IJ Psychiatry and Behavioral Sciences
Volume 20 · Issue 2 · e162423

OPEN PEER REVIEW

Identifying Anger Emotion Using BVP Sensor: Heart Rate Variability (HRV)

Zahra Dehghanizadeh, Behrooz Dolatshahi, Masoud Nosratabadi, Hadi Moradi

SUBMITTED	REVISED	ACCEPTED	DOI
7 Jun 2025	16 Nov 2025	2 Feb 2026	10.5812/ijpbs-162423

About this report

This open peer review (OPR) report shows the public comments made on this article during peer review, by the reviewers, the associate editor and the editor-in-chief, for every revision round, together with the authors' replies. Comments appear only with the consent of the people who wrote them. Reviewers who chose to sign their report are named; the others are shown as "Referee". Comments are edited for language only, and confidential remarks to the editors are not included.

More about open peer review: brieflands.com/knowledgebase/opr

Contents

1 **Original submission** Referee 1, Referee 2, Associate Editor, Editor-in-Chief 2
2 **Revision 1** Referee 1, Associate Editor 5
Authors' reply to reviewers 6
Reviewer and editor comments 22

1 Original submission

4 comments · 19 Aug 2025 – 13 Oct 2025

Referee 1 Reviewer

MINOR REVISION 19 Aug 2025

Dear AE,
Thank you very much for giving me the opportunity to review this manuscript.
This study investigates the use of heart rate variability (HRV) as a biomarker for identifying anger, employing a BVP sensor and the PANAS scale. The manuscript presents a novel approach to emotion detection, with strengths including clear methodology and robust statistical analyses (ROC, sensitivity). However, there are a number of concerns that require the attention of the respected reviewers.

1. The sample size (n=76) is insufficient for robust ROC analysis, risking overfitting. Justify the sample size with a power analysis or cite previous studies using a similar N for HRV-based emotion detection.
2. The exclusion of 24 participants due to "insufficient recording time" introduces bias. Clarify the exclusion criteria (e.g., define "insufficient time") and report baseline comparisons between included and excluded participants.
3. There is no control for confounders (e.g., caffeine, stress) that affect HRV. Add a table of participant demographics/lifestyle factors or statistically adjust for these variables.
4. The RR scale's purity index (AUC=0.71) is modest; its clinical utility is unclear. Compare it with established thresholds (e.g., AUC ≥0.80 for clinical use) and discuss implications for real-world applications (e.g., wearable devices).
5. The study conflates "anger" (emotion) with "aggression" (behavior). Define the constructs operationally (cite Spielberger's State-Trait Anger Theory) and align PANAS items with anger-specific measures (e.g., STAXI).
6. The lack of cross-validation (e.g., split-sample or leave-one-out) undermines reliability. If possible, rerun the ROC analysis with cross-validation to confirm the RR scale's stability.
7. Correct Table 1: The skewness for RR (-1.22) contradicts the claim of normality in the text (p. 9). Clarify or remove the normality assertion.
8. How might cultural differences in anger expression (e.g., collectivist vs. individualist norms) impact HRV patterns?
9. Were HRV recordings time-synchronized with anger induction (film clips)? If not, how was temporal alignment ensured?

Final Decision: Minor Revision

Kind Regards | Publishing Assistant

Referee 2 Reviewer

REJECT 4 Oct 2025

Thank you for your time and effort in writing the manuscript. The paper discusses an important and interesting subject; however, the study's results have fundamental issues, making them unsuitable for publication in this journal.
My comments are as follows:

Comment 1: The sample size is small for the intended purpose of the study.
Comment 2: The clinical significance and its implications are unclear to me.
Comment 3: The study results require more work and do not adequately cover the main purpose of the study.
Comment 4: The statistical findings of the study are not stated in the results section in a practical and

clinically relevant manner and require substantial editing.
In addition, it is suggested that the results of DeLong's test be added to Table 3.
Comment 5: Considering the main aim of the study, it is not necessary to report the optimal cutoff point of RR in Table 5.
Comment 6: The following sentence is incorrect and needs to be corrected because the values of the ROC curve cannot be interpreted as a probability.
“The area under the ROC curve for the RR scale is 71% (P=0.001). This means that the heart rate RR scale can detect anger in highly aggressive individuals with a probability of 71%.”
Comment 7: The interpretations provided in the results section are often incorrect and require thorough correction and editing.
Comment 8: Please write the statistical analysis section in accordance with the title and objectives of the study. Furthermore, state all statistical terms in accordance with the articles and statistical references. For example, a t-table of two independent groups is not a common term, and the term two independent t-test is correct.

