JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Improving Laryngeal Cancer Detection Accuracy Through Deep Learning and Attention Mechanisms

Puneet Misra[1], V. K. Saraswat[2], Mohd. Usman Siddharth[3], Rakesh Srivastava[4]

[1]University of Lucknow, Lucknow, PIN-226007

[2]Institute of Engineering & Technology, Dr B R Ambedkar University, Khandari campus Agra

[3] School of Data Science Symbiosis University of Applied Sciences, Indore, Madhya Pradesh

[4] ENT Department, Health City Vistar Hospital, Gomti Nagar, Lucknow, India,

Abstract: This study emphasizes the pressing need for improved laryngeal cancer detection. Current diagnostic methods, primarily endoscopic imaging, are hindered by the increasing volume of patient data and the potential for invasive biopsies to cause lasting damage. To mitigate these challenges, an artificial intelligence-driven framework is proposed, leveraging deep learning to enhance diagnostic accuracy and efficiency. This framework integrates VGG16 for robust feature extraction with an attention mechanism and an LSTM decoder, enabling precise identification of laryngeal cancer. By incorporating channel and spatial attention layers, the model focuses on relevant image features, improving detection accuracy. The suggested strategy is to address issues with prediction accuracy, resource use, and real-time performance to provide a more effective and assessable tool for early laryngeal cancer detection and better patient outcomes.

Keywords: Laryngeal cancer, endoscopic images, artificial intelligence, deep learning, VGG16, encoder, decoder, CBAM, LSTM

I. INTRODUCTION

Laryngeal cancer is the seventh most malignancy that affects the larynx and have significant impact on patient's life quality and overall health outcomes [1]. There are numerous risk factors for this type of cancer, such as extended exposure to sunlight, risky behaviours like smoking, binge drinking, and chewing betel nuts, infectious diseases like Epstein-Barr virus and human papillomavirus, lifestyle choices like eating few fruits and vegetables, and a family history of cancer[2][3].Therefore early and assessable detection of cancer is the demand for saving the life of patients.

Recently, endoscopic imaging has been regarded as the gold standard for sufficient human screening for anaccurate diagnosis of primary stage laryngeal cancer globally. Furthermore, the number of laryngeal cancer patients has been rising exponentially, the increasing number of patients brings with it a lot of data that professionals must analyze and handle.. However, the interpretation of medical visuals is aninefficient process and often leads to errors, might result in scarring and tissue damage. Particularly, the larynx's function is very susceptible to epithelial alterations and scarring, indicating that even minor tissue biopsies may result in long-term negative consequences like dysphonia [4].

To resolve the aforementioned problems Artificial intelligence techniques have been put forth to help physicians diagnose, evaluate risk, and make decisions regarding head and neck cancer in order to improve screening, detection, and prognosis. [5][6][7]. Moreover, due to the capability of these techniques to process a large amount of medical visuals lead to the early assessment of laryngeal cancer by using deep learning algorithms [8][9][10][11][12][13].The current techniques for detecting laryngeal cancer cells have poor real-

time performance, significant resource consumption, and low prediction accuracy. In this study an encoder decoder-based framework has been proposed to overcome these problem to some extent.

II. RELATED WORK

Artificial intelligence-based automated computing techniques, like as deep learning, do better in identifying voice diseases than picture categorization.

Sarah A et. al. [14] introduces a novel method called the Laryngeal Cancer Diagnosis using the Dandelion Optimizer Algorithm with Ensemble Learning (LCD-DOAEL). This method is specifically designed to analyse throat region images for detecting laryngeal cancer, showcasing the integration of advanced algorithms in medical diagnostics. Claudio Sampieri et. al. [15]proposed a deep learning model named SegMENT-Plus, specifically designed for the automatic segmentation of laryngeal cancer in endoscopic images. A new framework is presented in the study done by J. Sharmila Joseph et al. [16] that blends handcrafted features from Local Binary Pattern (LBP) and First-order statistics (STAT) with deep learning features from DenseNet 201. This hybridization enhances the representation of features extracted from endoscopic narrowband images of the larynx, leading to more effective classification of laryngeal tissues. Divya Rao et. al. [17] in her study explored the complexity of model as it helps in understanding the trade-offs between model complexity and diagnostic accuracy, this is necessary for creating detection systems that work..The study emphasises the importance of early detection of laryngeal cancer, which is critical for improving treatment outcomes. By focusing on this aspect.

