Proceedings of the 3rd International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/2025.22346

An Improved U-Net Model for Ultrasound Image Segmentation of Breast Cancer

Peihan Guo

School of Management and Engineering, Capital University of Economics and Business, Beijing, China
1762096608@qq.com

Abstract: Breast cancer is the most common and one of the most lethal malignant tumors among women worldwide. Early and accurate diagnosis plays a crucial role in improving patient survival rates. As one of the primary imaging modalities, breast ultrasound imaging has been widely employed in clinical screening due to its low cost and lack of radiation exposure. However, limited by its imaging mechanism, ultrasound images often suffer from severe speckle noise interference, blurred boundaries, and complex tissue structures, which significantly hinder the performance of automatic lesion segmentation. To address this challenge, this paper proposes an improved Attention U-Net model. By introducing Attention Gate modules into the conventional U-Net architecture, the model is guided to focus on salient regions associated with lesions while suppressing background interference. Moreover, the network depth is increased to enhance feature representation capabilities. As a result, the proposed model achieves improved segmentation accuracy and boundary fitting performance in complex scenarios.

Keywords: breast cancer, ultrasound imaging, image segmentation, U-Net, attention mechanism, deep learning

1. Introduction

Breast cancer is a tumor that ranks among the highest in both incidence and mortality among women worldwide [1]. Early detection and timely intervention are crucial for improving patient survival rates. Medical imaging plays an irreplaceable role in clinical practice as an essential tool for breast cancer screening, diagnosis, and therapeutic evaluation.

Among various medical imaging modalities, ultrasound imaging is widely used for the preliminary screening and diagnosis of breast cancer due to its advantages of being non-invasive, real-time, low-cost, and highly sensitive to dense breast tissue. However, ultrasound images are inherently challenged by strong speckle noise, blurred boundaries, and diverse lesion morphologies, making accurate segmentation of lesion regions a difficult task. High-quality segmentation of breast cancer lesions not only aids clinicians in the precise assessment of tumor size, location, and boundaries but also provides critical guidance for subsequent procedures such as surgical planning and radiotherapy target delineation.

Despite its many advantages, ultrasound imaging suffers from several inherent limitations stemming from its imaging principles. Ultrasound images typically exhibit low contrast, making the grayscale difference between lesions and surrounding normal tissues subtle and the visual boundaries

DOI: 10.54254/2755-2721/2025.22346

indistinct. In addition, boundary blurring is common, especially in the transition zone between tumor edges and breast parenchyma, where lesion contours often appear irregular, vague, or gradual, increasing the subjectivity of segmentation. Meanwhile, speckle noise, the most prevalent form of image degradation in ultrasound, is characterized by randomness and high frequency, which not only obscures important structural details but may also lead to the appearance of pseudo-lesions, misleading clinical interpretation. Furthermore, common ultrasound artifacts, such as mirror images, boundary enhancement, and posterior attenuation, further complicate accurate localization and contour extraction of lesion areas.

In recent years, the rapid development of deep learning techniques has introduced novel solutions for breast cancer image analysis. Particularly, end-to-end segmentation models represented by U-Net have gained widespread application in the field of medical image segmentation due to their simple architecture and outstanding performance [2]. U-Net leverages an encoder-decoder structure to achieve multi-level feature fusion and utilizes skip connections to effectively preserve spatial information. It has been extensively applied to the segmentation of various tissues and organs. However, when applied to the specific context of breast ultrasound images, the traditional U-Net model presents several limitations: first, its shallow architecture restricts the capacity for high-level semantic representation of complex lesion morphologies; second, the absence of an explicit attention mechanism hinders the model's ability to effectively focus on key regions within multi-scale contexts; finally, for images with blurred boundaries and complex anatomical structures, the original U-Net struggles to maintain both detailed accuracy and global consistency.

To address these challenges, this study proposes an improved U-Net model based on the Attention U-Net architecture, in which the network depth is increased to enhance feature representation capability. The attention module effectively strengthens the model's focus on critical information while suppressing irrelevant background noise, thereby improving both the accuracy and robustness of lesion segmentation. Experimental results demonstrate that the proposed model outperforms existing methods on multiple breast ultrasound image datasets, indicating strong application potential and clinical value.