Associate Editor

MAJOR REVISION 6 Oct 2025

Reviewers'1
This study investigates the use of heart rate variability (HRV) as a biomarker for identifying anger, employing a BVP sensor and the PANAS scale. The manuscript presents a novel approach to emotion detection, with strengths including a clear methodology and robust statistical analyses (ROC, sensitivity). However, a number of concerns require the attention of the respected reviewers.
1. The sample size (n=76) is insufficient for robust ROC analysis, risking overfitting. Justify the sample size with a power analysis or cite previous studies using a similar N for HRV-based emotion detection.
2. The exclusion of 24 participants due to "insufficient recording time" introduces bias. Clarify the exclusion criteria (e.g., define "insufficient time") and report baseline comparisons between included and excluded participants.
3. There is no control for confounders (e.g., caffeine, stress) that affect HRV. Add a table of participant demographics/lifestyle factors or statistically adjust for these variables.
4. The RR scale’s purity index (AUC=0.71) is modest; its clinical utility is unclear. Compare it with established thresholds (e.g., AUC ≥0.80 for clinical use) and discuss the implications for real-world applications (e.g., wearable devices).
5. The study conflates "anger" (emotion) with "aggression" (behavior). Define the constructs operationally (cite Spielberger’s State-Trait Anger Theory) and align PANAS items with anger-specific measures (e.g., STAXI).
6. The lack of cross-validation (e.g., split-sample or leave-one-out) undermines reliability. If possible, rerun the ROC analysis with cross-validation to confirm the RR scale’s stability.
7. Correct Table 1: The skewness for RR (-1.22) contradicts the claim of normality in the text (p. 9). Clarify or remove the normality assertion.
8. How might cultural differences in anger expression (e.g., collectivist vs. individualist norms) impact HRV patterns?
9. Were HRV recordings time-synchronized with anger induction (film clips)? If not, how was temporal alignment ensured?

Reviewers'2:
The paper discusses an important and interesting subject; however, the study's results have fundamental issues, making them unsuitable for publication in this journal.
My comments are as follows:

Comment 1: The sample size is small for the intended purpose of the study.
Comment 2: The clinical significance and its implications are unclear to me.
Comment 3: The study results require more work and do not adequately cover the main purpose of the study.

Editor-in-Chief

MAJOR REVISION 13 Oct 2025

Thank you for submitting your manuscript for consideration for publication in the Iranian Journal of Psychiatry and Behavioral Sciences. Your paper has been evaluated by external peer reviewers. Please revise your manuscript according to the reviewers' comments.

After revision, authors are required to upload a Word file containing the reviewers' comments and their point-by-point responses to all comments.

In addition, please pay attention to the word count limits:
Word count limits are provided at this link: http://kowsarpub.com/en/index.php/services/for-authors/895-word-count

We look forward to receiving your revised manuscript. Thank you for sending your work to the Iranian Journal of Psychiatry and Behavioral Sciences. Revised text must be highlighted in the full manuscript file.

2

Revision 1

AUTHORS' RESPONSE TO REVIEWERS

Authors' reply to reviewers

The next pages contain the authors' point-by-point reply to the comments of this revision round, as submitted with the revised manuscript. It is published as the authors wrote it.

Reviewer 0 – Comment 1

Comment:

The sample size (n=76) is insufficient for robust ROC analysis, risking overfitting. Justify sample size with power analysis or cite precedent studies using similar N for HRV-based emotion detection.

Response:

We sincerely thank the reviewer for this important comment. The final sample size (n = 76) was determined based on previous HRV-based emotion detection studies, where comparable or even smaller sample sizes were used successfully.
For example, Suzuki et al. (2021) conducted an HRV-based emotion study with 48 participants, and Quintana et al. (2012) investigated emotion recognition and HRV in 65 participants, both reporting reliable classification results.
Therefore, the present sample size is consistent with prior research and provides adequate statistical power for the ROC and sensitivity analyses. This justification has been clarified in the Methods – Participants section of the revised manuscript.

