Journal of Artificial Intelligence and Data Science (JAIDA)
e-ISSN: 2791-8335
https://dergipark.org.tr/pub/jaida

Vol. 6, No. 1, pp. 96-102, 2026
Research Article
Received: 11/06/2026 Accepted: 23/06/2026

Regression-Based Estimation of Wearable Sensor Signal Quality Using Statistical and Spectral Features

Gözde Çay*

Abstract

Physiological signals collected from wearable sensors are usually prone to noise due to motion artifacts, sensor displacement, and other sources. Since this signal degradation might cause problems in signal reliability, signal quality needs to be checked and determined. Although machine learning techniques have been successfully applied to signal quality assessment, the dominant features and their effects on different physiological modalities still need to be investigated. In this study, I performed an analysis on modality-specific signal degradation mechanisms in electrocardiography, electrodermal activity, and respiration signals using interpretable machine learning techniques including Random Forest, eXteme Gradient Boosting, and Light Gradient Boosting Machine. The models were trained using Leave-One-Subject-Out Cross-Validation to estimate the signal quality levels. In addition, feature importance analysis and SHapley Additive exPlanations were employed to investigate model interpretability. The results showed high prediction performance across all physiological modalities. EDA signals performed the highest coefficient of determination (R^2 = 0.996) followed by respiration (R^2 = 0.954) and ECG (R^2 = 0.908). Feature importance and SHAP analysis also confirmed these results, in addition they reflected the fact that the most dominant feature differs for EDA and ECG/respiration signals. These findings demonstrated that physiological signal degradation exhibited distinct modality-specific characteristics and interpretable machine learning techniques could provide valuable insight into the mechanisms of signal degradation.

Keywords: biomedical signals; machine learning; regression; signal quality; wearable sensors.

1. Introduction

Physiological sensors are widely used in wearable devices, remote monitoring systems, and healthcare applications to continuously measure vital signals such as electrocardiography (ECG), electroencephalography (EEG), electromyography (EMG), respiration, blood oxygen level (SpO_2), electrodermal activity (EDA) and so on. These signals provide valuable information regarding cardiovascular function, autonomic nervous system activity, stress levels and overall health status. The rising trend of using wearable sensing technologies allows continuous monitoring of these vital signals outside of the clinical environment [1], [2], [3], [4], [5].

Although wearable sensing has several advantages, these sensing technologies are frequently affected by various sources of signal degradation including motion artifacts, sensor displacement, poor skin contact, environmental interference and transmission errors. Low-quality signals may affect the signal processing, feature extraction, and machine learning decisions negatively. Therefore, signal quality assessment (SQA) becomes an important research topic in biomedical signal processing. Several studies have proposed signal quality indicators (SQI) to quantify the reliability of physiological signals. Traditional SQI methods often rely on statistical, or frequency-domain characteristics extracted from the signal [6], [7], [8], [9], [10], [11]. More recently, machine learning (ML) methods have been adapted to automate SQI assessments. Recent studies suggest that the ML models could successfully predict the low SQI; however, they lack a comprehensive understanding of the specific feature characteristics that dominate signal degradation and how these features vary across different physiological modalities [12], [13], [14], [15], [16].

Therefore, the objective of this study is to investigate whether signal degradation severity could be estimated from statistical and spectral features and whether the prediction performance differs among different physiological signals, namely ECG, EDA and respiration.

To achieve this objective, this research makes the following scientific contributions:

- A set of statistical and frequency-domain features were extracted from the degraded signals,
- Machine learning models were trained with these features,
- Model performances were compared through the coefficient of determination and model interpretability was investigated through feature importance analysis.