2. Previous Works

In recent years, deep learning techniques have achieved remarkable progress in the field of medical image segmentation. Among them, methods based on Convolutional Neural Networks (CNNs) have emerged as the mainstream approach for segmentation tasks. Compared to traditional methods relying on handcrafted feature extraction and rule-based design, deep learning models offer superior capabilities in feature representation and pattern recognition. These models can automatically learn effective features from large-scale datasets, making them well-suited to handle complex image structures and diverse visual tasks. In the segmentation of breast ultrasound images, numerous deep learning-based models have been proposed and have demonstrated promising experimental results.

One of the earliest deep networks applied to image segmentation tasks is the Fully Convolutional Network (FCN) [3]. This network replaces the fully connected layers in conventional classification networks with convolutional layers, enabling end-to-end pixel-level prediction. FCN revolutionized the conventional image segmentation pipeline by eliminating the need for additional post-processing steps and establishing a unified framework for semantic segmentation. Although FCN features a relatively simple architecture and performs well in low-resolution object segmentation, its performance in high-precision medical image processing remains suboptimal. The main issues include coarse resolution recovery during the up-sampling stage, insufficient detail preservation, and the failure to effectively integrate semantic and spatial features across multiple scales. These limitations lead to blurry boundaries and noticeable localization errors.

DOI: 10.54254/2755-2721/2025.22346

To address these shortcomings, U-Net was successfully introduced for biomedical image segmentation tasks [2]. U-Net adopts a symmetric encoder-decoder architecture and utilizes skip connections to fuse high-resolution shallow features with deep semantic features, significantly enhancing the model's ability to preserve boundary and spatial information. The model is characterized by its structural simplicity, strong trainability, and adaptability to small-scale datasets, making it widely used in the segmentation of medical images involving breast cancer, lung nodules, liver, and more. However, the original U-Net still encounters several limitations. Specifically, in the context of breast ultrasound images with vague lesion contours and indistinct inter-tissue boundaries, U-Net shows deficiencies in shallow feature extraction and semantic feature fusion. Furthermore, its limited network depth constrains the model's expressive capacity, impeding the capture of rich multi-scale contextual information and affecting the recognition of small lesions and fine-grained structures.

3. Dataset and Preprocessing

The breast ultrasound image dataset utilized in this study was collected in 2018 by Baheya Hospital for Early Detection and Treatment of Breast Cancer in Egypt [4]. The subjects were female patients aged between 25 and 75 years. The data collection process spanned approximately one year. All images were initially stored in DICOM format and were annotated and verified by experienced radiologists. The original dataset comprised 1,100 grayscale breast ultrasound images, acquired using ultrasound devices including the LOGIQ E9 and LOGIQ E9 Agile systems. These high-end systems are widely employed in radiology, cardiology, and vascular imaging. The images had a resolution of 1280×1024 pixels. A ML6-15-D Matrix linear array probe with a frequency range of 1–5 MHz was used, providing high-resolution tissue imaging suitable for detailed visualization of breast structures.

To ensure the effectiveness of subsequent model training, the images underwent a rigorous preprocessing pipeline, which involved the removal of duplicate scans and samples with annotation errors. Multiple rounds of manual review and correction were conducted by experienced radiologists at Baheya Hospital. After preprocessing, a total of 780 images were retained. The average image size was standardized to 500×500 pixels, and all images were converted to PNG format for model input and training.

The dataset was categorized into three classes based on lesion type: Normal, Benign, and Malignant. Each image was accompanied by a corresponding ground truth mask, which delineated the actual boundaries of the lesion region. Throughout the data collection and usage process, the research team strictly adhered to medical ethical standards. All patient information was anonymized to ensure privacy protection. Prior to data acquisition, informed consent was obtained from both Baheya Hospital and the participating patients. It was explicitly communicated that the data would be used exclusively for scientific research purposes, would not be publicly disclosed, and would be stored and utilized under strict confidentiality protocols.