Yi-Fan Kang et. al. [18] suggested the Intelligent Laryngeal Cancer Detection System (ILCDS), a deep learning-based solution specifically designed for effective laryngeal cancer screening in resource-constrained rural areas. This system addresses the challenges posed by a shortage of laryngologists and limited computer resources in these regions. The study also emphasized that the use of lightweight models, particularly MobileNet, which excels in compact size and fast inference speed. This makes it suitable for integration into the ILCDS, guaranteeing that even with constrained computational resources, the system can function efficiently. Wei Wang et.al. [19] successfully created five survival analysis models tailored for glottic carcinoma and non-glottic carcinoma (which includes supraglottic and subglottic types). This distinction makes it possible to make prognostic forecasts based on particular cancer subtypes that are more accurate. The study highlighted the effectiveness of the Random Survival Forest (RSF) model, which outperformed traditional models. The C-index values achieved were 0.687 for glottic carcinoma and 0.657 for non-glottic carcinoma, indicating a strong predictive capability.Hyun-Bum Ki et.al. [20] proposed an artificial intelligence model specifically designed to differentiate between healthy voices, laryngeal cancer voices, and those affected by other laryngeal conditions such as vocal cord paralysis and benign mucosal diseases. This model aimed to enhance early detection of laryngeal cancer based on voice analysis, which is a novel approach in this domain.

III. STUDY DESIGN

Dataset:

The study has been performed on the Zenodo's publicly available dataset, based on benign and malignant tissues. The dataset is made up of visualisation of 210 adult patient's sub epithelial tissues with different classification. Table 1 describes the distribution of images in the dataset.

Table 1: Distribution of images in the dataset

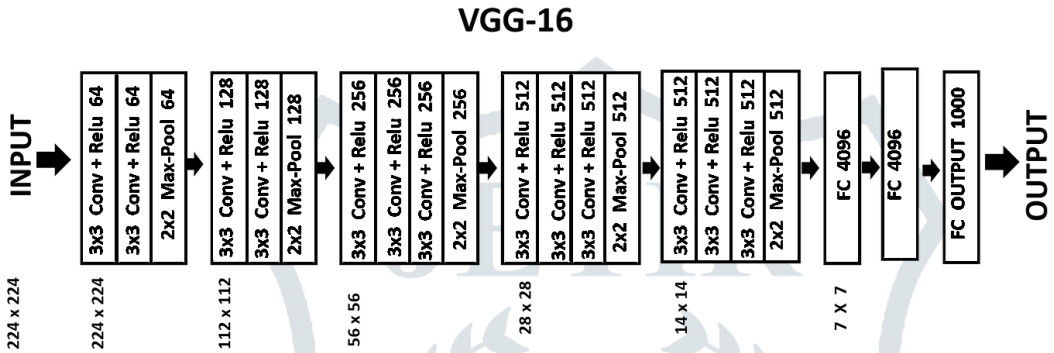
Class	Number of Images
Benign	7,657
Malignant	3,487

IV. MODEL ARCHITECTURE

The study utilizes the strength of encoder and decoder based neural network architecture. Encoder module extracts the characteristics of the image data and decoder module decodes the features for generating the probabilities.

To obtain details from the input picture, the encoder VGG16 [21] was utilized.VGG16 stands out among the well-known CNN architectures due to its straightforward design and contribution to the comprehension of network depth in feature representation. There are sixteen weight layers, thirteen convolution layers, and three fully linked layers in the network. Pretrained VGG16, utilizing the weights of image net, having the strength of transfer learning that reduces the training time of the network and optimize the computational power. The VGG16 network processes the input feature map up to the block5_conv3 layer.