*Corresponding author

GÖZDE ÇAY; Izmir Katip Celebi University, Faculty of Engineering and Architecture, Department of Computer Engineering, Türkiye; e-mail: gozde.cay@ikc.edu.tr ; 0000-0003-1985-1772

2. Dataset and Analysis

2.1. Wearable sensors dataset

This study investigated the publicly available WeSAD dataset [17]. This dataset consists of physiological and motion data, recorded from both a wrist- (Empatica E4 [18]) and a chest-worn (RespiBAN [19]) device, of 15 subjects. The data includes blood volume pulse, electrocardiogram (ECG), electrodermal activity (EDA), electromyogram (EMG), respiration, body temperature, and three-axis acceleration. In this study, I used ECG, EDA and respiration signals since these physiological signals are widely used in wearable health monitoring and represent different signal characteristics.

2.2 Analysis

2.2.1. Feature extraction

Statistical features were extracted from the degraded sensor signals using Python version 3.14 (Python Software Foundation, Fredericksburg, VA, USA). Prior to the feature extraction, the signals were segmented into fixed-length windows of 60 seconds and 50% overlap. Then synthetic degradation was applied to each window by replacing missing values by signal mean and adding Gaussian noise according to signal-to-noise ratio (SNR). The SNR in decibels was defined as in Eq. (1) where P_{signal} and P_{noise} denote the average powers of the original signal and additive noise, respectively. The noise power was obtained by rearranging the Eq. (1) according to predefined SNR values.

$$SNR_{dB} = 10\log_{10}\left(\frac{P_{signal}}{P_{noise}}\right) \quad (1)$$

The degradation framework produced stable and signal-specific changes in statistical and spectral features across ECG, EDA, and respiration signals. Five predefined Signal Quality Index (SQI) levels were determined according to progressively increasing levels of additive noise and missing samples as summarized in *Table 1*. Therefore, SQI values were treated as ordinal labels representing controlled degradation conditions rather than values calculated from a mathematical expression.

Table 1: Signal degradation metrics

SQI	Missing ratio	SNR
1	0.00	None
0.75	0.05	25
0.5	0.10	20
0.25	0.20	15
0	0.30	10

Then, the statistical features were calculated from each degraded window. The extracted features included mean, standard deviation, variance, minimum, maximum, range, root mean square (RMS), skewness, kurtosis, entropy, dominant frequency, and spectral entropy. These features were stored in a .csv file and used as inputs for subsequent machine learning analyses.

2.2.2. Regression

Random Forest (RF) [20], eXtreme Gradient Boosting (XGBoost) [21], and Light Gradient Boosting Machine (LightGBM) [22] regression models were used to investigate whether the signal degradation could be estimated using statistical features extracted from the raw signal. The training and testing sample size of the machine learning model were split using Leave-One-Subject-Out Cross-Validation (LOSO-CV) [23] and models with 300 estimators were trained on the extracted features. The performance of model was compared using the coefficient of determination (R^2) score and statistical significance between them using Wilcoxon signed-rank test [24].

In addition, feature importance ranking analysis was performed to determine what physical properties of the signal contributed mostly to the regression performance and to provide insight into the physical characteristics of the ECG, EDA and respiration signals. SHapley Additive exPlanations (SHAP) [25] analysis was additionally performed to investigate feature contributions.

The machine learning process was conducted using Python version 3.14 (Python Software Foundation, Fredericksburg, VA, USA), leveraging NumPy and Pandas libraries for data handling, and scikit-learn, xgboost and lightgbm libraries for machine learning functionalities and statistical analysis.

3. Results and Discussion

3.1. Feature behavior under signal degradation

Signal degradation was performed successfully. Visual inspection confirmed progressive deterioration of signal as the amount of additive noise and missing samples increased. *Figure 1* depicts an ECG signal window as a representative sample of the signal degradation.