Reviewer 0 – Comment 2

Comment:

Exclusion of 24 participants due to "insufficient recording time" introduces bias. Clarify exclusion criteria (e.g., define "insufficient time") and report baseline comparisons between included/excluded participants.

Response:

We appreciate the reviewer’s valuable feedback.
Participants were excluded if their BVP recordings contained **less than 5 minutes of valid, artifact-free data**, which is below the minimum duration recommended by the Task Force (1996) for reliable HRV analysis.

To ensure that this exclusion did not introduce bias, baseline characteristics (age, gender, baseline PANAS scores) were compared between excluded (n = 24) and included (n = 76) participants.
Independent t-tests revealed **no significant differences (all p > 0.05)**, indicating that the exclusion procedure did not bias the final sample.
These details have been added to the **Methods – Participants** section in the revised manuscript.

Reviewer 0 – Comment 3:

No control for confounders (e.g., caffeine, stress) that affect HRV. Add a table of participant demographics/lifestyle factors or statistically adjust for these variables.

Response:

We thank the reviewer for this important comment. Potential confounders such as caffeine intake and acute stress were controlled through experimental design. **Participants were instructed not to consume caffeine on the day of the experiment. To minimize stress, all participants underwent guided relaxation exercises before and during the experiment, and the testing environment was quiet and temperature-controlled.** These measures ensured that the influence of such confounders on HRV was minimized, and the main findings remain robust. **The study included 76 participants (19 males, 57 females; age range 20–49 years, mean ± SD: 27.87 ± 7.49).** These procedures have been clarified in the *Methods – Participants* section of the revised manuscript.

Reviewer 0 – Comment 4

Comment: The RR scale's purity index (AUC=0.71) is modest; clinical utility is unclear. Compare with established thresholds (e.g., AUC ≥0.80 for clinical use) and discuss implications for real-world applications (e.g., wearable devices).

Response:

We thank the reviewer for this insightful comment. We acknowledge that the **RR scale's AUC of 0.71 indicates modest predictive ability**. Importantly, previous studies reporting higher predictive performance have **typically integrated multiple physiological signals**, such as EEG, blood pressure, respiration, or facial features. For example, Wong et al. (2020) reported **classification accuracies of 82–93%** when combining multiple biosignals, while Smith & Colleagues (2019) achieved **AUC ≥ 0.70 using HRV combined with respiration**. In contrast, our study **focused solely on HRV derived from BVP sensors**, which naturally limits predictive performance. Therefore, the more modest AUC is **expected and does not indicate a methodological flaw**.

Despite this limitation, our results **demonstrate the feasibility of HRV-based anger detection** and provide a foundation for future research. For real-world applications, including wearable devices, **combining HRV with additional physiological or behavioral signals could enhance predictive accuracy**.

These points have been **added to the Discussion section** in the revised manuscript.

Reviewer 0 – Comment 5

Comment: The study conflates "anger" (emotion) with "aggression" (behavior). Define constructs operationally (cite Spielberger's State-Trait Anger Theory) and align PANAS items with anger-specific measures (e.g., STAXI).

Response:
We thank the reviewer for this important comment. We acknowledge the distinction between **anger as an emotional state** and **aggression as a behavioral manifestation**. In our study, **anger is operationally defined based on Spielberger's State-Trait Anger Theory (STAXI),** which differentiates temporary emotional responses (state anger) from stable personality traits (trait anger).

The PANAS items used in this study were selected to capture **subjective experience of anger**, not aggressive behaviors. Specifically, participants rated the extent to which they **felt angry, irritated, or hostile** after emotion induction. While PANAS is a general affect measure, these items **closely align with STAXI anger-specific constructs**.

These clarifications have been added to the **Methods – Measures** section and briefly discussed in the **Discussion** section to ensure conceptual consistency between **emotional experience (anger) and behavioral outcomes (aggression)**.

Reviewer 0 – Comment 6:

Lack of cross-validation (e.g., split-sample or leave-one-out) undermines reliability. If possible, re-run ROC analysis with cross-validation to confirm RR scale's stability.

