China and WTO Review (CWR) Journal
2026, (13)
ISSN: 2384-4388
https://cwto.net
Research Article

Visibility-Aware Patch Scheduling for Underwater Object Recognition

Minjun Kim [1], Jihoon Park [1*] Seoyeon Lee [1]

[1] Department of Mechanical Engineering, Korea Advanced Institute of Science and Technology (KAIST), 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea

* **Corresponding Author:** minjun.kim@kaist.ac.kr

ARTICLE INFO

ABSTRACT

Underwater object recognition is affected by light attenuation, turbidity, backscatter, color distortion, and uneven visibility. Efficient visual models are needed for autonomous underwater vehicles, but aggressive patch reduction may remove weak object cues in low-visibility scenes. This study investigates visibility-aware patch scheduling for underwater object recognition. We propose AquaPatch-Net, which estimates local visibility from contrast decay, blue-green color shift, edge attenuation, and suspended-particle noise. Clear background regions are processed through a compact convolutional path, while low-visibility object candidates are retained in a transformer path with enhanced contextual attention. A confidence recovery module further revisits uncertain patches when the initial prediction is unstable. The evaluation combined public underwater image datasets and field data collected during 27 autonomous underwater vehicle dives. The final dataset included 62,400 labeled images, 11 object categories, four turbidity levels, and depth records from 3 to 46 meters. Under clear-water conditions, AquaPatch-Net reduced average computation by 35.2% with no measurable loss in top-1 accuracy. Under high turbidity above 18 NTU, object recall reached 81.7%, compared with 74.9% for a fixed-compression transformer. In field deployment, the model processed 18.5 hours of continuous video and reduced onboard GPU energy consumption by 23.6%. Missed detections of small marine debris decreased from 17.8% to 11.3%, while false alarms from sand ripples and bubbles decreased by 14.6%. The average decision latency remained below 34 ms per frame, meeting real-time navigation requirements. These results show that visibility-aware patch scheduling can support efficient and reliable underwater perception in resource-limited robotic platforms.

Keywords: Underwater Object Recognition, Visibility Estimation, Patch Schedulin, Autonomous Underwater Vehicle, Efficient Vision Model, Real-Time Perception, Robotic Inspection.

INTRODUCTION

Underwater object recognition plays an important role in marine environmental monitoring, seabed exploration, infrastructure inspection, aquaculture management, and autonomous underwater vehicle (AUV) navigation. Reliable recognition allows underwater robots to identify marine organisms, debris, cables, pipelines, structural defects, and other targets without continuous human intervention. However, underwater visual perception remains substantially more difficult than object recognition in terrestrial environments. Light is selectively absorbed as it propagates through water, with longer wavelengths attenuating more rapidly than blue and green components. Scattering caused by suspended particles further degrades image quality, while uneven artificial illumination often produces strong local brightness differences. These effects jointly lead to color distortion, reduced contrast, blurred boundaries, loss of texture, and decreased visibility, particularly in deep, turbid, or particle-rich environments. Existing studies have attempted to alleviate these problems using color correction, contrast enhancement, physical image-formation models, and learning-based restoration techniques [1,2]. Recent vision models have also demonstrated that computational resources do not necessarily need to be distributed uniformly across all image regions. Globally informed token-pruning strategies, for example, can estimate token importance from broader feature relationships and remove redundant visual information while preserving task-relevant representations [3]. This principle is particularly relevant to underwater perception because the information density and visibility of different regions within the same frame can vary considerably. Transformer-based underwater enhancement and recognition methods have further improved long-range feature modeling under degraded imaging conditions [4,5].