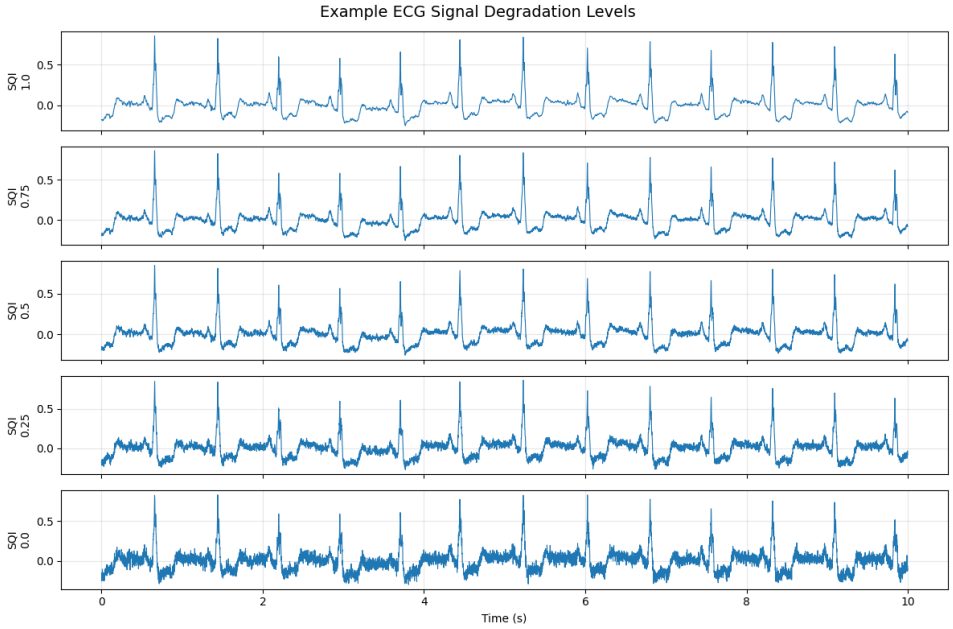


Figure 1: Signal degradation on an ECG signal window

The generated signal quality dataset contains 43,110 observations of features with balanced SQI levels. Analysis of extracted features revealed clear relationships between signal quality and signal characteristics. For ECG signals, spectral entropy increased from 6.79 at SQI = 1.00 to 8.61 at SQI = 0.00 (*Figure 2*), indicating that signal degradation increased spectral complexity and reduced signal regularity.

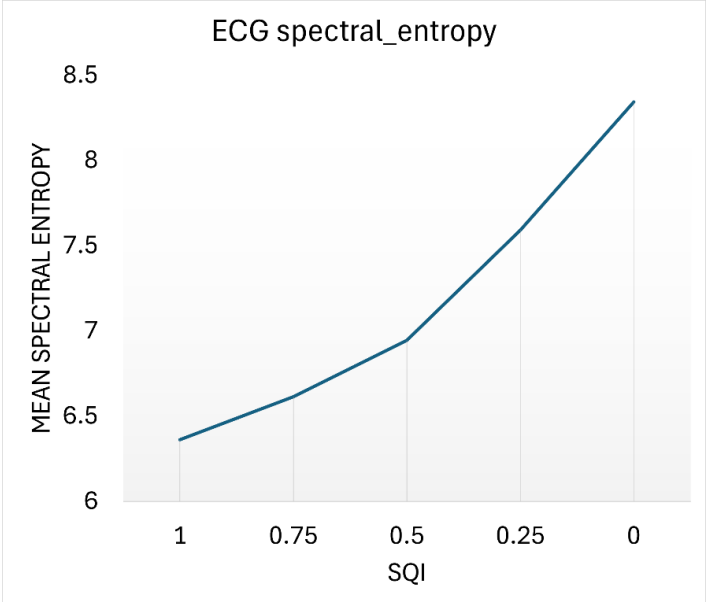


Figure 2: ECG Spectral Entropy vs SQI - Average spectral entropy increases as signal quality decreases.

EDA signals exhibited a different degradation pattern. The minimum signal value decreased substantially with increasing degradation (*Figure 3*), suggesting that EDA quality was strongly influenced by amplitude-related characteristics. Spectral entropy and kurtosis also changed systematically across SQI levels but to a lesser extent than the minimum value.

