Dear Editor, Dear Reviewers,

We sincerely want to thank you for you deep, detailed and insightful feedback concerning our manuscript titled "Anomaly Detection in Multi-Wavelength Photoplethysmography using Lightweight Machine Learning Algorithms". The comments are all valuable and constructive for revising and improving our paper.

We have addressed all comments by modifying the manuscript and the models accordingly. The corresponding response and adaptations are individually described in more depth below. Finally, the modifications are highlighted in the manuscript.

Yours sincerely,

Vlad-Eusebiu, Joan, Angel, Juan Carlos, Johan and Bruno

Reviewer 3

The study focus is to test effect of temporal window sizes, signal feature sets and different ML algorithms (DT, RF, SVM, and RCA) on accuracy and F1 of the anomaly detection in MW-PPG signals. The study is interesting and has applications especially in wearables where light weight and optimized algorithms are needed for PPG analysis.

The authors identified the optimal signal feature sets, window sizes for different algorithms and demonstrated that simultaneous analysis of multichannel signal outperforms single channel signal analysis.

The article is well written and includes all needed information to understand the study and its results. However, the article is a bit lengthy therefore I'd suggest to reduce it by moving the non-necessary content to Appendix or removing it completely. There are many repetitions, some figures are not needed, and some well known algorithms or approaches are unnecessarily explained in details. An interested reader without proper background would easily find needed information in the literature.

1. **P.1 par2: "other valuable information resides in the PPG pulse shape" Please write which additional info can be extracted from PPG signals (e.g. respiration rate)**

   Response: Thank you for your valuable feedback. We appreciate your suggestion to clarify the additional information that can be extracted from PPG signals. In addition to heart rate, respiration rate, and SpO2, blood pressure or blood glucose can be estimated. We included this information in the revised version of the manuscript to provide a more comprehensive understanding of the valuable information that can be derived from PPG signals.

   Modifications: We explicitly mentioned in Section 1 the respiration rate and other two markers that are a hot research topic nowadays.

2. **The descriptions of PPG artifacts, ML algorithms and evaluation metrices are very detailed, which is an advantage for a newcomer in the field. While they are unnecessary for experienced reader. I'd suggest the authors to reduce the section to the minimum necessary information, while the extensive descriptions can be moved to Appendix.**

   Response: Thank you for your feedback. We agree that providing a concise and focused presentation of the essential information is important. Since another reviewer had the same comment, we decided to comply with his suggestion, and we reduced the size of the manuscript to 17 pages.

   Modifications: We have successfully condensed the article to 17 pages by reevaluating its structure and focusing on highlighting the most salient contributions and striking results of our research. We included the rest of the content that we considered necessary in Appendix section.

3. **P.8 par. 4: RCA – write full name, "robust collaborative autoencoder"**

   Response: Thank you for your remark. Since we trained the autoencoder only on clean PPG windows, we decided to drop the RCA abbreviation. Both autoencoders and RCA are based on the same basic principle of encoding and decoding, but RCA introduces additional mechanisms to handle noise and corruption in the input data. In response to the confusion between RCA and the autoencoder, we apologize for any inaccuracies in our previous statement. We must clarify that we unintentionally mixed up the concepts.

   Modifications: We decided to drop the RCA abbreviation for the autoencoder.

4. **P9. Par. 3: Provide information about models and manufacturers, actual distances between LEDs and LDs, signal acquisition.**

   Response: Thank you for your feedback. This information will help to provide a more comprehensive understanding of the experimental setup and data collection procedures.

   Modifications: In Section 4.1, we will provide additional information regarding the models and manufacturers of the components used, the actual distances between LEDs and PDs, as well as details about the signal acquisition process.

5. **P9. Par. 4: "recorded data measured from two subjects" – Provide more information about the subjects (skin color, age, gender). Including only two subjects does not cover population variation, therefore the study results should be discussed considering the small sample size. A statement about the medical ethics committee approval is missing, since the study was performed on human subjects.**

   Response: Thank you for your comment.

   Due to the small-scale experiment, and since we use de-identified data that target anomaly detection that is not subject-dependent, we consider that experiments do not pose potential risks or ethical considerations. We adhere to ethical principles and obtain informed consent from the participants (one of them is also the author).

   Modifications: TBA

6. **P.9 par. 4: Please provide information how the artifacts were introduced to the signal.**

   Response: Thank you for your valuable comment. In our experimental setup, we artificially introduced different types of artifacts by manipulating specific parameters during the data acquisition process. For instance, we introduced motion artifacts by intentionally moving the finger during signal recording. We also introduced contact force artifacts by applying varying levels of pressure to the sensor.

   Modifications: TBA

7. **P. 15: "Peak variance" calculation – How noise effects the peak location determination?**

   Response: Thank you for your comment. The temporal distribution of peaks can be influenced by the presence of noise, with the extent of this influence depending on the amplitude of the noise. Some types of anomalies (motion artifacts) might lead to increased variability in peak variance. Such anomalies cause unpredictable patterns in the occurrence of peaks. Contact force artifacts, or signal channel saturation, could lead to a decrease in peak variance.

   Modifications: No modifications are done regarding this comment.