Object detection research has also introduced attention mechanisms, feature pyramids, multi-scale fusion, and Transformer blocks to strengthen representations of weak and small underwater targets [6,7]. These methods are useful because many objects of practical interest occupy only a small portion of an underwater image and may have weak contrast relative to the surrounding water. Small debris, distant organisms, partially buried objects, and thin structural components can easily lose discriminative features after repeated downsampling. Multi-scale feature aggregation can retain spatial details across different resolution levels, while attention mechanisms can emphasize informative regions and suppress irrelevant backgrounds [8,9]. Transformer structures provide an additional advantage by modeling wider contextual relationships, which may help distinguish a weak target from visually similar background patterns. Nevertheless, most existing models still process image regions using nearly identical computational procedures. A high-contrast background patch may therefore receive the same amount of attention and feature refinement as a low-visibility region containing a small target [10,11]. Such uniform computation does not fully reflect the highly non-uniform visual conditions of underwater scenes. The computational requirement of underwater recognition is also a practical concern for AUV deployment. Onboard systems generally operate under strict limitations in processing capability, memory, battery capacity, heat dissipation, and mission duration. Lightweight underwater recognition models have therefore been developed using compact convolutional modules, efficient attention mechanisms, GSConv, partial convolution, feature reuse, and reduced detection backbones [12,13]. Transformer-based detectors have also been adapted to improve the recognition of small underwater objects while maintaining manageable computational cost [14,15]. These approaches can significantly reduce model size and floating-point operations, often with only a limited reduction in average detection accuracy. However, their computational structures are usually determined before inference. Network depth, channel width, token number, or compression ratio remains largely fixed regardless of local image quality [16,17]. This limitation is important because underwater visibility rarely changes uniformly across an entire frame [18,19]. One region may contain relatively clear open water, while another may be degraded by backscatter, illumination decay, shadows, suspended particles, or strong color attenuation. Applying the same compression strategy to both regions may discard information that is redundant in clear areas but essential in difficult ones [20]. Aggressive feature reduction is particularly risky for small debris, partly occluded targets, weak edges, and distant objects because their visual evidence is already limited. A computationally efficient underwater model should therefore determine not only how much information can be removed, but also where information can be safely compressed [21,22]. Such a mechanism requires local visibility assessment together with adaptive feature allocation so that easy regions consume fewer resources while ambiguous or degraded regions retain sufficient representational capacity. Another limitation concerns evaluation conditions. Many underwater detection studies are primarily validated using public datasets such as URPC, DUO, and RUOD [23,24]. These datasets provide important benchmarks for comparing recognition methods, but their images cannot fully represent the environmental variability encountered during long-term AUV missions. Real deployments may involve continuous changes in water depth, turbidity, sunlight penetration, artificial illumination, camera orientation, vehicle motion, sediment disturbance, and particle density. The same object can therefore appear very different between dives or even between consecutive video frames. Models that perform well on a fixed benchmark may exhibit reduced recall or increased false detections when these conditions change. Moreover, most published evaluations emphasize mean average precision, precision, recall, or model complexity, whereas onboard inference latency and energy consumption are reported less frequently. For autonomous platforms with limited battery reserves, however, computational efficiency is directly related to mission duration and operational range. These observations indicate the need for an underwater recognition framework that adapts computation to the actual visual difficulty of local image regions rather than relying on globally fixed compression. The key challenge is to reduce redundant computation in clear regions without weakening the features required to recognize difficult targets. This requires a mechanism capable of estimating local visibility, assigning different processing paths according to regional image quality, and recovering patches whose importance is uncertain. It also requires evaluation under both controlled benchmark conditions and real AUV deployments so that improvements in computational efficiency can be assessed together with recognition reliability and energy use.

This study proposes AquaPatch-Net, a visibility-aware patch scheduling framework for underwater object recognition. The model estimates local visual quality using complementary indicators associated with contrast decay, blue-green color shift, edge attenuation, and suspended-particle interference. Patches with relatively clear visual information are assigned to a compact convolutional path to avoid unnecessary computation, whereas low-visibility regions with a higher probability of containing difficult targets are retained in the Transformer path for richer contextual modeling. Patches with ambiguous confidence are further examined through a confidence recovery module to reduce the risk of prematurely discarding weak but informative features. The framework is evaluated using public underwater datasets together with field data obtained from 27 AUV dives, comprising 62,400 labeled images collected at depths ranging from 3 to 46 m. The experiments assess recognition accuracy, object recall, small-debris detection, false-alarm rate, inference time, and onboard GPU energy consumption. The objective is not simply to construct a smaller detector, but to determine whether computation can be allocated according to local underwater visibility while preserving information in visually difficult regions. By linking environmental degradation with adaptive patch-level processing, this study aims to provide a more practical balance between recognition reliability and computational efficiency. The resulting framework may support longer AUV missions, more energy-efficient onboard perception, and more robust underwater recognition under changing visibility conditions, while also providing a general strategy for dynamic visual computation in resource-constrained robotic systems.