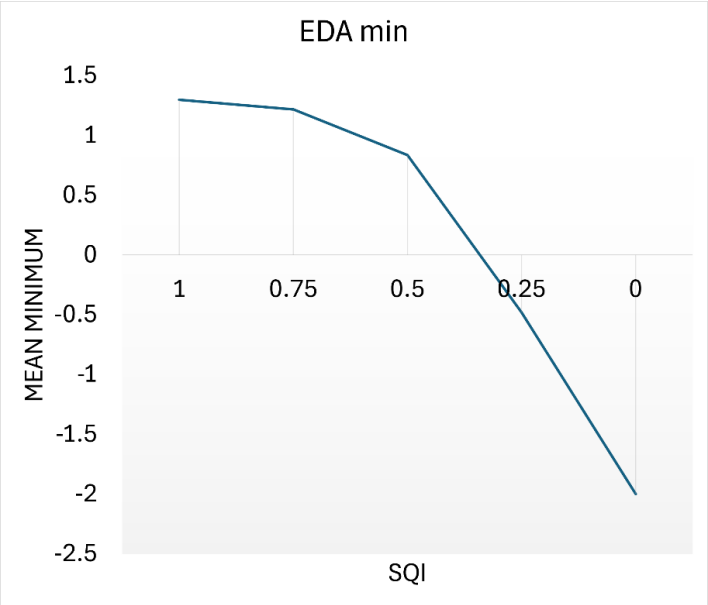


Figure 3: EDA Minimum Value vs SQI - Minimum signal value decreases as degradation increases.

Respiratory signals demonstrated progressive increases in spectral entropy from 2.72 at SQI = 1.00 to 6.58 at SQI = 0.00. Kurtosis and entropy also increased as signal quality deteriorated, indicating greater signal irregularity and complexity.

These observations suggest that different physiological modalities exhibit distinct degradation mechanisms and may require different signal quality indicators.

3.2. Machine learning-based signal quality estimation

EDA exhibited the strongest degradation-dependent feature changes and achieved the highest quality-estimation performance, followed by respiration signals and ECG among all models. For ECG signals, RF achieved the highest performance, followed by XGBoost and LightGBM. It was also seen that XGBoost achieved the highest performance, followed by LightGBM and RF for respiration signals. *Table 2* summarizes the regression performance of the evaluated models.

Table 2: Regression performance of evaluated models

Signal	RF			XGB			LGBM		
	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE
ECG	0.908	0.037	0.062	0.899	0.048	0.067	0.877	0.047	0.069
EDA	0.995	0.004	0.018	0.996	0.005	0.017	0.977	0.024	0.045
Resp	0.951	0.041	0.078	0.954	0.044	0.076	0.953	0.042	0.076

Overall, the results proved that the statistical and frequency-domain features could effectively capture information related to signal degradation.

3.3. Statistical comparison of machine learning models

For ECG signals, RF significantly outperformed XGBoost ($p = 0.035$, Cohen's $d = 0.51$) with a moderate effect size whereas no significant difference was observed between RF and LightGBM and XGBoost and LightGBM ($p = 0.229$ and 0.804, respectively). Analysis of respiration signals revealed, both XGBoost and LightGBM significantly outperformed RF ($p = 0.001$, Cohen's $d = -1.09, -1.14$, respectively) with a large effect size. It was seen that although the absolute performance difference between the models was small ($\Delta R^2 \approx 0.003$), the improvement was consistently observed across subjects and reached statistical significance. For the EDA signals, RF and XGBoost demonstrated comparable performance, while XGBoost significantly outperformed LightGBM ($p = 0.008$). However, the effect size is negligible (Cohen's $d = -0.07$), indicating that the practical performance difference between the models was minimal despite the consistency of the observed improvement across subjects. Therefore, I conclude that the optimal machine learning model depends on the physiological signal modality.