4. Model

As one of the most representative deep neural network architectures for medical image segmentation tasks, U-Net has demonstrated strong performance in feature extraction and spatial information recovery due to its symmetric encoder-decoder structure and skip connection mechanism. It has been widely applied across various medical imaging scenarios. However, when faced with the complexities of breast cancer ultrasound images—such as blurred lesion boundaries, indistinct tissue transitions, and severe background noise—the conventional U-Net still exhibits limitations, particularly in terms of accurate localization of critical regions and effective feature focus. These shortcomings often lead to imprecise segmentation or fragmented lesion boundaries.

DOI: 10.54254/2755-2721/2025.22346

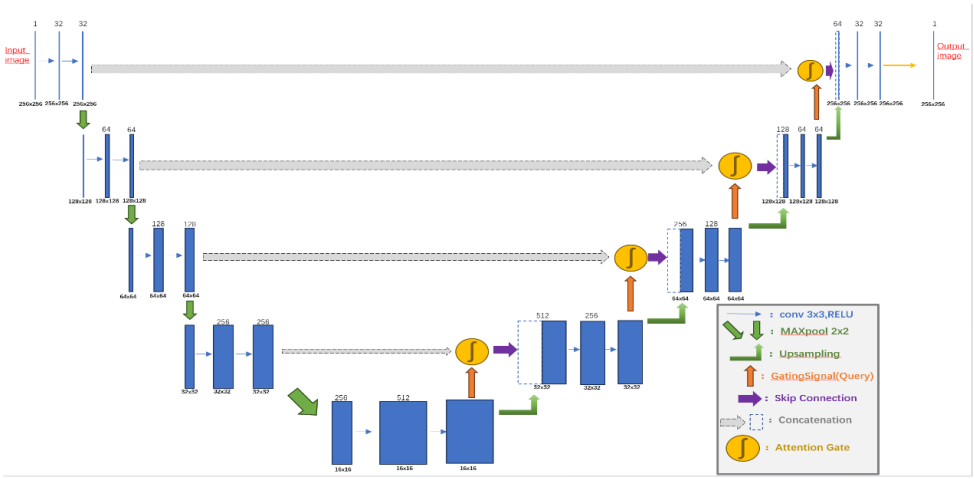

Figure 1: Model Architecture.

To address these issues, this study proposes an improved Attention U-Net model (Figure 1). The model guides the network to automatically focus on discriminative lesion regions while suppressing background interference. By retaining the structural advantages of the original U-Net, it significantly enhances segmentation performance in complex lesion regions commonly found in breast ultrasound images.

The proposed model architecture consists of three core components: the Encoder Block, Decoder Block, and the Attention Gate (AG) mechanism. The encoder adopts a four-level progressive structure, where each level comprises two 3×3 convolutional layers followed by a max-pooling operation. All convolutional layers utilize ReLU activation functions, and Dropout regularization is introduced to improve generalization capability. Normal initialization is employed for weight initialization to stabilize the training process. The max-pooling operation progressively compresses the spatial dimensions of feature maps, thereby expanding the receptive field and enabling the extraction of higher-level semantic features. The output features at each encoder level are not only forwarded to the next layer but also retained for skip connections to be used in the decoder.

The Decoder Block performs progressive up-sampling of the compressed feature maps from the encoder to restore the spatial resolution of the original image. Each decoding layer first applies an up-sampling operation to increase the feature map size and then incorporates the corresponding skip connection features from the encoder. These are fused via feature concatenation, integrating low-level spatial details with high-level semantic representations. Subsequently, two 3×3 convolutional layers are used to further refine the fused features and enhance the representation of segmentation boundaries.

One of the key innovations of this model is the Attention Gate (AG) mechanism. The AG module dynamically computes attention weight maps by combining skip connection features from the encoder with the up-sampled features from the decoder. This process involves several convolutional operations, batch normalization, non-linear activations, and up-sampling. The fused features are normalized using a Sigmoid function to produce an attention map, which is then used to re-weight the skip connection features. This allows the network to adaptively enhance feature responses in lesion areas while suppressing irrelevant background information during training, thus improving both the accuracy and robustness of segmentation in complex breast ultrasound images.