MATERIALS AND METHODS

Study Samples and Underwater Data

The study used public underwater image datasets and field data collected during 27 autonomous underwater vehicle dives. After data cleaning and label checking, 62,400 images were included. The dataset covered 11 object classes, four turbidity levels, and water depths from 3 to 46 m. The main objects included marine debris, fish, shells, cables, underwater structures, and other common targets. Field images were collected under clear water, moderate turbidity, high turbidity, uneven lighting, and strong backscatter. Frames with missing labels, serious camera errors, or large blocked regions were removed. Images from the same dive were assigned to only one data subset to avoid overlap between training and testing. Depth and turbidity records were matched with image frames by time stamp.

Experimental Design and Control Models

AquaPatch-Net was compared with a full-patch Transformer, a fixed-compression Transformer, and a lightweight convolutional model. The full-patch Transformer kept all image patches and served as the accuracy control. The fixed-compression Transformer removed the same proportion of patches from each frame. It was used to test whether visibility-based scheduling was better than fixed patch reduction. The convolutional model was used to compare the two-path design with a small single-path network. Ablation tests removed contrast information, color shift, edge strength, particle noise, and the confidence recovery module one at a time. All models used the same training data, input size, optimizer, augmentation settings, and stopping rule. Separate tests were conducted for different turbidity levels and water depths.

Image Measurement and Quality Control

All images were resized to the same input size and normalized before training. Random flipping, small-angle rotation, brightness adjustment, contrast adjustment, and mild blur were used only for the training data. Turbidity was recorded in nephelometric turbidity units and divided into four levels. Water depth was obtained from the AUV pressure sensor. Sensor records were matched with image frames by time stamp. Object labels were checked by two trained reviewers, and unclear samples were reviewed again before analysis. Small marine debris was examined separately because these objects are more difficult to detect under low contrast and high turbidity. Recognition performance was measured by top-1 accuracy, recall, precision, and false-alarm rate. Model cost was measured by computation, GPU energy use, and decision latency. All field tests used the same onboard GPU, input size, and software settings. Warm-up frames were excluded from latency measurement.

Visibility Scoring and Patch Scheduling

For each patch Pi, AquaPatch-Net calculated a visibility score Vi from local contrast, blue-green color shift, edge strength, and suspended-particle noise. Each term was normalized before calculation. The visibility score was defined as:

$$V_i=w_1C_i+w_2B_i+w_3E_i+w_4P_i$$

where Ci is local contrast, Bi is the blue-green color term, Ei is edge strength, and Pi is the suspended-particle term. The parameters w1, w2, w3, and w4 control the weight of each term. A routing score R_i was then calculated from local visibility and prediction uncertainty:

$$R_i=\lambda(1-V_i)+(1-\lambda)U_i$$

where Ui is the prediction uncertainty of patch Pi, and λ controls the contribution of visibility information. Patches with Ri<Tr were sent to the compact convolutional path. Patches with Ri ≥ Tr remained in the Transformer path. Patches close to Tr were checked again by the confidence recovery module. The values of w1, w2, w3, w4, λ, and Tr were selected on the validation set and kept fixed during final testing.

Performance Evaluation and Statistical Analysis

Model performance was tested under clear water, moderate turbidity, high turbidity, and field deployment conditions. The main measures were top-1 accuracy, object recall, small-debris miss rate, false-alarm rate, decision latency, and onboard GPU energy use. Samples with turbidity above 18 NTU were analyzed separately. Continuous AUV video was also used to test performance during long missions. Results from repeated runs were reported as mean ± standard deviation. Paired comparisons were made using the same test frames. The paired t-test was used when the differences were close to a normal distribution. The Wilcoxon signed-rank test was used otherwise. Statistical significance was set at $p<0.05$. Recognition accuracy, energy use, and latency were considered together when comparing the models.

RESULTS AND DISCUSSION

Recognition Accuracy and Patch Reduction

Under clear-water conditions, AquaPatch-Net reduced average computation by 35.2% without a clear loss in top-1 accuracy. This result shows that clear background regions do not need full Transformer processing. The compact convolutional path handled regions with strong contrast and clear edges, while more computation was kept for possible object areas. Fig.1 compares recognition accuracy and computational cost across the tested models. Rani et al. (2026) also reduced underwater detection cost through dynamic Transformer sampling, while Bao et al. (2026) reported a good balance between model size and detection accuracy on embedded hardware. AquaPatch-Net differs from these methods because patch reduction changes with local visibility instead of using one fixed lightweight structure [25,26].