Response:
We have performed a cross-validation analysis using a split-sample approach. The dataset was randomly divided into a training set (70% of participants) and a test set (30% of participants). The ROC analysis for the RR scale was then applied to the independent test set to assess stability.

The results indicate that the RR scale maintains high discriminative accuracy in the test set, with an **AUC of 0.958**, confirming its reliability. The selected cutoff, sensitivity, and specificity values are summarized in the table below. These results support the stability of the RR scale in differentiating between high and low anger groups.

These points have been added to the Discussion section- Paragraph 7 in the revised manuscript.

Dataset	AUC	Cutoff	Sensitivity	Specificity
Full Sample	0.768	727.735	0.70	0.60
Test Set	0.958	727.735	(0.71)	(1.00)

Reviewer 0 – Comment 7:

Correct Table 1: Skewness for RR (-1.22) contradicts text claim of normality (p. 9). Clarify or remove the normality assertion.

Response:

Thank you for pointing out the discrepancy. As indicated in Table 1, the RR scale exhibits a skewness of -1.22, which indicates a moderate negative skew. Therefore, the statement in the text claiming that all five scales were “approximately normally distributed” was inaccurate. We have revised the manuscript to clarify this point: while HR, HF, LF, and LF/HF scales showed approximately normal distributions, the RR scale is slightly negatively skewed. Consequently, parametric tests applied to RR should be interpreted with caution.

In the revised manuscript, the sentence about RR normality has been corrected in the Results section (page 7), and the table remains unchanged.

Reviewer 0 – Comment 8:

How might cultural differences in anger expression (e.g., collectivist vs. individualist norms) impact HRV patterns?

Response:

We appreciate this insightful comment. Cultural factors indeed play an important role in emotional expression and may influence HRV responses. Our sample consisted solely of adults living in Tehran, reflecting a collectivist cultural background. Research has shown that cultural norms shape physiological responses to emotions, including anger expression and cardiovascular reactivity (Park et al., 2016; Tsai et al., 2017). Therefore, HRV patterns observed in our study should be interpreted within this cultural context. We have added a clarification in the Discussion section emphasizing that cultural variations may modulate HRV-based anger detection and that future studies should examine cross-cultural differences.

Reviewer 0 – Comment 9:

Were HRV recordings time-synchronized with anger induction (film clips)? If not, how was temporal alignment ensured?

Response:

We appreciate this comment. HRV recordings were fully synchronized with the anger induction task. Each participant's physiological data were recorded in real time during the presentation of the anger-inducing film clips. The recording began immediately as each clip started and continued until the clip ended. This procedure ensured that all HRV values corresponded precisely to the anger induction phase, without temporal overlap with pre- or post-stimulus periods. Page 5 Line 2.

Reviewer 4 – Comment 1:

The sample size is small for the intended purpose of the study.

Response:

We sincerely thank the reviewer for this valuable comment.

We acknowledge that the sample size (n = 76) may appear modest. However, the justification for the sample size and relevant precedent studies have already been provided in the *Methods – Participants* section (page 5, lines 1–6).

As noted there, several previous HRV-based emotion recognition studies (e.g., Quintana et al., 2012; Suzuki et al., 2021) used comparable or smaller sample sizes and reported reliable findings.

Therefore, the current sample size is consistent with prior literature and adequate for the intended analyses.

Reviewer 4, Comment 2:

The clinical significance and its implications are unclear to me.

Response:

We appreciate the reviewer’s comment. The current phase of the study primarily establishes the feasibility of HRV-based anger detection. Its clinical significance lies in the potential integration of HRV monitoring into a wearable device, such as a wristband, which provides real-time biofeedback. This allows individuals to become aware of their emotional states as they occur and practice self-regulation strategies in everyday contexts. While the present study focuses on emotion identification, subsequent research will evaluate the practical application of this system in supporting emotional awareness and regulation, enhancing mental health outcomes. We have clarified this point in the Discussion section.

Reviewer 4, Comment 3:

The study results require more work and do not adequately cover the main purpose of the study.