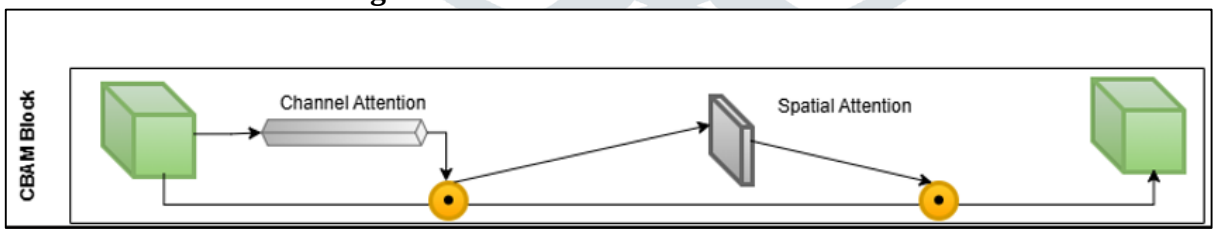
Figure 1: Architecture of VGG16 architecture



Convolutional block attention module (CBAM) [22] receives the encoder's features and uses them to focus on the pertinent information, reducing the dimensionality of the feature maps.CBAM is a useful neural network architecture attention module. CBAM module helps to focus on the important features of the input image by suppressing the unnecessary noisy data from the input image. The Channel Attention module and the Spatial Attention module are the two distinct modules that make up the attention module. Every channel is regarded as a feature descriptor in the Channel Attention module. The aggregate spatial feature map was created by increasing the representational capacity of the input data via the use of maxpooling and averagepooling processes. The spatial attention module assesses the inter-spatial connection of features to concentrate on the pertinent information.

The input feature maps are multiplied by the attention maps generated by these two channels in order to enhance feature extraction. The structure of CBAM module can be represented as follows.

Figure 2: Architecture of CBAM module



The CBAM block's output feature map is used by Long Short-Term Memory (LSTM) [23]. There are concealed short-term memory states and a large number of intermediate long-term memory cells in LSTM. This architecture handles the flow of information effectively. The main component of LSTM architecture are input, forest and output gate, ensuring to regulate the information flow and discard the irrelevant information.

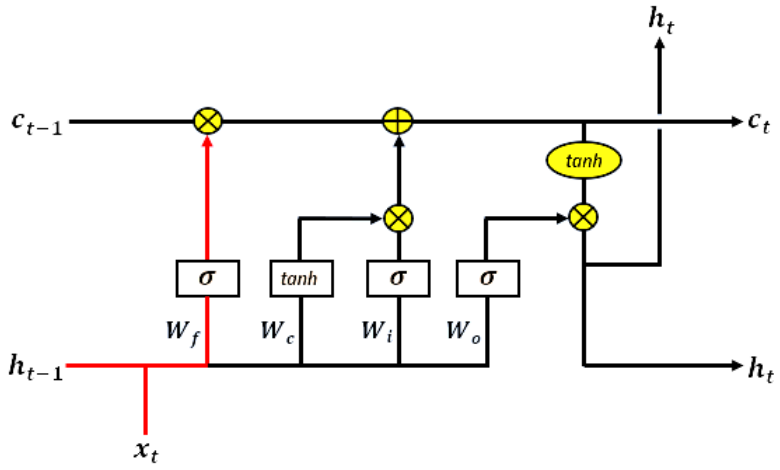
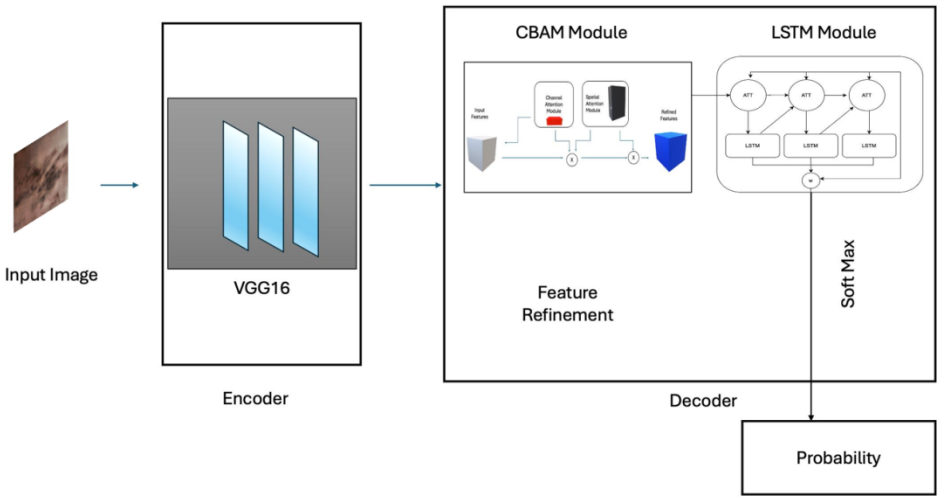


Figure 3: Architecture of LSTM

The output probabilities produced by the LSTM output gate used to predict the correct class. Figure 4 depicts the overall architecture of the proposed network.