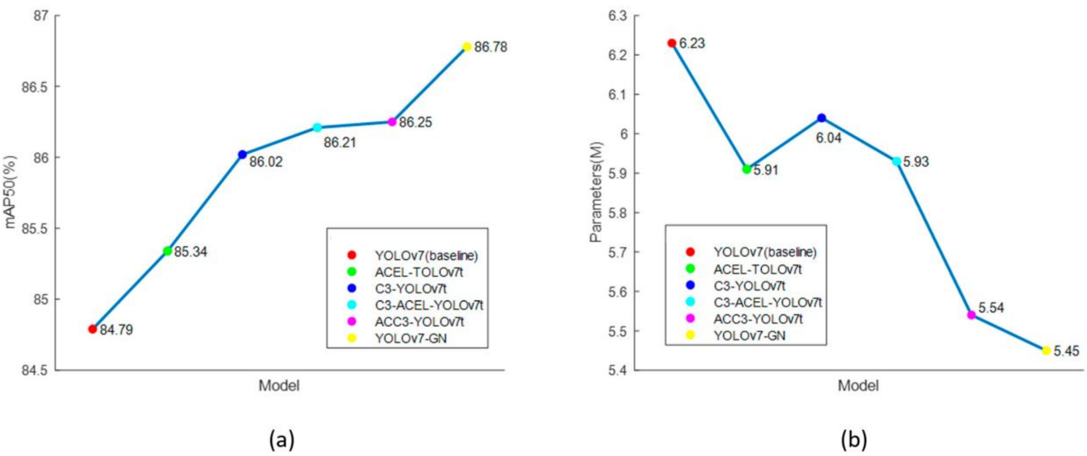

Figure 1. Recognition Performance And Computational Cost Of Aquapatch-Net Under Different Visibility Levels

Performance Under High Turbidity

At turbidity above 18 NTU, AquaPatch-Net reached an object recall of 81.7%, compared with 74.9% for the fixed-compression Transformer. The difference was 6.8 percentage points. This result indicates that fixed patch reduction can remove useful object features when contrast and edge strength are weak. AquaPatch-Net keeps more low-visibility patches in the Transformer path, so nearby context can support recognition when the local signal is poor. Aslam et al. (2025) also found that global context and fine-scale features are important for weak and small underwater objects. Hong et al. (2026) reported similar problems under poor visibility and complex backgrounds. The present results show that patch reduction should change with water conditions rather than use the same ratio in clear and turbid scenes [27,28].

Field Deployment, Energy Use, and Latency

AquaPatch-Net was tested during 18.5 h of continuous AUV video processing. Onboard GPU energy use decreased by 23.6%, while average decision latency remained below 34 ms per frame. The energy reduction was smaller than the drop in computation because total power use also includes memory access, camera input, and data transfer. Even so, the reduction is useful for AUV missions with limited battery capacity. Previous lightweight underwater models have mainly reported parameters, FLOPs, or short-term inference speed. Bao et al. (2026) reported lower model cost through dynamic sampling, while Du et al. (2026) tested lightweight detection on embedded underwater hardware. The present field test shows that patch reduction can also lower energy use during long underwater missions [29,30].

Small Debris Detection and False Alarms

The missed-detection rate for small marine debris decreased from 17.8% to 11.3%, while false alarms from sand ripples and bubbles decreased by 14.6%. As shown in Fig.2, most errors from fixed compression occurred in low-contrast objects and background patterns with similar edges. The confidence recovery module checked uncertain patches again and reduced early recognition errors. Huang et al. (2026) also reported that small underwater objects remain difficult to detect under poor visibility, occlusion, and complex backgrounds. AquaPatch-Net reduces this problem by keeping more processing for uncertain regions instead of increasing computation for every patch. However, the 11.3% miss rate shows that small-object recognition is still not fully solved. Further tests should include stronger turbidity, deeper water, more camera types, and longer AUV missions.