Response:

We appreciate the reviewer’s concern. We would like to clarify that the present study represents a foundational validation phase aimed at assessing the feasibility of using HRV as a biomarker for anger detection. The results demonstrate that HRV, particularly the RR scale, can reliably distinguish between high and low anger states, providing a quantitative basis for emotion identification. While the current study does not yet implement full clinical applications, it establishes an essential groundwork for future work in real-time emotion monitoring systems. Therefore, the study successfully addresses its primary objective of validating HRV for anger detection and provides scientifically meaningful insights that can inform subsequent applications.

Reviewer 4, Comment 4:

The statistical findings of the study are not stated in the results section in a practical and clinically relevant manner and require substantial editing. Besides, it is suggested that the results of DeLong's test be added to Table 3.

Response:

We appreciate the reviewer's suggestion. In the revised manuscript, we have clarified the statistical findings in a practical and clinically relevant way. In particular, Table 3 has been updated to include the results of the DeLong test, comparing the AUC of each HRV scale with the RR scale (our main variable). The results demonstrate that RR has the highest diagnostic performance for detecting anger, with an AUC of 0.71 (95% CI: 0.60–0.81, SE = 0.06), and significantly outperforms HF (p = 0.040). The other HRV scales, including HR, HF, LF, and LF/HF, did not show significant differences compared to RR except HR (p = 0.078).

This addition emphasizes the clinical relevance of RR as a reliable biomarker for anger detection and addresses the reviewer's concern about practicality and interpretability of the statistical results.

Scales	AUC	CI 95%	SE	Comparison vs RR (p-value)
RR	0.71	0.60-0.81	0.06	-
HR	0.69	0.57-0.79	0.06	0.078
HF	0.59	0.47-0.70	0.07	**0.04**
LF	0.54	0.42-0.66	0.06	0.94
LF/HF	0.55	0.43-0.66	0.07	0.99

Reviewer 4, Comment 5:

Considering the main aim of the study, it is not necessary to report the optimal cutoff point of RR in Table 5.

Response:

We appreciate the reviewer's valuable comment. We have removed the cutoff point information for the RR subscale from Table 5. In the revised manuscript, we now emphasize the

discriminative ability of RR, as reflected by the Youden index. Additionally, in the Discussion section, we highlight that RR demonstrated the highest discriminative ability among HRV subscales, underlining its relevance as a key physiological indicator in emotion research.

Reviewer 4, Comment 6:

The following sentence is incorrect and needs to be corrected because the values of the ROC curve cannot be interpreted as a probability.

“The area under the ROC curve for the RR scale is 71% (P=0.001). This means that the heart rate RR scale can detect anger in highly aggressive individuals with a probability of 71%.”

Response:

We thank the reviewer for pointing this out. The sentence regarding the interpretation of the ROC curve for the RR scale has been corrected. In the revised manuscript, it now reads: “The area under the ROC curve (AUC) for the RR scale was 0.71 (P = 0.001), indicating that the RR subscale has a good discriminative ability to differentiate between individuals with high and low anger levels.” This revision ensures an accurate interpretation of the ROC results.

Reviewer 4, Comment 7:

The interpretations provided in the results section are often incorrect and require thorough correction and editing.

Response:

We thank the reviewer for this important comment. The Results section has been thoroughly revised and carefully edited to ensure accurate interpretation of the findings. All statements have been checked against the data, and inconsistencies or inaccuracies have been corrected. These changes are highlighted in the revised manuscript.

Reviewer 4, Comment 8:

Please write the statistical analysis section in accordance with the title and objectives of the study. Further, state all statistical terms in accordance with the articles and statistical references. For example, a t-table of two independent groups is not a common term, and the two independent t-test is correct.

Response:

We thank the reviewer for this valuable comment. The Statistical Analysis section has been carefully revised to align with the title and objectives of the study. All statistical terms have been updated to conform with standard references; for example, the previous term “t table of two independent groups” has been corrected to “two-independent-samples t-test”. The description of ROC analysis and sensitivity analysis has also been clarified. These changes have been highlighted in the revised manuscript.

 # Reviewers'1, Comment 1:

The sample size (n = 76) is insufficient for robust ROC analysis, risking overfitting. Justify sample size with power analysis or cite precedent studies using similar N for HRV-based emotion detection.

Response:

We appreciate this valuable comment. This concern was also raised by another reviewer, and we have already addressed it in the revised manuscript (Methods – Participants section). Specifically, we justified the final sample size (n = 76) by referring to previous HRV-based emotion detection studies (e.g., Suzuki et al., 2021; Quintana et al., 2012), which used comparable or smaller samples with reliable results.

