Breast Ultrasound Image Segmentation Based on Attention U-Net

Yiming Lu

{2330157381@qq.com}

East China University of Science and Technology, No. 130, Meilong Road, Shanghai, China

Abstract. Breast cancer remains a significant health challenge, with early diagnosis critical for improving patient outcomes. This study explores the application of the Attention U-Net model for breast ultrasound image segmentation, aiming to enhance diagnostic accuracy in breast cancer detection. Experimental results demonstrated the model's accuracy stabilizing at approximately 0.95, with predicted masks showing close alignment to expert-annotated ground truths. Additionally, Grad-CAM visualizations illustrated the model's capability to concentrate on critical regions, enhancing interpretability. Despite its computational demands, the Attention U-Net model offers significant potential for medical applications, providing a robust framework for improving clinical diagnosis and treatment planning in breast cancer care.

Keywords: Breast cancer, Ultrasound image, Image segmentation, Attention U-Net.

1 Introduction

Cancer has long been one of the significant health challenges faced by humanity [1], with many patients discovering their condition only when it is too late, significantly reducing the chances of successful treatment. Therefore, early diagnosis of cancer is crucial, as it not only improves patient survival rates but also provides better timing and conditions for subsequent treatment. Among various diagnostic technologies, ultrasound has played an important role in cancer diagnosis due to its unique advantages.

Ultrasound technology offers significant benefits in the diagnosis of breast cancer [2]. Firstly, it enhances diagnostic accuracy. Secondly, compared to traditional radiological detection methods, ultrasound is safer as it does not involve radioactive substances or contrast agents, thereby avoiding radiation exposure or allergic reactions in patients. Furthermore, ultrasound enables real-time imaging, allowing physicians to dynamically observe the patient's internal structure and changes during the examination, which is vital for assessing the benign or malignant nature of tumors and determining the presence of metastasis. The application of precise image segmentation in medical image processing holds vast potential. Through accurate image segmentation, doctors can more reliably identify pathological tissues, leading to more effective treatment plans. Additionally, image segmentation techniques are also crucial in fields such as autonomous driving, robotic vision, and remote sensing analysis. However, the implementation of image segmentation technologies faces numerous challenges, primarily including issues such as blurred edges, occlusion, and noise, all of which can impact segmentation accuracy.

CONF-MLA 2024, November 21, Adana, Turkey

DOI 10.4108/eai.21-11-2024.2354629

The U-Net model, proposed by Olaf Ronneberger et al. in 2015, is a deep learning model widely used in biomedical image processing [3]. U-Net, with its symmetric U-shaped architecture and efficient feature extraction capability, has become one of the preferred models for medical image segmentation tasks. The model consists of two main parts: the encoder and the decoder. The encoder is responsible for feature extraction, capturing high-level features from images through a combination of convolutional layers and max pooling layers. The decoder, on the other hand, reconstructs the feature maps into high-resolution segmentation images through deconvolution and upsampling. The skip connections in U-Net allow for effective concatenation of feature maps from corresponding layers in the encoder and decoder, thereby integrating semantic and spatial information at different levels.

Despite the significant achievements of U-Net in medical image segmentation, the model may lose critical detail information when handling complex images, particularly in the decoder phase. Additionally, while the skip connections in U-Net help retain spatial information, they may also introduce redundant features, increasing computational load. To address these issues, Ozan Oktay and his team proposed the Attention U-Net model [4]. This model introduces an attention mechanism on the basis of U-Net, assigning weights to each pixel through a soft attention module, enabling the model to focus more on features relevant to the target area while suppressing activations from irrelevant regions. This mechanism helps reduce redundant information and improve segmentation accuracy. Therefore, when addressing medical image segmentation challenges, particularly in scenarios requiring precise segmentation of complex anatomical structures or pathological regions, the recognition accuracy of Attention U-Net surpasses that of traditional U-Net.

