ISSN(O): 2320-5407 | ISSN(P): 3107-4928

ISSN NO. 2320-5407

International Journal of Advanced Research

Publisher's Name: Jana Publication and Research LLP

www.journalijar.com

REVIEWER'S REPORT

Manuscript No.: IJAR-59759

Title: Asynchronous, Self-Supervised and Uncertainty-Aware Audio Visual Fusion for Vector Surveillance in Resource-Constrained Environments.

Recommendation:

Accept as it is .....

Accept after minor revision.....

Accept after major revision .....yes.....

Do not accept (*Reasons below*) .....

Rating	Excel.	Good	Fair	Poor
Originality			Y	
Techn. Quality			Y	
Clarity			Y	
Significance			y	

Reviewer's ID: JPR-Dr.Shaweta Sachdeva

Detailed Reviewer's Report

1. The authors should clearly highlight the novelty of the proposed framework compared with existing self-supervised multimodal fusion and uncertainty-aware learning techniques.
2. The manuscript does not provide sufficient information about the datasets used for experimentation. The authors should specify dataset names, sample sizes, mosquito species, class distributions, and training-testing partitions to ensure reproducibility.
3. The proposed framework achieves 92.8% classification accuracy, outperforming the supervised multimodal baseline. The authors should provide stronger experimental evidence and statistical validation to establish the significance of this improvement.
4. The authors should include an ablation study to evaluate the individual contributions of self-supervised learning, asynchronous temporal matching, and uncertainty-aware fusion.
5. The uncertainty estimation mechanism requires further clarification. The authors should explain how uncertainty modules are trained and how their predictions accurately reflect the reliability of acoustic and visual inputs.
6. The manuscript should provide a detailed justification for selecting ResNet-18 and MobileNetV2 as feature extraction architectures instead of other lightweight deep learning models.
7. The temporal matching mechanism uses a predefined time window for identifying positive audio-visual pairs. The authors should specify the selected window size and analyze its impact on classification accuracy and synchronization robustness.
8. The experimental evaluation considers acoustic noise, low-light conditions, and temporal misalignment. Additional experiments under more challenging real-world environmental conditions would strengthen the practical applicability of the proposed approach.
9. The authors should clarify whether the audio and visual recordings represent genuinely corresponding mosquito events. This is particularly important for validating the proposed contrastive learning strategy.
10. The manuscript reports an inference time of 42 ms on a four-core CPU. The authors should provide detailed hardware specifications, memory consumption, model size, and computational complexity to substantiate the suitability of the framework for resource-constrained environments.
11. The comparison with existing approaches should be strengthened by including recent state-of-the-art multimodal learning techniques and presenting a comprehensive performance comparison.
12. The authors should provide additional evaluation metrics, including precision, recall, class-wise F1-score, and confusion matrices, to demonstrate the effectiveness of the proposed approach across different mosquito species.

ISSN(O): 2320-5407 | ISSN(P): 3107-4928
International Journal of Advanced Research
Publisher's Name: Jana Publication and Research LLP
www.journalijar.com

REVIEWER'S REPORT

13. The methodology should clearly describe the percentage of labeled data used during linear probing and explain whether identical training and evaluation conditions were maintained for all baseline models.
14. The authors should investigate how the proposed system performs when one modality becomes completely unavailable, rather than merely degraded.
15. The manuscript should include repeated experimental trials with different random seeds and report standard deviations or confidence intervals to establish the reliability of the findings.

Minor Reviewer Comments

16. The abstract should provide more precise information regarding the datasets, experimental setup, and primary contributions of the proposed framework.
17. The architecture diagram should be improved by clearly distinguishing the training and inference stages and explaining the role of uncertainty estimation in adaptive fusion.
18. The authors should provide a clearer explanation of the mathematical formulations used for contrastive learning and uncertainty-weighted feature fusion.
19. The literature review should be strengthened by critically discussing the limitations of existing audio-visual fusion approaches rather than primarily describing previous studies.
20. The results section would benefit from additional graphical representations illustrating training convergence, comparative classification performance, and robustness under different noise conditions.
21. The discussion section should elaborate on the practical challenges associated with deploying the proposed system using low-cost acoustic and visual sensors.
22. The authors should ensure consistency in mathematical symbols, abbreviations, figure captions, table formatting, and reference presentation throughout the manuscript.
23. The limitations and future scope section should discuss possible extensions involving real-time field deployment, additional environmental sensors, and adaptation to previously unseen mosquito species.
24. The manuscript should undergo careful proofreading to eliminate minor grammatical errors, improve sentence clarity, and maintain a consistent academic writing style.

Overall Comments to the Authors

The manuscript addresses a relevant research problem and proposes a promising framework for intelligent mosquito surveillance using self-supervised multimodal learning and uncertainty-aware fusion. The integration of acoustic and visual information is a potentially valuable contribution, particularly for resource-constrained environments. However, insufficient dataset descriptions, limited implementation details, the absence of ablation studies, and the need for stronger statistical and real-world validation affect the reproducibility and reliability of the findings. The authors should address these issues and strengthen the experimental analysis before the manuscript is considered for publication.

Final Recommendation: Major Revision.