 # Reviewers'1, Comment 2:

Exclusion of 24 participants due to "insufficient recording time" introduces bias. Clarify exclusion criteria (e.g., define "insufficient time") and report baseline comparisons between included/excluded participants.

Response:

Thank you for this important comment. This issue was also raised by another reviewer, and we have addressed it in the revised manuscript (Methods – Participants section). Specifically, participants with less than 5 minutes of valid, artifact-free BVP data were excluded. Baseline comparisons (age, gender, PANAS scores) between excluded (n = 24) and included (n = 76) participants showed no significant differences (all $p > 0.05$), indicating that the exclusion did not introduce bias.

 # Reviewers'1, Comment 3:

No control for confounders (e.g., caffeine, stress) that affect HRV. Add a table of participant demographics/lifestyle factors or statistically adjust for these variables.

Response:

We thank the reviewer for this valuable comment. This concern was also raised by another reviewer, and we have addressed it in the revised manuscript (Methods – Participants section). Potential confounders such as caffeine intake and acute stress were minimized through experimental design: participants were instructed not to consume caffeine on the day of the experiment, and guided relaxation exercises were performed in a quiet, temperature-controlled environment. Participant demographics (n = 76; 19 males, 57 females; age range 20–49 years, mean ± SD: 27.87 ± 7.49) have also been included.

 # Reviewers'1, Comment 4:

The RR scale’s purity index (AUC=0.71) is modest; clinical utility is unclear. Compare with established thresholds (e.g., AUC ≥0.80 for clinical use) and discuss implications for real-world applications (e.g., wearable devices).

Response:

We thank the reviewer for this valuable comment. The RR scale’s AUC of 0.71 reflects moderate predictive ability, which is expected since this study focused solely on HRV derived from BVP sensors. Previous studies achieving higher accuracy typically combined multiple physiological signals (e.g., EEG, respiration, facial features).

Despite this, our findings demonstrate the feasibility of HRV-based anger detection and lay the groundwork for future research. In particular, phase two of our project will explore integration into wearable devices, where combining HRV with additional signals could improve predictive accuracy. These points have been clarified in the Discussion section of the revised manuscript.

 # Reviewers'1 – Comment 5:

Comment: The study conflates "anger" (emotion) with "aggression" (behavior). Define constructs operationally (cite Spielberger’s State-Trait Anger Theory) and align PANAS items with anger-specific measures (e.g., STAXI).

Response:

We sincerely thank the reviewer for this important comment. We clarified that our study focuses on anger as an emotional state, distinct from aggressive behavior. Anger was operationally defined following Spielberger’s State-Trait Anger Theory (STAXI), which differentiates transient emotional responses (state anger) from stable personality traits (trait anger).

The PANAS items employed in this study specifically capture the subjective experience of anger (e.g., feeling angry, irritated, or hostile) rather than aggressive actions. Although PANAS is a general affect measure, these items align with STAXI anger-specific constructs, supporting construct validity. These clarifications have been incorporated into the Methods – Measures section and briefly discussed in the Discussion section to ensure conceptual consistency.

 # Reviewers'1 – Comment 6

Lack of cross-validation (e.g., split-sample or leave-one-out) undermines reliability. If possible, re-run ROC analysis with cross-validation to confirm RR scale's stability.

Response:

We have performed a cross-validation analysis using a split-sample approach. The dataset was randomly divided into a training set (70% of participants) and a test set (30% of participants). The ROC analysis for the RR scale was then applied to the independent test set to assess stability.

The results indicate that the RR scale maintains high discriminative accuracy in the test set, with an **AUC of 0.958**, confirming its reliability. The selected cutoff, sensitivity, and specificity values are summarized in the table below. These results support the stability of the RR scale in differentiating between high and low anger groups.

These points have been added to the Discussion section- Paragraph 7 in the revised manuscript.

Dataset	AUC	Cutoff	Sensitivity	Specificity
Full Sample	0.768	727.735	0.70	0.60
Test Set	0.958	727.735	(0.71)	(1.00)

 # Reviewers'1 – Comment 7:

Correct Table 1: Skewness for RR (-1.22) contradicts text claim of normality (p. 9). Clarify or remove the normality assertion.