Figure 4: Proposed framework for the model



The proposed network utilizes the VGG16 network processes the input image up to the block5_conv3 layer to extract the features. CBAM module refines these features by highlighting important channels and spatial locations. Global average pooling has been applied that converts the spatial features to a 1D representation. A dense layer is added to the architecture to create the final feature embeddings. The features are processed through an LSTM (as a sequence of length 1). A final dense layer with sigmoid activation has been applied to produce the probabilities.

V. Implementation Details and Results

Experimental setup

The current study is carried out using Kaggle P100 GPU having 16GB RAM. Additionally, python, for its vast library of machine learning task, is used to develop the model. Several libraries have been utilised including numpy for numerical calculations, pandas to build the data frames, and matplotlib to plot the results.

Performance metrics:

The study has been evaluated on the several metrics including precision, recall, accuracy and F1 Score which can be calculated as follows

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive\ + False\ Negative}$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negaive}$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$F1\ Score = 2\ X\ \frac{Precision\ X\ Recall}{Precision + Recall}$$

Results:

The study suggested an encoder-decode based framework to improve the categorization of benign and malignant lesions by utilizing VGG16 and an attention mechanism with LSTM. The accuracy of the proposed approach reaches 95%. Table 2 shows the quantitative result obtained from the proposed method and Table 3 shows the hyperparameter settings of the experiment.

Table 2: Quantitative analysis of precision, recall, and F1 score obtained from the proposed network

Class	Precision(%)	Recall(%)	F1 Score(%)
Benign	97	96	96
Malignant	90	94	92

Table 3: Hyperparameter settings used in the experiment

Hyperparameter	**Setting**
Batch size	32
Epoch	100
Activation Function	reLu in each convolutional layer and softmax in Dense Layer
Input Shape	224, 224, 3
Loss Function	Categorical cross entropy
Optimizer	Adamx with learning rate 0.0001

Figure 5 shows the graph of training process. The training graph depicts that model learns linearly with respect to the epochs. Figure 6 shows the confusion matrix obtained from the proposed network after training. The confusion matrix demonstrates how effectively the model distinguishes between cancerous and benign cells.

