org/10.17762/ijritcc.v11i3s.6182