This study aiming to explore the application and effectiveness of this model in the processing of breast cancer ultrasound images. Specifically, we will analyze the performance of Attention U-Net in breast ultrasound image segmentation tasks, assessing its ability to enhance segmentation accuracy and reduce false detection rates. Additionally, the study will discuss the advantages of the model in handling complex pathological regions, including how the attention mechanism effectively focuses on key features to more accurately identify and segment different types of tumors. We hope that the results of this research will contribute to the further development of breast cancer ultrasound image analysis, providing more precise solutions for clinical diagnosis and treatment.

2 Previous works

Traditional image segmentation methods have historically held an important position in the field of image processing. These early segmentation techniques employed manually designed features and rules to delineate different regions within images, encompassing various classic methods such as thresholding, region growing, and edge detection. These techniques have undergone extensive development, laying a solid foundation for the field of image segmentation.

Thresholding is one of the simplest and most commonly used image segmentation methods [5]. This technique divides the pixels in an image into foreground and background based on one or more predefined thresholds. The advantages of thresholding include its rapid computation and straightforward implementation, making it suitable for images with clear gray value differences.

However, thresholding often performs poorly when dealing with images that exhibit uneven lighting or substantial noise, leading to inaccurate segmentation results.

Region growing operates by selecting one or more initial seed points and expanding the region based on similarity criteria [6]. This method can effectively segment areas with similar characteristics; however, its performance is highly dependent on the choice of initial seed points. Additionally, region growing typically requires significant computational resources, resulting in longer processing times when handling complex structures. It may also encounter issues such as over-segmentation or under-segmentation when processing images with ambiguous boundaries or intricate textures.

Edge detection is another prevalent segmentation technique aimed at identifying the presence of object edges within an image [7]. By calculating the gradients of the image, edge detection can effectively locate strong edges. Nonetheless, this method is very sensitive to noise, and its performance is often suboptimal when dealing with discontinuous or blurred edges, frequently resulting in incomplete edge detection outcomes.

Despite achieving satisfactory results in certain specific scenarios, traditional image segmentation methods face numerous limitations and challenges. Firstly, these methods often rely on manually designed features, which may fail to adequately represent the essential information of complex and variable image data. For instance, some features might not effectively differentiate similar regions or handle images with complex textures, leading to inaccuracies in segmentation results. Secondly, the performance enhancement of traditional methods is often constrained by computational resources and algorithm design. As the volume of data increases dramatically, traditional algorithms frequently exhibit inefficient computation when processing large-scale datasets. Many traditional algorithms require substantial time for feature extraction and parameter tuning, making it difficult to meet the demands of real-time processing. Furthermore, traditional methods struggle to adapt to changing environments and scenarios when addressing complex tasks; for instance, the segmentation performance can significantly degrade under varying lighting conditions and noise influences. Lastly, the limitations of traditional image segmentation methods become increasingly apparent when handling high-dimensional data. With the rise in image resolution and the emergence of multimodal data, traditional methods often fail to fully leverage the rich contextual information available, resulting in suboptimal segmentation outcomes.

In this context, the introduction of deep learning technologies offers new possibilities for addressing these challenges. Deep learning models can automatically learn feature representations, enabling them to capture complex patterns and structures within images more effectively, thereby achieving more precise segmentation results [8]. This shift not only enhances the accuracy of image segmentation but also paves the way for handling complex scenes and large-scale datasets.

With the explosive growth of data and the enhancement of computational capabilities, the application of deep learning in the field of image segmentation has become increasingly widespread. Deep learning models have demonstrated significant advantages in processing complex image data, as they can automatically learn feature representations from vast amounts of data without the need for manual feature design. This capability allows deep learning models to adapt to various types of images and their diverse features, resulting in more accurate and efficient image segmentation. The introduction of deep learning technologies marks a

significant advancement in image segmentation techniques. By constructing deep neural networks, these models can extract richer feature hierarchies, capturing complex patterns and structures in images from low-level features to high-level semantic information. This multi-layered feature learning ability enables deep learning models to maintain high segmentation accuracy even in the face of complex backgrounds, lighting variations, and noise disturbances. Furthermore, the scalability of deep learning models provides broad prospects for their application in image segmentation. Through transfer learning and data augmentation techniques, researchers can train efficient models on limited labeled data, significantly improving segmentation performance. This capability is particularly beneficial in fields such as medical image segmentation, where the costs of sample acquisition and annotation are high; deep learning can effectively overcome these challenges and promote advancements in related research.