Figure 5: Training and validation accuracy graph of the proposed architecture

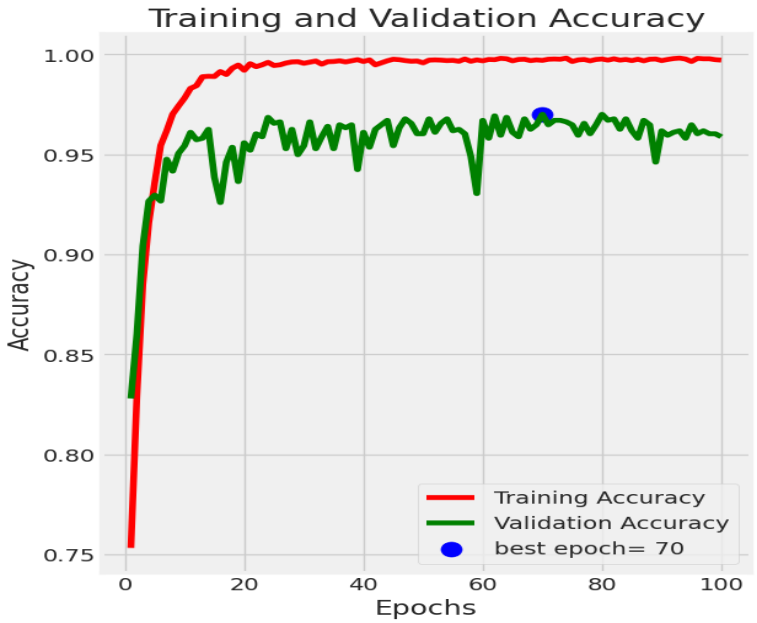
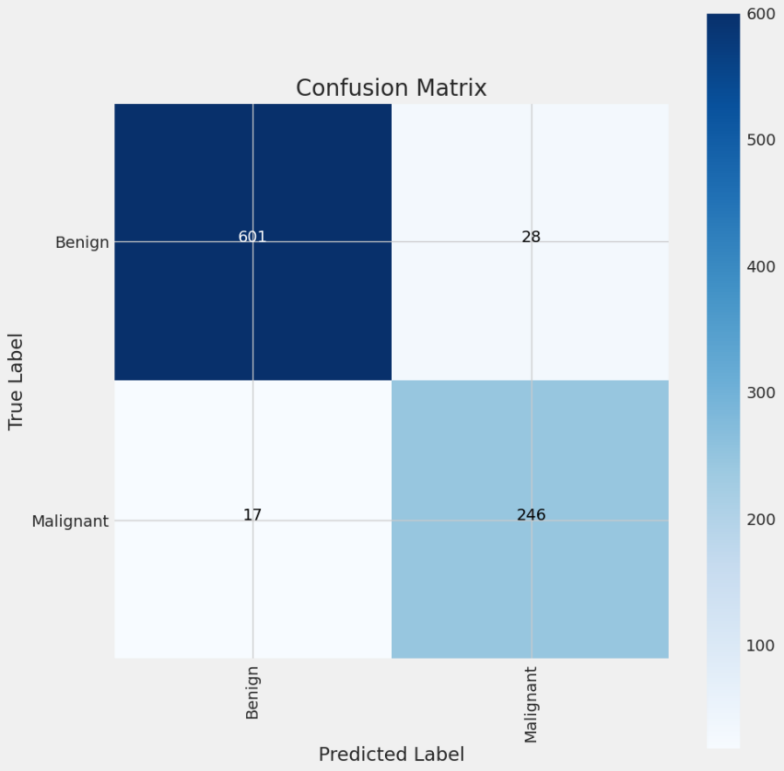


Figure 6: Confusion matrix obtained from the proposed architecture



VI. Conclusion

The study has effectively integrated VGG16 with attention mechanism and LSTM decoder, achieving a breakthrough in the identification of laryngeal cancer. Additionally, the proposed structure is capable for better feature mining and focuses on the relevant features by utilizing channel and spatial attention layers. The result of proposed framework reflects in better accuracy, precision, recall, and F1 Score and can be further evaluated with real time clinical data under observation.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020:GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 Countries. CA Cancer J Clin.2021;71(3):209–49.
2. Pfister DG, Ang K-K, Brizel DM, Burtness BA, Cmelak AJ, Colevas AD, et al. Head and neck cancers. J Natl ComprCancNetw. 2011;9(6):596–650.
3. Adeoye J, Thomson P. Strategies to improve diagnosis and risk assessment for oral cancer patients. Faculty Dental J. 2020;11(3):122–7.
4. M. D. Weller, P. C. Nankivell, C. McConkey, V. Paleri, H. M.Mehanna, Clin. Otolaryngol. 2010, 35, 364.
5. Adeoye J, Tan JY, Choi S-W, Thomson P. Prediction models applying machine learning to oral cavity cancer outcomes: A systematic review. Int J Med Inform. 2021;154: 104557.
6. Volpe S, Pepa M, Zaffaroni M, Bellerba F, Santamaria R, Marvaso G, et al. Machine Learning for Head and Neck Cancer: A Safe Bet?—A Clinically Oriented Systematic Review for the Radiation Oncologist. Front Oncol. 2021;11:89.
7. Mahmood H, Shaban M, Rajpoot N, Khurram SA. Artificial Intelligence-based methods in head and neck cancer diagnosis: an overview. Br J Cancer. 2021;124(12):1934–40.
8. Nuha Qais Abdulmajeed. (2022). Belal Al-Khateeb, and Mazin Abed Mohammed, “a review on voice pathology: Taxonomy, diagnosis, medical procedures and detection techniques, open challenges, limitations, and recommendations for future directions,.” Journal of Intelligent Systems, 31(1), 855–875.