The entire model is trained in an end-to-end fashion. The input consists of preprocessed breast ultrasound images, and the output is a segmentation mask of the same size. A 1×1 convolution layer is employed at the output to generate a single-channel prediction, followed by Sigmoid activation for normalization. The model uses binary cross-entropy loss as the objective function and the Adam

DOI: 10.54254/2755-2721/2025.22346

optimizer, which ensures stable convergence and avoids local minima during training. For performance evaluation, the primary metric is Mean Intersection over Union which quantitatively measures the overlap between the predicted segmentation and the ground truth.

5. Results

To validate the effectiveness of the proposed improved Attention U-Net model in breast ultrasound image segmentation, a set of visualized segmentation results on the test data was presented and analyzed. These visualizations include the Original Mask, Predicted Mask, Processed Mask, and intermediate results in the form of heatmaps. By comparing these images, a more intuitive assessment of the model's practical performance in medical imaging can be achieved.

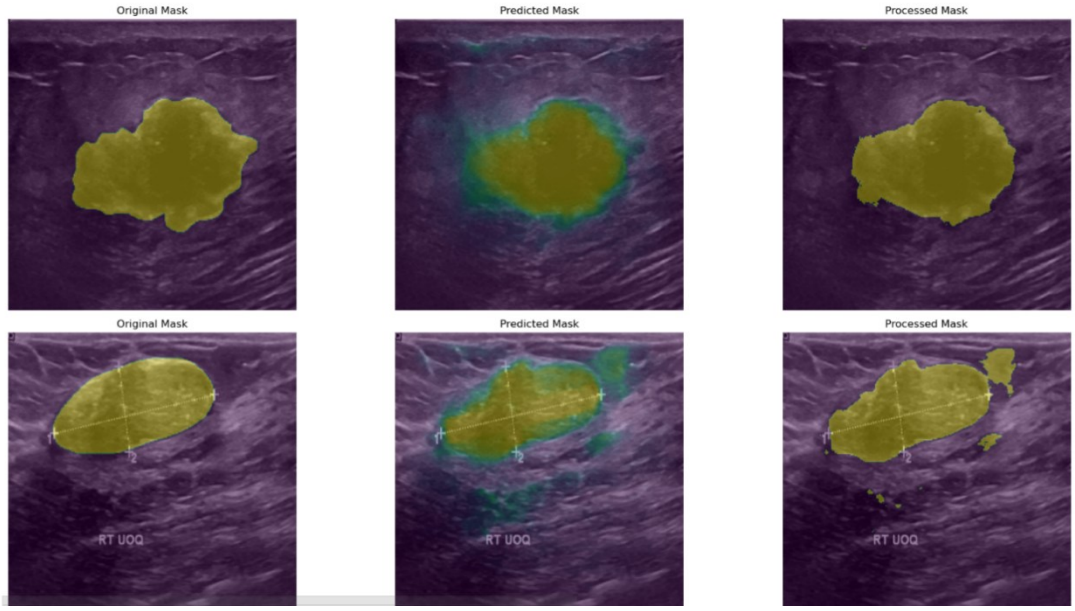

Figure 2: The segmentation results.

As illustrated in Figure 2, from an overall visual perspective, the model demonstrates a strong ability to accurately delineate lesion regions in most cases. The predicted masks exhibit a high degree of spatial alignment with the original masks, indicating that the model possesses robust feature extraction capabilities and can effectively recognize the morphological structure of lesions in breast ultrasound images. This high degree of contour alignment provides a reliable image basis for subsequent quantitative analysis and computer-aided diagnosis.

In terms of lesion localization, the model shows strong spatial awareness. Across different image sets, both the predicted and processed masks align well with the lesion regions annotated in the original masks, suggesting that the model can consistently capture the spatial distribution characteristics of lesions within the images. This localization capability is of critical importance for breast cancer screening and early diagnosis, particularly in clinical applications involving lesion tracking or change analysis based on spatial information.