3 Dataset and Preprocessing

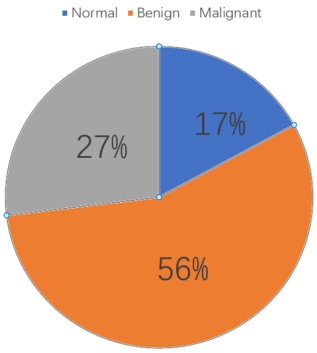

Fig. 1. Data distribution.

The dataset used in this study is derived from Breast Ultrasound Images [2]. This dataset is designed to facilitate the early diagnosis and detection of breast cancer, encompassing three categories of breast images: normal, benign, and malignant. It supports research on classification, detection, and segmentation based on deep learning techniques.

In terms of sample size, the dataset is relatively balanced, consisting of 133 normal images, 437 benign images, and 210 malignant images. The overall composition is illustrated in Figure 1, which reflects the distribution of samples across different categories. This diverse sample type provides rich data support for model training, enhancing the model's generalization ability and adaptability. All images were acquired at high resolution to ensure optimal image quality during analysis and processing. The resolution of the images is 256x256 pixels, which not only provides sufficient detail for deep learning models but also enables effective recognition and differentiation of various breast tissue characteristics. High-resolution images capture finer variations during feature extraction, thus improving the accuracy of both segmentation and classification.

Before training the deep learning model, several preprocessing steps were undertaken to ensure data quality and model efficacy. Firstly, to address the dataset's diversity, data augmentation

techniques were applied, including random rotation, scaling, and flipping. These augmentation methods aim to increase the number of training samples, enabling the model to maintain high accuracy when encountering images of varying orientations, sizes, and shapes. Additionally, data augmentation effectively reduces the model's dependence on specific samples, preventing overfitting. Secondly, all images were normalized to a range of 0 to 1 by dividing the pixel values by 255. This normalization step not only facilitates faster convergence of the model but also enhances training efficiency and stability. The standardized data effectively reduces variations in brightness and contrast among different images, allowing the model to focus more on learning features rather than being disturbed by noise.

Subsequently, corresponding mask images were generated from the original images. These mask images, which share the same resolution as the original images, are used to identify the locations of regions of interest. The masks were manually annotated by experts to ensure that the model receives accurate target region information during training, thereby improving segmentation precision and accuracy. Finally, to evaluate the model's performance, the dataset was divided into training, validation, and testing sets, with the training set comprising 80% of the data for model training, the validation set accounting for 10% for parameter tuning, and the testing set making up the remaining 10% for final performance assessment. This division ensures the effectiveness and reliability of the model training process while enabling comprehensive performance evaluation across different datasets, thereby confirming the model's efficacy in practical applications.

4 Model

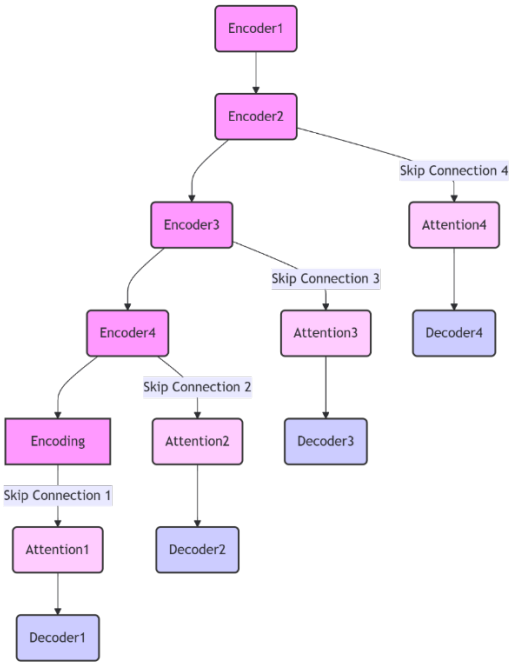

Fig. 2. Model architecture.