9. Fahad Taha Al-Dhief, Marina Mat Baki, NurulMu'azzah Abdul Latiff, Nik Noordini Nik Abd. Malik, NaseerSabri Salim, Musatafa Abbas AbboodAlbader, Nor MuzlifahMahyuddin, and Mazin Abed Mohammed, 2021. "Voice Pathology Detection and Classification by Adopting Online Sequential Extreme Learning Machine," IEEE Access, 9: 77293–77306,
10. Abdulmajeed, N. Q., Al-Khateeb, B., & Mohammed, M. A. (2023). Voice pathology identification system using a deep learning approach based on unique feature selection sets. Expert Systems. https://doi.org/10.1111/exsy.13327
11. Bera, K., Schalper, K.A., Rimm, D.L., Velcheti, V. and Madabhushi, A. (2019) Artificial Intelligence in Digital Pathology—New Tools for Diagnosis and Precision Oncology. Nat Rev Clin Oncol, 16, 703-715.https://doi.org/10.1038/s41571-019-0252-y
12. Campanella, G., Hanna, M.G., Geneslaw, L., Miraflor, A., Silva, V.W.K., Busam, K.J., Brogi, E., Reuter, V.E., Klimstra, D.S. and Fuchs, T.J. (2019) Clinical-Grade Computational Pathology Using Weakly Supervised Deep Learning on Whole Slide Images. Nature Medicine, 25, 1301-1309. https://doi.org/10.1038/s41591-019-0508-1
13. Fourcade, A. and Khonsari, R.H. (2019) Deep Learning in Medical Image Analysis A Third Eye for Doctors. Journal of Stomatology, Oral and Maxillofacial Surgery, 120, 279-288. https://doi.org/10.1016/j.jormas.2019.06.002
14. Alzakari, S. A., Maashi, M., Alahmari, S., Arasi, M. A., Alharbi, A. A., & Sayed, A. (2024). Towards laryngeal cancer diagnosis using Dandelion Optimizer Algorithm with ensemble learning on biomedical throat region images. *Scientific Reports*, *14*(1), 19713.
15. Sampieri, C., Azam, M. A., Ioppi, A., Baldini, C., Moccia, S., Kim, D., ... & Peretti, G. (2024). Real-time laryngeal cancer boundaries delineation on white light and narrow-band imaging laryngoscopy with deep learning. *The Laryngoscope*, *134*(6), 2826-2834.
16. Joseph, J. S., Vidyarthi, A., & Singh, V. P. (2024). An improved approach for initial stage detection of laryngeal cancer using effective hybrid features and ensemble learning method. *Multimedia Tools and Applications*, *83*(6), 17897-17919.
17. Rao, D., Singh, R., Koteshwara, P., &Vijayananda, J. (2024). Exploring the Impact of Model Complexity on Laryngeal Cancer Detection. *Indian Journal of Otolaryngology and Head & Neck Surgery*, *76*(5), 4036-4042.
18. Kang, Y. F., Yang, L., Xu, K., Hu, B. B., Cai, L. J., Liu, Y. H., & Lu, X. (2024). A lightweight intelligent laryngeal cancer detection system for rural areas. *American Journal of Otolaryngology*, *45*(6), 104474.
19. Wang, W., Wang, W., Zhang, D., Zeng, P., Wang, Y., Lei, M., ... & Cai, C. (2024). Creation of a machine learning-based prognostic prediction model for various subtypes of laryngeal cancer. *Scientific Reports*, *14*(1), 6484.
20. Kim, H. B., Song, J., Park, S., & Lee, Y. O. (2024). Classification of laryngeal diseases including laryngeal cancer, benign mucosal disease, and vocal cord paralysis by artificial intelligence using voice analysis. *Scientific Reports*, *14*(1), 9297.
21. Tammina, S. (2019). Transfer learning using vgg-16 with deep convolutional neural network for classifying images. *International Journal of Scientific and Research Publications (IJSRP)*, *9*(10), 143-150.
22. Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 3-19).
23. Zhang, B., Wang, Q., Gao, Z., Zeng, R., & Li, P. (2022). Temporal grafter network: Rethinking lstm for effective video recognition. *Neurocomputing*, *505*, 276-288.