3.4. Feature importance analysis

For ECG signals, spectral entropy consistently emerged as the most influential predictor across machine learning models. SHAP analysis also confirmed that increases in spectral entropy were associated with reductions in signal quality as shown in *Figure 4*.

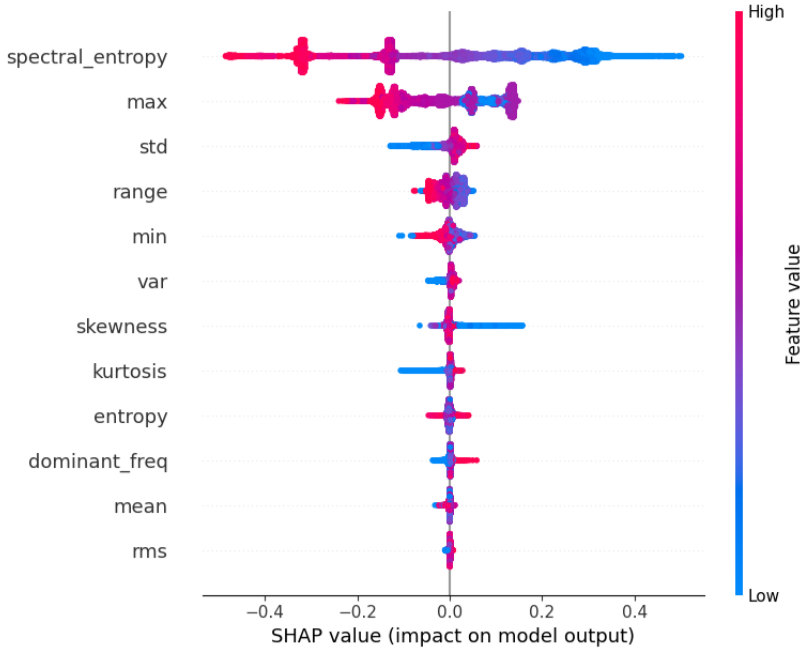


Figure 4: SHAP summary plot for ECG quality estimation using XGBoost. Spectral entropy exhibited the largest impact on model predictions, with higher entropy values consistently associated with lower signal quality. Secondary contributions were observed from amplitude-related features including maximum value, standard deviation, and signal range.

Conversely, for EDA signals, the minimum signal value was consistently identified as the most important predictor by Random Forest, XGBoost, LightGBM, and SHAP analyses. This finding suggests that EDA degradation is primarily reflected through changes in signal amplitude rather than spectral complexity.