This study employs the Attention U-Net model (Figure 2), which integrates an attention mechanism and builds upon the classic U-Net architecture. The core advantage of Attention U-Net lies in its ability to identify and focus on key regions within images, thereby significantly enhancing the accuracy of segmentation tasks. This feature is particularly important in the field of medical image segmentation, as it allows for more precise identification of pathological areas, providing robust support for clinical diagnosis.

The choice of Attention U-Net in this research is based on several reasons. First, by introducing the attention mechanism, Attention U-Net can better recognize and concentrate on critical features within images, which is essential for improving the segmentation accuracy of breast ultrasound images. Second, within the encoder-decoder structure of U-Net, Attention U-Net reduces unnecessary information transmission through attention gates, thereby enhancing both the efficiency and accuracy of the model. Despite the incorporation of the attention mechanism, Attention U-Net is designed to maintain computational efficiency, making it more practical for real-world applications. Furthermore, the model demonstrates stable performance improvements across different datasets and training scales, indicating strong generalization capabilities.

In the implementation of this project, the training process and hyperparameter settings of the model have been meticulously planned. The dataset is divided into a training set (80%) and a validation set (20%) to facilitate model training and evaluation. This division ensures that the model receives ample data support during training while allowing for an assessment of its generalization ability during validation. The batch size is set to 8, a choice determined by memory capacity and training efficiency, effectively balancing training speed and memory usage. The model training is conducted over 20 epochs to ensure sufficient learning of data features. An appropriate number of iterations aids in achieving a balance between stability and accuracy, helping to avoid overfitting. The steps per epoch are calculated based on the batch size and the size of the training set, ensuring that all data is processed during each training cycle. This setup guarantees that the model can effectively utilize all samples for learning in each epoch.

In terms of loss function selection, we employed the binary cross-entropy loss function, a common choice for binary classification problems, particularly suitable for scenarios where the model outputs probability values. This loss function effectively measures the discrepancy between the predicted probability distribution and the true labels, guiding the model training process. To accommodate the characteristics of the loss function, we selected the Adam optimizer. As an adaptive learning rate optimization algorithm, Adam combines the advantages of both RMSProp and Momentum, allowing for flexible adaptation to different training scenarios, thereby enhancing training efficiency and performance.

5 Results

On the test set, the model's accuracy stabilized at approximately 0.95 in the later stages of training, demonstrating its progressive adaptation to the medical image segmentation task and continuous optimization in segmentation performance. This improvement was particularly evident when compared to the standard U-Net, which achieved a lower accuracy of only 0.83 on the same task. The key advantage of the Attention U-Net over the traditional U-Net lies in

its ability to learn spatially adaptive attention maps. These attention maps help the model focus on the most relevant regions of the image, suppressing irrelevant or background information. In medical image segmentation, where precision is critical and the structures of interest may be small or surrounded by complex backgrounds, this ability to focus attention significantly improves segmentation accuracy. In contrast, the standard U-Net relies on fixed, non-adaptive convolutional layers, which can be less effective in handling the complex and varying features in medical images, leading to lower segmentation performance.

To further illustrate the results, Figure 3 compares the model's predicted masks with the original masks. As shown, the predicted masks align closely with the original masks, accurately segmenting target regions in the medical images. This high degree of alignment indicates that the model achieves precise segmentation and can effectively identify regions of interest. To gain deeper insights into the model's decision-making process, we applied Grad-CAM [9] to visualize the model's attention regions (Figure 3). The results show that during segmentation, the model successfully focuses on key areas within the images, suggesting that it leverages critical information when making segmentation decisions. This visualization aids in understanding the model's behavior and provides a useful guide for further optimization.