Further comparison between the predicted and processed masks reveals the positive impact of post-processing on segmentation accuracy. Post-processed masks exhibit smoother boundaries and more regular shapes, which help eliminate noise and edge artifacts present in the raw predictions. For instance, in the second row of Figure 2, the processed mask more accurately conforms to the actual shape outlined in the original mask, resulting in a segmentation output that better meets the clinical demands for boundary clarity and regional accuracy in medical image analysis.

DOI: 10.54254/2755-2721/2025.22346

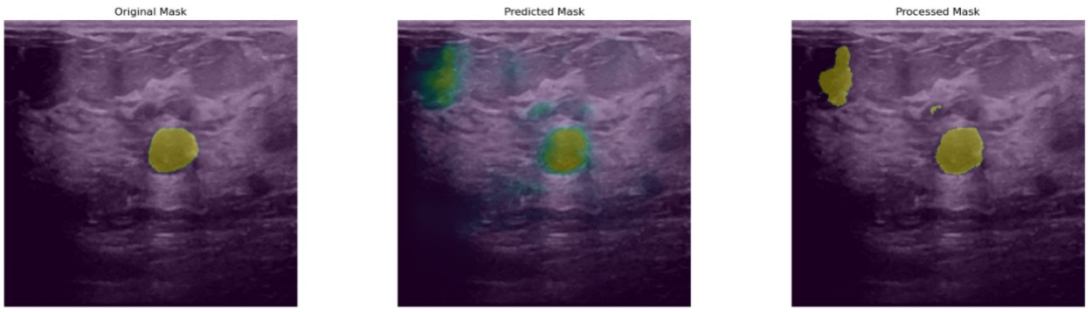

Figure 3: The illustration of poor results.

Despite the encouraging results, several limitations were observed during experimentation, as shown in Figure 3. First, in some images, the boundaries of the processed masks still deviate significantly from those of the original masks, manifesting as inconsistent expansions or contractions along the edges. This indicates that the model's ability to capture fine-grained boundary details remains suboptimal. Such imprecise boundary fitting could hinder clinical tasks that rely on accurate quantification of lesion size and shape, highlighting the need for further enhancement of the model's edge-awareness capabilities.

Second, certain images continue to exhibit misclassified regions and noisy predictions. These erroneous areas often occur in regions with complex shadowing or where tissue echogenicity closely resembles that of the lesions, leading the model to falsely identify background structures as pathological. For example, some predicted masks contain additional green spot-like areas, reflecting the model's misinterpretation of high-echogenicity background tissue as lesion regions. This indicates that when dealing with complex or highly heterogeneous structures, the model still requires improvement in feature discrimination and robustness to minimize false positives and misclassifications.

In summary, the proposed Attention U-Net model demonstrates strong overall performance in breast ultrasound image segmentation, with notable advantages in lesion contour recognition, spatial localization, and feature attention. Nevertheless, future work should focus on optimizing boundary delineation and reducing erroneous predictions, which remain critical challenges for improving clinical applicability.

6. Discussion and Conclusion

This study proposes an improved U-Net model incorporating attention mechanisms, aimed at enhancing the accuracy of automatic lesion segmentation in breast ultrasound images. Building upon the advantages of the original U-Net encoder-decoder structure, the model introduces Attention Gates to dynamically weight the skip connection features. This allows the model to more effectively focus on critical regions within the image while suppressing background interference unrelated to the lesions. Experimental results demonstrate that the proposed model performs well in contour fitting, lesion localization, and spatial perception, providing an effective solution for the automatic analysis of complex breast ultrasound images.

Despite its promising performance, the model still exhibits some limitations in scenarios involving ambiguous tissue boundaries, high structural heterogeneity, or significant image noise. Issues such as inaccurate boundary fitting and false positive predictions may still occur. Furthermore, training the model relies on high-quality, expert-annotated datasets, which incurs considerable data preparation costs. This poses challenges to scalability and generalizability, especially when adapting the model to other diseases or imaging modalities, where large volumes of annotated samples are also required.