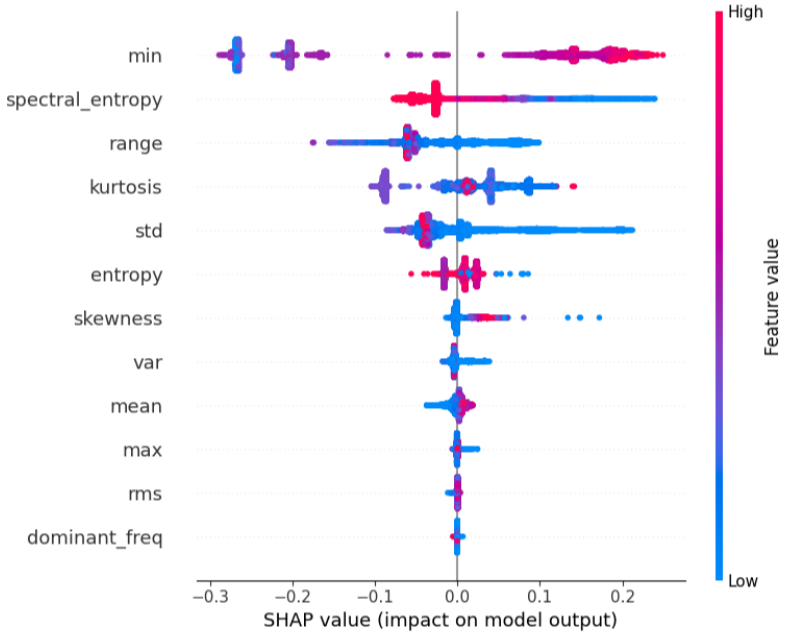


Figure 5: SHAP summary plot for EDA quality estimation using XGBoost. Minimum exhibited the largest impact on model predictions. Secondary contributions were observed from spectral and signal complexity including spectral entropy, signal range and kurtosis.

Regarding respiratory signals, spectral entropy again emerged as the dominant feature. Additional contributions from entropy, kurtosis, and dominant frequency were observed, particularly in the XGBoost model. SHAP analysis confirmed that increasing spectral entropy was strongly associated with lower signal quality levels.

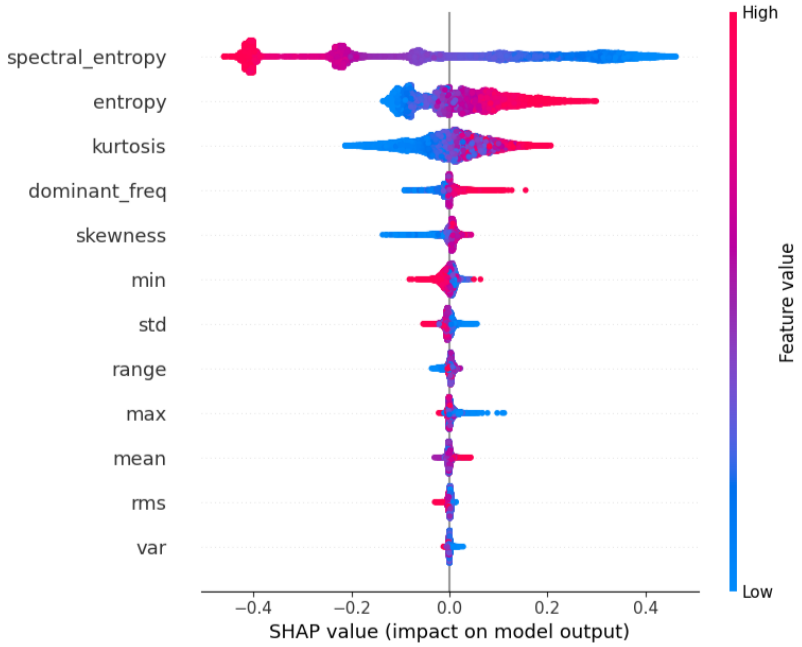


Figure 6: SHAP summary plot for respiration quality estimation using XGBoost. Spectral entropy exhibited the largest impact on model predictions. Secondary contributions were observed from entropy, kurtosis and dominant frequency.

The consistency of these findings across RF, XGBoost, LightGBM and SHAP analyses proved that while ECG degradation is primarily expressed through increasing spectral complexity, EDA degradation is primarily reflected through amplitude-boundary changes and respiration degradation is primarily characterized by increasing spectral disorder and signal complexity.

Overall, these observations highlight the importance of modality-specific signal quality assessment strategies since different physiological modalities demonstrate distinct degradation signatures.

4. Conclusion

In this paper, the regression models, namely Random Forest, XGBoost and LightGBM, on estimating signal degradation from wearable sensor signals were investigated. Twelve features including mean, standard deviation, variance, minimum, maximum, range, root mean square (RMS), skewness, kurtosis, entropy, dominant frequency, and spectral entropy were extracted from degraded ECG, EDA and respiration signals and the models were trained with these features. Signal degradation signatures varied distinctly across physiological modalities. Across all models, as well as SHAP analyses, spectral entropy consistently emerged as the dominant predictor of signal quality for both ECG and respiration signals. In contrast, EDA signal degradation was primarily influenced by the minimum signal value. The results confirmed that signal degradation severity could be estimated from statistical and spectral features, and prediction performance differs among ECG, EDA, and respiration signals based on their physiological modalities. Ultimately, this study offers a novel approach to evaluating wearable sensor signal degradation to design and develop more reliable monitoring systems.