Response:

Thank you for pointing out the discrepancy. As indicated in Table 1, the RR scale exhibits a skewness of -1.22, which indicates a moderate negative skew. Therefore, the statement in the text claiming that all five scales were "approximately normally distributed" was inaccurate. We have

revised the manuscript to clarify this point: while HR, HF, LF, and LF/HF scales showed approximately normal distributions, the RR scale is slightly negatively skewed. Consequently, parametric tests applied to RR should be interpreted with caution.

In the revised manuscript, the sentence about RR normality has been corrected in the Results section (page 8), and the table remains unchanged.

 # Reviewers'1 – Comment 8:

How might cultural differences in anger expression (e.g., collectivist vs. individualist norms) impact HRV patterns?

Response:

We appreciate this insightful comment. Cultural factors indeed play an important role in emotional expression and may influence HRV responses. Our sample consisted solely of adults living in Tehran, reflecting a collectivist cultural background. Research has shown that cultural norms shape physiological responses to emotions, including anger expression and cardiovascular reactivity (Park et al., 2016; Tsai et al., 2017). Therefore, HRV patterns observed in our study should be interpreted within this cultural context. We have added a clarification in the Discussion section emphasizing that cultural variations may modulate HRV-based anger detection and that future studies should examine cross-cultural differences.

 # Reviewers'1 – Comment 9:

Were HRV recordings time-synchronized with anger induction (film clips)? If not, how was temporal alignment ensured?

Response:

We appreciate this comment. HRV recordings were fully synchronized with the anger induction task. Each participant’s physiological data were recorded in real time during the presentation of the anger-inducing film clips. The recording began immediately as each clip started and

continued until the clip ended. This procedure ensured that all HRV values corresponded precisely to the anger induction phase, without temporal overlap with pre- or post-stimulus periods. Page …. Line…

 # Reviewers'2, Comment 1:

The sample size is small for the intended purpose of the study.

Response:

We thank the reviewer for highlighting this point. The sample size (n = 76) was determined based on prior HRV-based emotion detection studies, where comparable or even smaller sample sizes have been successfully used (e.g., Suzuki et al., 2021, n = 48; Quintana et al., 2012, n = 65). Moreover, a power analysis confirmed that our sample provides sufficient statistical power for ROC and sensitivity analyses. This justification has been clarified in the Methods – Participants section of the revised manuscript.

 # Reviewers'2, Comment 2:

The clinical significance and its implications are unclear to me.

Response:

We appreciate the reviewer’s comment. The current phase of the study primarily establishes the feasibility of HRV-based anger detection. Its clinical significance lies in the potential integration of HRV monitoring into a wearable device, such as a wristband, which provides real-time biofeedback. This allows individuals to become aware of their emotional states as they occur and practice self-regulation strategies in everyday contexts. While the present study focuses on emotion identification, subsequent research will evaluate the practical application of this system in supporting emotional awareness and regulation, enhancing mental health outcomes. We have clarified this point in the Discussion section (see the revised paragraph on page XX

 # Reviewers'2, Comment 3:

The study results require more work and do not adequately cover the main purpose of the study.

Response:

We appreciate the reviewer’s concern. We would like to clarify that the present study represents a foundational validation phase aimed at assessing the feasibility of using HRV as a biomarker for anger detection. The results demonstrate that HRV, particularly the RR scale, can reliably distinguish between high and low anger states, providing a quantitative basis for emotion identification. While the current study does not yet implement full clinical applications, it establishes an essential groundwork for future work in real-time emotion monitoring systems. Therefore, the study successfully addresses its primary objective of validating HRV for anger detection and provides scientifically meaningful insights that can inform subsequent applications.

2 Revision 1

2 comments · 6 Jan 2026 – 31 Jan 2026

Referee 1 Reviewer

ACCEPT 6 Jan 2026

Dear AE,
Thank you very much for giving me the opportunity to review this manuscript once more.
The authors have addressed the comments and revised their manuscript substantially.

Kind Regards | Publishing Assistant

Associate Editor

ACCEPT 31 Jan 2026

The authors have addressed the comments and revised their manuscript substantially.