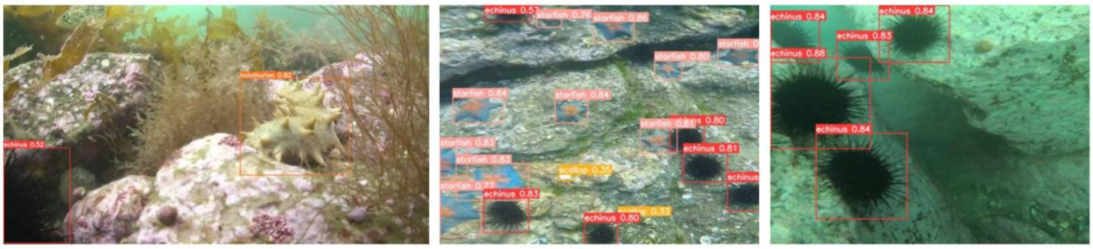

Figure 2. Detection Results for Small Underwater Objects Under Poor Visibility

CONCLUSION

This study developed AquaPatch-Net for underwater object recognition under different visibility conditions. The model controls patch processing using local contrast, color shift, edge strength, particle noise, and prediction uncertainty. Under clear-water conditions, AquaPatch-Net reduced average computation by 35.2% without a clear drop in top-1 accuracy. At turbidity above 18 NTU, object recall reached 81.7%, which was 6.8 percentage points higher than that of the fixed-compression Transformer. During 18.5 h of AUV video processing, onboard GPU energy use decreased by 23.6%, while decision latency remained below 34 ms per frame. The missed-detection rate for small marine debris decreased from 17.8% to 11.3%, and false alarms from sand ripples and bubbles decreased by 14.6%. These results show that patch processing can change with local visibility while keeping useful object information. AquaPatch-Net may be used for real-time underwater inspection and AUV navigation with limited computing power and battery capacity. However, small-object misses still occurred in poor visibility, and the field tests covered a limited range of cameras, depths, and water conditions. Future studies should test the model in deeper water, stronger turbidity, more sensor settings, and longer field missions.

REFERENCES

Xu, D., Chen, H., & Gui, H. (2026, March). Unified Online Estimation Method for SOC, SOH, and Power Capacity Considering Safety Boundary Consistency in Battery Management Systems. In 2026 International Conference on Electrical, Control and Information Technology (ECITech) (pp. 133-137). IEEE.

Rani, K., & Bhandari, A. K. (2026). Advances in Underwater Image Enhancement Models: A Comprehensive Review, Challenges, and Applications. Circuits, Systems, and Signal Processing, 1-102.

Yang, J. (2026). Stage‐Coupled Computational Framework for Stratified Accessibility and Equity Analysis in Community‐Based Elderly Care Services.

Yao, G., Zheng, J., Wang, Z., Zhang, W., Han, R., Zhao, C., ... & Liu, R. (2026, March). V-pruner: A fast and globally-informed token pruning framework for vision transformer. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 40, No. 40, pp. 34396-34404).

Midhunchakkaravarthy, D., RangaNarayana, K., & Narayana, V. L. (2026). Underwater low-light image enhancement using a hybrid CNN–transformer framework with multi-scale feature fusion and cheetah optimization. Discover Computing, 29(1), 61.

Zhang, Z. (2026). A Study on the Impact of Cost Presentation on Decision-making Bias and Cancellation Behavior in Online Subscriptions. Available at SSRN 7349839.

Du, Y., Liu, Q., Dong, Y., & Chen, Z. (2026). Cross-Domain Transfer and Few-Shot Recognition of Defect Images in Complex Industrial Settings.

Liang, S., Hsu, K. H., & Liu, Z. (2026). Rolling Scheduling and Capacity Balancing for Server Assembly Under Engineering Change Disturbances.

Aslam, W., Hussain, J., Aslam, M. Z., Jan, S., Riaz, T. B., Iqbal, A., ... & Khan, I. (2025). Enhanced brain tumor segmentation in medical imaging using multi-modal multi-scale contextual aggregation and attention fusion. Scientific Reports, 15(1), 37308.

Bao, Y., & Wang, H. (2026). Large-scale Metrics Migration: A Systematic Approach to Rebuilding Production Measurement Infrastructure. Available at SSRN 7185660.