8. **P. 15: "Low-Frequency/High=Frequency (LF/HF) ratio" – "=" symbol should be replaced by "-"**

   Response: Thank you for your remark. We decided to rename the feature since another reviewer suggested renaming the LF/HF feature to a more suggestive name that does not create any confusion with other concepts.

Modifications: We have updated the concept name to "Frequency band ratio," and it is consistently referred to by this name throughout the text.

9. **P.16 Fig 16: What is the x-axis? Is it a window number? What are the features in the graphs (peaks/dips)?**

   Response: Thank you for your comment. The x-axis represents the window number or, more specifically, the LF/HF ratio of a specific window.

   Modifications: We decided to move the figure to the Appendix due to space constraints. We updated the figure to include a clear and descriptive x-axis label to improve the clarity and understanding of the plot.

10. **P.16: "Number of peaks" is not explained.**

    Response: Thank you for your feedback. We apologize for the lack of explanation regarding the term.

    Modifications: We modified Subsection 4.5 and included a clear definition of "Number of peaks" and ensured that it is properly explained.

11. **TSFEL and tsfresh feature sets are not explained**

    Response: Thank you for your remark. In response to the reviewer's request, we have included a table in the manuscript that provides a detailed explanation of all three feature sets used in our study. This table serves as a comprehensive reference, highlighting the key characteristics and components of each feature set. By including this information, we aim to enhance the clarity and understanding of the feature extraction process in our research.

    Modifications: In Subsection 5.1, we have incorporated a table that elucidates the three feature sets determined in our study, namely the custom feature set, TSFEL feature set, and tsfresh feature set. This table provides a comprehensive overview of each feature set, outlining their distinctive characteristics and components.

12. **P.17 par. 4: The "random forest" method is described two times, but once it should be the "decision tree" method.**

    Response: Thank you for your remark. Indeed, there was an incorrect reference to the "random forest" method instead of the intended "decision tree" method. We apologize for the confusion caused by this oversight.

    Modifications: In Subsection 5.2 we provide the appropriate correction to accurately reflect the parameters of "decision tree" method.

13. **P.21 Table 2: Explain the difference between custom feature set and the other two.**

    Response: Thank you for your comment. In response to the reviewer's request, we have included a table in the manuscript that provides a detailed explanation of all three feature sets used in our study.

    Modifications: In Subsection 5.1, we have incorporated a table that elucidates the three feature sets determined in our study. This clarifies the difference between the custom feature set and the other two sets.

14. **The authors should discuss effect of a very small subject group size (two volunteers) on the study results. How does it affect them? What would they expect if a darks or obese volunteer would participate?**

    Response: Thank you for your remark. Indeed, it has been seen in previous research that inaccurate PPG measurements occur more frequently in dark skin as compared to light skin, and more in obese subjects. We believe that better sensor selection, optimal wavelengths, and sensor configurations should be studied.

    In the current study, we are not focusing on detecting anomalies in the PPG pulse shape or disease-related anomalies but more general and subject-independent artifacts such as body gross movement, signal dropout, signal attenuation due to contact force or low temperature. Therefore, if such volunteers were included in the test set, assuming that the current sensor configuration is not performing well in measuring the PPG pulse from such subjects, we could detect this as an anomaly. This is because such samples would already be considered anomalies in the training set, either due to dropout, contact force, or low temperature, as observed in other subjects that have lighter skin and low adipose tissue.

    However, we acknowledge the very small subject group size could represent a drawback due to the limited generalizability of the results. Anomaly detection algorithms rely on learning patterns and identifying deviations from those patterns. With only two subjects, the algorithm may not have enough diverse data to accurately capture the full range of motion artifacts and contact force variations that can occur across different individuals. However, we tried to leverage that by applying data augmentation methods, such as scaling, warping, and noise addition, to introduce different instances of artifacts based on measured data.

    Modifications: We have put a large amount of thought into this remark and decided to add in Section 6 a limitation of this study.

15. **How the results of this study would impact the wearables design?**

    Response: Thank you for your comment. We believe that the feature wearables design will shift towards a multichannel and a multi-wavelength (MW) approach. This shift can be due to the enhanced signal quality that is offered by a multichannel and MW-PPG system. However, by only increasing the number of measurement channels, without other channel fusion algorithms, this advantage cannot be explored. For instance, to make a PPG system more suitable for measuring dark skin tones and obese individuals, powerful signal processing techniques should be used (signal decomposition, independent component analysis). However, when applying these techniques, the assumption is that all channels contain relevant PPG information and that the motion artifacts are present in all channels to some extent. This thing might not be true for all the channels, and that is why we believe an anomaly detection stage would improve the overall accuracy of the system and avoid distorted signal reconstruction.

    Moreover, for the current wearable devices that already employ MW-PPG for SpO_2, heart rate and respiration rate estimation, the classical signal quality index (SQI) can be replaced by an anomaly detection stage that takes into account all the channels and has an increased accuracy. Several studies have demonstrated that HR measurements from wearable devices are often less accurate during physical activity or cyclic wrist motions. We believe that an anomaly detection stage increases the confidence level of the extracted parameters.

    Modifications: In Section 6, we presented a practical use case on how PPG devices can benefit from anomaly detection in MW-PPG signals.