Although the promising results, the study has several limitations. The signal degradation was synthetically generated so it may not fully represent the complexity of real-world artifacts encountered in wearable sensing applications. Also, only ECG, EDA, and respiratory signals were included in the study. Additional physiological modalities should be considered in future studies. Finally, only statistical and frequency-domain features were examined. Deep learning models may be investigated in future studies.

Declaration of Interest

The author declares that there is no conflict of interest.

References

[1] Y. Jiang, C. Spies, J. Magin, S. J. Bhosai, L. Snyder, and J. Dunn, “Investigating the accuracy of blood oxygen saturation measurements in common consumer smartwatches,” *PLOS Digital Health*, vol. 2, no. 7, p. e0000296, Jul. 2023, doi: 10.1371/JOURNAL.PDIG.0000296.

[2] A. Anma and M. Komori, “The Efficacy of a Wearable Device-based Vital Sign Monitoring System in a Japanese Home Medical Care Clinic,” *Internal Medicine*, vol. 64, no. 24, pp. 3502–3508, Dec. 2025, doi: 10.2169/INTERNALMEDICINE.5213-24.
[3] M. Joshi *et al.*, “Performance of Continuous Digital Monitoring of Vital Signs with a Wearable Sensor in Acute Hospital Settings,” *Sensors 2025, Vol. 25, Page 2644*, vol. 25, no. 9, p. 2644, Apr. 2025, doi: 10.3390/S25092644.
[4] M. J. M. Breteler *et al.*, “Reliability of an all-in-one wearable sensor for continuous vital signs monitoring in high-risk patients: the NIGHTINGALE clinical validation study,” *Journal of Clinical Monitoring and Computing 2025 39:5*, vol. 39, no. 5, pp. 1087–1100, Mar. 2025, doi: 10.1007/S10877-025-01279-X.
[5] A. Angelucci, M. Greco, M. Cecconi, and A. Aliverti, “Wearable devices for patient monitoring in the intensive care unit,” *Intensive Care Medicine Experimental 2025 13:1*, vol. 13, no. 1, pp. 26-, Feb. 2025, doi: 10.1186/S40635-025-00738-8.
[6] S. Fallet *et al.*, “Assessing ECG signal quality indices to discriminate ECGs with artefacts from pathologically different arrhythmic ECGs,” *Physiol. Meas.*, vol. 37, no. 8, p. 1370, Jul. 2016, doi: 10.1088/0967-3334/37/8/1370.
[7] R. He *et al.*, “Reducing false arrhythmia alarms in the ICU using novel signal quality indices assessment method,” *Comput. Cardiol. (2010).*, vol. 42, pp. 1189–1192, Feb. 2015, doi: 10.1109/CIC.2015.7411129.
[8] S. Fallet *et al.*, “Assessing ECG signal quality indices to discriminate ECGs with artefacts from pathologically different arrhythmic ECGs,” *Physiol. Meas.*, vol. 37, no. 8, p. 1370, Jul. 2016, doi: 10.1088/0967-3334/37/8/1370.
[9] C. Orphanidou, “Signal Quality Assessment in Physiological Monitoring: Requirements, Practices and Future Directions,” pp. 1–14, 2018, doi: 10.1007/978-3-319-68415-4_1.
[10] A. Ignácz, S. Földi, P. Sótonyi, and G. Cserey, “NB-SQI: A novel non-binary signal quality index for continuous blood pressure waveforms,” *Biomed. Signal Process. Control*, vol. 70, p. 103035, Sep. 2021, doi: 10.1016/J.BSPC.2021.103035.