Xiong, W., Guo, Y., & Zeng, Y. (2026). A Study on the Energy Efficiency of MEP Systems and Coordinated Dispatch Mechanisms for Multi-energy Systems in High-density Urban Buildings. Available at SSRN 7120518.

Hong, Z., Liang, S., Xu, T., & Chen, H. (2026). A Study on Failover Verification and Recovery Objective Prediction for Cross-Region Cloud Services.

Rashid, M., Wang, J., Ahmed, F., Ahmed, N., Mohsan, S. A. H., Alabdulkreem, E., & Mostafa, S. M. (2026). YOLOv8n-PFA: a parallel fusion attention network for enhanced target detection in challenging environments. Frontiers in Marine Science, 13, 1762170.

14.Huang, J., Xu, T., Yin, J., & Yang, J. (2026). Data Quality Monitoring and Intelligent Scheduling Optimization in Financial Data Workflow Automation. Available at SSRN 7179338.

Jiao, Y., Shi, T., Zhao, B., & Wang, A. (2026). The Impact of VR Sketching Simulations on the Assessment of Scale and Site Suitability in Public Art Spaces.

Liang, S., Du, Y., Chen, W., & Liu, Z. (2026). Order Allocation and Emergency Transportation Decisions Amid Multi-Source Procurement Disruptions.

Gwilliam, M., Zhang, R., Padmanabhan, N., Du, H., & Shrivastava, A. (2026, March). How to design and train your implicit neural representation for video compression. In 2026 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) (pp. 729-739). IEEE.

You, S. (2026). Verifiable Audit Mechanisms in AI Compliance Automation: Scalability. Available at SSRN 6547458.

Yuan, Y., Yin, J., Huang, J., & Xu, T. (2026). Weakly Supervised Anomaly Detection and Privacy Risk Scoring in Community Message Threads.

Bao, Y., & Wang, H. (2026). Edge-to-Cloud Infrastructure Continuum: A Unified Framework for Optimizing Industrial IoT Systems. Available at SSRN 7185578.

Sundaram, M., T, K., Kumar, R. L., & Punitha, P. (2026). A Gradient‐Based Approach for Sustainable Data Compression and Edge Processing in Underwater IoT Networks. International Journal of Communication Systems, 39(6), e70459.

Wang, S., Feng, Y., & Fang, X. (2026, May). A Large Language Model-Enabled Multi-Agent Collaboration Method for Complex Task Solving. In 2026 6th International Symposium on Computer Technology and Information Science (ISCTIS) (pp. 253-256). IEEE.

Yang, Q., & Du, Y. (2026). Predicting Rollback Risks and Controlling Gradual Rollout Traffic for Online Ranking Feature Deployments.

Awad, A., Saleem, A., Aljnadi, Y., Lucas, E., Paheding, S., & Havens, T. C. (2026). RUOD-R: A High-Fidelity Re-Annotated Benchmark for Underwater Object Detection. IEEE Access.

Zhao, Z., & Welsch, R. E. (2026). Point-in-Time Financial RAG with Frozen LLMs and Market-Feedback Adaptive Retrieval. arXiv preprint arXiv:2605.31201.

Yuan, Y., Xu, T., Yin, J., & Huang, J. (2026). Modeling Loading Anomalies and Identifying Root Causes in Media Consumption Workflows on Large Social Media Platforms. Available at SSRN 7393318.

Nurfahanie, A. H., & Prayogo, L. M. (2026). Satellite-Derived Bathymetry in Tropical Shallow Waters: Methods, Sensors, Accuracy, and Applications. JOMAFISH: Journal of Marine and Fisheries Science, 1(1), 62-79.

Zhang, Z., Wang, J., Li, Z., Wang, Y., & Zheng, J. (2025). Anncoder: A mti-agent-based code generation and optimization model. Symmetry, 17(7), 1087.

Ramesh, S., Mann, S., & Stumpf, A. (2025). A comprehensive review of bio-inspired approaches to coordination, communication, and system architecture in underwater swarm robotics. Journal of Marine Science and Engineering, 14(1), 59.

Zhao, J., Fan, J., & Li, L. (2026). A Study on an Explainable Causal-Enhanced LLM Agent for Predicting the Forming Quality of Automotive Component Materials.

ETHICAL DECLARATION

Conflict of interest: No declaration required. **Financing:** No reporting required. **Peer review:** Double anonymous peer review.