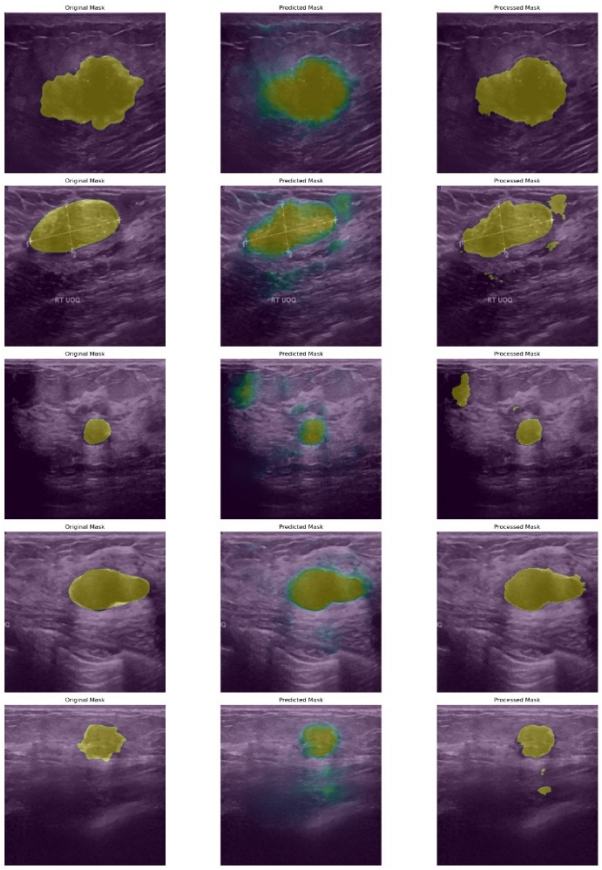

Fig. 3. The illustration of manual segmentation (left), our results (middle) and Grad-CAM results (right).

In conclusion, our study demonstrates the strong potential of deep learning in medical image segmentation. By integrating an attention mechanism, we further improved the model's performance. These findings not only provide valuable technical support for medical image analysis and clinical diagnosis but also pave the way for future research in this field.

6 Conclusion

This paper presents an Attention U-Net-based deep learning model, which incorporates an attention mechanism to effectively enhance segmentation accuracy and generalization capability in medical imaging. Experimental results demonstrate the model's outstanding performance, underscoring its adaptability to various datasets. Furthermore, compared with other complex deep learning models, the Attention U-Net achieves high computational efficiency without compromising precision, making it more feasible for real-world medical applications.

However, the Attention U-Net model does have certain limitations, particularly in terms of computational resource demands and sensitivity to specific datasets. While the introduction of the attention module significantly improves segmentation accuracy, it also increases computational complexity and resource consumption. Implementing the attention mechanism requires intensive matrix operations and feature map processing, resulting in high memory usage and computational costs. This is especially pronounced when handling large-scale datasets or high-resolution images, where resource demands may hinder real-time clinical applications. Furthermore, the added complexity due to the attention mechanism can reduce the interpretability of the model. Although attention weights highlight feature importance, their direct impact on the final segmentation result is not always intuitive, complicating the model’s interpretability and its adaptability to different datasets. Additionally, training is sensitive to various hyperparameters, such as parameter tuning, optimizer selection, and loss function configuration. Misconfiguration can lead to overfitting or underfitting, limiting the model’s generalization and application range.

Looking ahead, several promising directions for improvement include advancements in model structure, data handling, and training strategies. First, collecting more diverse and high-quality data could significantly enhance the model's generalization and accuracy. Given the variability in medical imaging, including disease type, pathological features, and imaging modalities, expanding dataset diversity can help the model capture subtle image characteristics comprehensively. Data augmentation techniques [10], such as rotation, scaling, and mirroring, can further enrich datasets, improving robustness across different scenarios. Exploring efficient large-scale pretrained models and multi-task learning [11] is also critical for future advancements. By leveraging pretrained large models, such as Transformers or deep convolutional neural networks, models can gain valuable prior knowledge from a broad range of visual tasks, enhancing performance in medical image segmentation. Multi-task learning frameworks enable the model to address multiple tasks, such as segmentation and classification, within a single structure, increasing resource efficiency and adaptability. Transfer learning [12] and zero-shot learning [13] also hold great potential for medical imaging analysis. Transfer learning allows pretrained models on natural images to adapt to medical images, offering promising performance even with limited data. Zero-shot learning extends this adaptability,

enabling the model to make accurate predictions for novel pathology types or imaging modalities, thus addressing emerging or rare cases [14]. In terms of model architecture, exploring flexible adaptive attention mechanisms and lightweight designs may help maintain high precision while reducing computational demands. Adaptive attention mechanisms dynamically allocate attention across different regions [15], focusing on relevant information rather than uniformly processing all feature maps. Coupled with pruning and quantization, this approach can achieve a lightweight model structure, making it suitable for embedded systems and resource-constrained medical environments. Enhanced interpretability techniques should also be prioritized in future studies. The inherent "black box" nature of deep learning models limits their clinical applicability. Developing methods for visualizing the model's decision-making processes can offer clinicians more intuitive information. Improved model transparency, achieved through methods like Grad-CAM or other attention weight visualizations, can help clinicians better understand which regions and features the model prioritizes, increasing trust in the model's outputs.