DOI: 10.54254/2755-2721/2025.22346

As artificial intelligence continues to advance in medical image processing, the emergence of large-scale pretrained models offers new technical directions for segmentation tasks. Models such as Vision Transformer [5], Swin Transformer [6], and the Segment Anything Model [7] possess powerful capabilities in global modeling and contextual understanding, enabling them to effectively capture complex structural relationships and long-range dependencies within images. In the future, integrating our proposed architecture with these vision foundation models is expected to not only improve segmentation accuracy and semantic understanding but also reduce dependence on manual annotations through the advantages of pretraining, thereby improving training efficiency in low-data scenarios. Moreover, transfer learning will play a vital role in enhancing model generalization [8]. By pretraining on large-scale, general-purpose medical imaging datasets and then fine-tuning on specific disease types or imaging modalities, models can mitigate the issue of data scarcity in target domains. In cross-disease, cross-modality (e.g., transferring from ultrasound to MRI or mammography), or cross-population (e.g., across different genders or ethnicities) tasks, transfer learning provides broader adaptability and accelerates deployment timelines [9]. Multimodal fusion is also expected to be a key future research direction. The diagnosis of breast cancer often depends on the comprehensive assessment of multiple imaging modalities. A single modality is insufficient to fully capture the characteristics of a lesion. By combining ultrasound, mammography, MRI, and even histopathological images, joint modeling of these multi-source datasets can uncover more discriminative pathological representations and enhance the model's performance under multi-dimensional information integration. Such strategies will push the boundaries from simple "image segmentation" toward comprehensive image-based diagnosis. Additionally, the incorporation of self-supervised learning paradigms is a promising path to reduce annotation costs and improve model adaptability [10]. By designing tasks such as contrastive learning, image reconstruction, and context prediction, models can learn structural features in unsupervised or weakly supervised settings, achieving strong downstream performance even with limited labeled data. This is crucial for enhancing the scalability of medical imaging models.

In conclusion, the proposed attention-enhanced U-Net model achieves strong results in breast cancer ultrasound image segmentation and shows considerable potential for practical clinical application. Future work involving integration with foundation models, multimodal fusion, transfer learning, and self-supervised pretraining is expected to further improve the model's precision, generalizability, and applicability—driving the continued advancement and clinical translation of intelligent medical imaging technologies in breast cancer and beyond.

References

[1] *M. Arnold et al., "Current and future burden of breast cancer: Global statistics for 2020 and 2040," The Breast, vol. 66, pp. 15–23, Dec. 2022, doi: 10.1016/j.breast.2022.08.010.*

[2] *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.*

[3] *J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," presente d at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440. Accessed: Mar. 29, 2025. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2015/html/Long_Fully_Convolutional_Networks_2015_CVPR_paper.html*

[4] *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.*

[5] *K. Han et al., "A Survey on Vision Transformer," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 1, pp. 87–110, Jan. 2023, doi: 10.1109/TPAMI.2022.3152247.*

[6] *Z. Liu et al., "Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows," presented at the Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 10012–10022. Acces*

DOI: 10.54254/2755-2721/2025.22346

sed: Mar. 29, 2025. [Online]. Available: https://openaccess.thecvf.com/content/ICCV2021/html/Liu_Swin_Transformer_Hierarchical_Vision_Transformer_Using_Shifted_Windows_ICCV_2021_paper

[7] A. Kirillov et al., "Segment Anything," presented at the Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 4015–4026. Accessed: Mar. 29, 2025. [Online]. Available: https://openaccess.thecvf.com/content/ICCV2023/html/Kirillov_Segment_Anything_ICCV_2023_paper.html

[8] 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.

[9] Z. Li et al., "Cross-modality representation and multi-sample integration of spatially resolved omics data," Jun. 11, 2024, bioRxiv. doi: 10.1101/2024.06.10.598155.

[10] R. Krishnan, P. Rajpurkar, and E. J. Topol, "Self-supervised learning in medicine and healthcare," Nat. Biomed. Eng, vol. 6, no. 12, pp. 1346–1352, Dec. 2022, doi: 10.1038/s41551-022-00914-1.