

Peer Review

Review of: "Temporal Feature Weaving for Neonatal Echocardiographic Viewpoint Video Classification"

Linh T. Duong¹¹. Department of Applied Physics, Royal Institute of Technology, Sweden

The manuscript, titled 'Temporal Feature Weaving for Neonatal Echocardiographic Viewpoint Video Classification,' presents significant technical contributions to the field. However, its quality and impact could be further improved by addressing the following points:

- 1. Abstract:** While informative, the abstract needs more specific quantitative results. Including the baseline accuracy alongside the reported 4.33% improvement is crucial for context. Adding key metrics like the F1-score and briefly stating the significance of these results compared to existing methods would strengthen the abstract.
- 2. Introduction:** The introduction establishes a good foundation but could benefit from a more in-depth analysis of prior work's limitations. Specifically, elaborating on why single-frame approaches fall short and how this motivates the proposed method would be beneficial. While the dataset's novelty is mentioned, the introduction should explicitly state how it addresses the shortcomings of current datasets.
- 3. Dataset:** The dataset description, while detailed, needs a stronger justification for excluding classes like "Ductal Cut." Explaining their reduced clinical relevance or increased difficulty in classification would be helpful. Although the class imbalance is acknowledged, the paper lacks a discussion of the strategies used to mitigate its effects during training (e.g., oversampling, class weighting).

4. **Model Architecture:** The explanation of the Temporal Feature Weaving (TFW) method could be simplified for broader audience comprehension. A more intuitive description or visual aid would improve understanding. The choice of ResNet-50 as the backbone requires further justification. A brief comparison with other architectures like EfficientNet or Vision Transformers, highlighting the reasons for selecting ResNet-50, would strengthen the argument.
5. **Experiments:** While training parameters are detailed, the paper omits a discussion of hyperparameter tuning. Explaining how values for the learning rate, dropout, and regularization were selected is important. Adding statistical significance tests (e.g., t-tests, confidence intervals) to the results would validate the reported improvements in accuracy and F1-score.
6. **Results:** The results table would benefit from additional metrics like confusion matrices or per-class performance to identify areas of strength and weakness. The discussion of the results should be expanded. A deeper analysis of why TFW outperforms attention-based methods, perhaps with visualizations of feature importance or attention maps, would be insightful.
7. **Discussion:** The discussion should address potential limitations of the proposed method, such as its reliance on high-quality video data or its performance on unseen datasets. The clinical implications of this work should be explored further. Discussing its potential impact on real-world neonatal echocardiography workflows would be valuable.
8. **Figures and Tables:** All figures, especially Figure 2 (architecture diagram), need more detailed explanations in the text. Figure 1 (class distribution) should clearly highlight the class imbalance and discuss its implications.
9. **Conclusion:** The conclusion should offer specific, actionable directions for future research. Suggesting potential expansions of the dataset or exploring alternative architectures (e.g., Transformers) would be helpful.
10. **References:** The reference section should be updated to include more recent publications in video classification, temporal modeling, and neonatal echocardiography.

11. General Suggestions

- **Language and Grammar:** Ensure the manuscript is free of grammatical errors and awkward phrasing. For example, "key contirbutions are" should be corrected to "key contributions are."

- **Clarity:** Avoid overly technical jargon where possible, or provide brief explanations for terms like "temporal weaving" and "attention heads."
- **Reproducibility:** While the code is mentioned as being available on GitHub, ensure that all details (e.g., preprocessing steps, hyperparameters) are included for full reproducibility.

Declarations

Potential competing interests: No potential competing interests to declare.