In summary, the Attention U-Net demonstrates immense potential in medical image segmentation. Accurate segmentation can assist clinicians in more precisely delineating tumor boundaries, enhancing breast cancer diagnostic accuracy, informing treatment planning, and supporting treatment monitoring. These advancements provide a solid foundation for future medical imaging analysis and personalized healthcare, paving the way for broader AI integration in healthcare.

References

[1] J. Ferlay *et al.*, "Cancer statistics for the year 2020: An overview, " *International Journal of Cancer*, vol. 149, no. 4, pp. 778–789, 2021, doi: 10.1002/ijc.33588.

[2] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, "Dataset of breast ultrasound images, " *Data in Brief*, vol. 28, p. 104863, Feb. 2020, doi: 10.1016/j.dib.2019.104863.

[3] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation, " in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., Cham: Springer International Publishing, 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.

[4] O. Oktay *et al.*, "Attention U-net: Learning where to look for the pancreas, " May 20, 2018, *arXiv*: arXiv:1804.03999. doi: 10.48550/arXiv.1804.03999.

[5] M. Cheriet, J. N. Said, and C. Y. Suen, "A recursive thresholding technique for image segmentation, " *IEEE Transactions on Image Processing*, vol. 7, no. 6, pp. 918–921, Jun. 1998, doi: 10.1109/83.679444.

[6] J. Tang, "A color image segmentation algorithm based on region growing, " in *2010 2nd International Conference on Computer Engineering and Technology*, Apr. 2010, pp. V6-634-V6-637. doi: 10.1109/ICCET.2010.5486012.

[7] H. G. Kaganami and Z. Beiji, "Region-based segmentation versus edge detection, " in *2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, Sep. 2009, pp. 1217–1221. doi: 10.1109/IIH-MSP.2009.13.

[8] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey, " *IEEE Transactions on Pattern Analysis and*

Machine Intelligence, vol. 44, no. 7, pp. 3523–3542, Jul. 2022, doi: 10.1109/TPAMI.2021.3059968.

[9] R. R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, and D. Batra, "Grad-CAM: Why did you say that?, " Jan. 25, 2017, *arXiv*: arXiv:1611.07450. doi: 10.48550/arXiv.1611.07450.

[10] C. Shorten, T. M. Khoshgoftaar, and B. Furht, "Text data augmentation for deep learning, " *J Big Data*, vol. 8, no. 1, p. 101, Jul. 2021, doi: 10.1186/s40537-021-00492-0.

[11] Y. Zhang and Q. Yang, "An overview of multi-task learning, " *National Science Review*, vol. 5, no. 1, pp. 30–43, Jan. 2018, doi: 10.1093/nsr/nwx105.

[12] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning, " *J Big Data*, vol. 3, no. 1, p. 9, May 2016, doi: 10.1186/s40537-016-0043-6.

[13] Y. Xian, B. Schiele, and Z. Akata, "Zero-shot learning - the good, the bad and the ugly, " presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4582–4591. Accessed: Nov. 14, 2024. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2017/html/Xian_Zero-Shot_Learning_-_CVPR_2017_paper.html

[14] F. Pourpanah *et al.*, "A review of generalized zero-shot learning methods, " *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1–20, 2022, doi: 10.1109/TPAMI.2022.3191696.

[15] W. Li, K. Liu, L. Zhang, and F. Cheng, "Object detection based on an adaptive attention mechanism, " *Sci Rep*, vol. 10, no. 1, p. 11307, Jul. 2020, doi: 10.1038/s41598-020-67529-x.