[11] M. Elgendi, “Optimal Signal Quality Index for Photoplethysmogram Signals,” *Bioengineering 2016, Vol. 3, Page 21*, vol. 3, no. 4, p. 21, Sep. 2016, doi: 10.3390/BIOENGINEERING3040021.
[12] A. Mondal, M. S. Manikandan, and R. B. Pachori, “Automatic ECG signal quality assessment using convolutional neural networks and derivative ECG signal for false alarm reduction in wearable vital signs monitoring devices,” *Biomed. Signal Process. Control*, vol. 108, p. 107876, Oct. 2025, doi: 10.1016/J.BSPC.2025.107876.
[13] Y. ; Liu *et al.*, “An Automatic ECG Signal Quality Assessment Method Based on Resnet and Self-Attention,” *Applied Sciences 2023, Vol. 13, Page 1313*, vol. 13, no. 3, p. 1313, Jan. 2023, doi: 10.3390/APP13031313.
[14] F. Mohagheghian *et al.*, “Optimized Signal Quality Assessment for Photoplethysmogram Signals Using Feature Selection,” *IEEE Trans. Biomed. Eng.*, vol. 69, no. 9, pp. 2982–2993, Sep. 2022, doi: 10.1109/TBME.2022.3158582.
[15] E. Idrobo-Avila, G. Bognar, D. Krefting, T. Penzel, P. Kovacs, and N. Spicher, “Quantifying the Suitability of Biosignals Acquired During Surgery for Multimodal Analysis,” *IEEE Open J. Eng. Med. Biol.*, vol. 5, pp. 250–260, 2024, doi: 10.1109/OJEMB.2024.3379733.
[16] Y. Jiang *et al.*, “An Optimized Signal Quality Assessment Method for Noncontact Capacitive ECG,” *IEEE Trans. Instrum. Meas.*, vol. 74, 2025, doi: 10.1109/TIM.2025.3533644.
[17] P. Schmidt, A. Reiss, R. Duerichen, and K. Van Laerhoven, “Introducing WeSAD, a multimodal dataset for wearable stress and affect detection,” *ICMI 2018 - Proceedings of the 2018 International Conference on Multimodal Interaction*, pp. 400–408, Oct. 2018, doi: 10.1145/3242969.3242985.
[18] J. Stritzelberger *et al.*, “Validity of Empatica E4 Wristband for Detection of Autonomic Dysfunction Compared to Established Laboratory Testing,” *Diagnostics*, vol. 15, no. 20, p. 2604, Oct. 2025, doi: 10.3390/DIAGNOSTICS15202604.
[19] A. Simons, T. Doyle, D. Musson, and J. Reilly, “Impact of Physiological Sensor Variance on Machine Learning Algorithms,” *Conf. Proc. IEEE Int. Conf. Syst. Man Cybern.*, vol. 2020-October, pp. 241–247, Oct. 2020, doi: 10.1109/SMC42975.2020.9282912.
[20] L. Breiman, “Random forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.
[21] “XGBoost.” Accessed: Jun. 10, 2026. [Online]. Available: https://xgboost.ai/
[22] “Welcome to LightGBM’s documentation! — LightGBM 4.6.0 documentation.” Accessed: Jun. 10, 2026. [Online]. Available: https://lightgbm.readthedocs.io/en/stable/
[23] “LeaveOneOut — scikit-learn 1.9.0 documentation.” Accessed: Jun. 10, 2026. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.LeaveOneOut.html
[24] B. Rosner, R. J. Glynn, and M. L. T. Lee, “The Wilcoxon signed rank test for paired comparisons of clustered data,” *Biometrics*, vol. 62, no. 1, pp. 185–192, 2006, doi: 10.1111/J.1541-0420.2005.00389.X.
[25] “Welcome to the SHAP documentation — SHAP latest documentation.” Accessed: Jun. 10, 2026. [Online]. Available: https://shap.readthedocs.io/en/latest/