FOREX Publication
Open Access | Rapid and quality publishing

International Journal of
Electrical and Electronics Research (IJEER)
Research Article | Volume 13, Issue 4 | Pages 704-711 | e-ISSN: 2347-470X

Hybrid CNN-BiLSTM Autoencoder for Anomaly Detection in Remote Photoplethysmography(rPPG) Signals

Moussa Mmadi[1*], George Kamucha[2], and Ciira wa Maina[3]

[1]*Department of Electrical Engineering, Pan African University Institute for Basic Sciences, Technology and Innovation, Nairobi, Kenya; moussa.mmadi@students.jkuat.ac.ke*
[2]*Department of Electrical and Information Engineering, University of Nairobi, Nairobi, Kenya; gkamucha@uonbi.ac.ke*
[3]*Department of Electrical Engineering, Dedan Kimathi University of Technology, Nyeri, Kenya; ciira.maina@dkut.ac.ke*

***Correspondence:** *Moussa Mmadi; moussa.mmadi@students.jkuat.ac.ke*

ABSTRACT- Remote photoplethysmography (rPPG) is becoming increasingly popular as a non-contact method for tracking physiological parameters like heart rate and respiration rate. However, the accuracy of rPPG signals is often compromised by various factors, including movement, lighting variations, and sensor noise. These challenges can severely impact signal quality, leading to unreliable measurements and hindering the practical application of rPPG-based systems. In this research, we introduced and assessed the effectiveness of a hybrid Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory (BiLSTM) Autoencoder model specifically designed for reliable anomaly detection in rPPG time-series data. This model aims to detect anomalies such as sensor noise and motion artifacts within the remote photoplethysmography signal. The UBFC-rPPG, COHFACE, and the PURE datasets were utilized for training and testing, demonstrating excellent performance in distinguishing clean segments from noisy ones, with high precision, recall, F1-score, and low false positive rates.

Keywords: Remote Photoplethysmography, Convolutional Neural Network, Anomaly Detection, Time-Series, BiLSTM.

ARTICLE INFORMATION

Author(s): Moussa Mmadi, George Kamucha, and Ciira wa Maina;
Received: 30/06/2025; **Accepted**: 28/10/2025; **Published**: 10/12/2025;
E- ISSN: 2347-470X;
Paper Id: IJEER 3006-22;
Citation: 10.37391/ijeer.130410
Webpage-link:
https://ijeer.forexjournal.co.in/archive/volume-13/ijeer-130410.html

Publisher's Note: FOREX Publication stays neutral with regard to jurisdictional claims in Published maps and institutional affiliations.

1. INTRODUCTION

The medical field has experienced significant changes over time, leading to improvements in its dependability and effectiveness in patient care. The creation of techniques for assessing physiological parameters has become essential. A recent study [20] suggests that noncontact technologies can be used to gather information about an individual's physiological signals. This remote monitoring approach, called remote photoplethysmography (rPPG), is a groundbreaking non-contact method for estimating vital signs by analyzing the color variations in human skin captured by video cameras. Through video processing techniques, an algorithm extracts skin pixel data from each frame, and then the spatial averages of these pixels are calculated for each image. This process produces a series of temporal rPPG signals derived from spatial averages. By examining these temporal data, it is possible to extract the physiological information contained within the rPPG signal. This technology enables the monitoring of physiological metrics such as heart rate (HR), body temperature (BT), respiratory rate (RR), blood oxygen saturation (SpO2), heart rate variability (HRV), and blood pressure (BP) [6–12], without the need for physical contact, offering significant.

Anomalies are a fundamental aspect of nearly every system in today's world, which is inundated with numerous Internet of Things (IoT) devices generating vast amounts of data [10]. The process of anomaly detection, which involves identifying unexpected items or events within data, has garnered significant interest among researchers and practitioners. It is now a primary focus in data mining and quality assurance [5]. This field has been explored across various application domains and has seen considerable advancements. In the realm of rPPG signals, anomalies refer to unexpected, irregular, or abnormal patterns within the extracted physiological signal that deviate from the anticipated heart-related waveform. Such anomalies can greatly affect the accuracy of vital sign estimation. The rPPG signals are particularly susceptible to various types of anomalies that can degrade signal quality and lead to incorrect heart rate or respiratory rate estimations. These anomalies are caused by user movement, environmental conditions, sensor limitations, and physiological variability [10],[11]. *Table 1* presents the main common types of anomalies that can impair rPPG signal quality and their causes.

Some studies collectively highlighted the crucial role of pre-processing methods and anomaly detection systems in enhancing the dependability of rPPG signals for real-world use. However, there are still obstacles to overcome, especially when dealing with complex real-world scenarios that involve changing lighting conditions, facial obstructions, and significant head movements.

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 13, Issue 4 | Pages 704-711 | e-ISSN: 2347-470X

To address these obstacles, it is crucial to obtain an accurate rPPG signal. Therefore, developing an algorithm that can identify irregularities and assess signal quality in real-time is necessary to enhance the use of rPPG technology in practical applications. Due to the scarcity of research on anomaly detection in rPPG signals, our study focuses on these aspects. The main objective of this paper is to develop a Hybrid CNN–BiLSTM Autoencoder model designed for robust anomaly detection in the rPPG time series data. Unlike methods that rely solely on CNNs, which focus on capturing local spatial characteristics, or those that use only LSTMs, which concentrate on temporal relationships, our model combines the strengths of both approaches. The CNN layers are adept at detecting local skin color variations due to changes in blood volume. In contrast, the BiLSTM layers are capable of modeling long-term sequential patterns over time. This combination enhances the model's ability to differentiate between clean physiological signals and anomalies such as motion artifacts or sensor noise. An ablation study reveals that our hybrid model significantly outperforms models based solely on CNNs or LSTMs.

The key contributions are as follows: *(i)* To preprocess the data from the video and compute rPPG signals. *(ii)* To segment remote photoplethysmography signals into short windows. *(iii)* To build and train a hybrid CNN-BiLSTM Autoencoder model for identifying segments where the anomaly significantly distorts the signal. *(iv)* To use the reconstruction error to detect anomalies.

The rest of the work is organized as follows: *Section 2* covers related works, *section 3* details the methodology we used, *section 4* provides the experiments and results, *section 5* presents a discussion, and *section 6* offers the conclusion.

Table 1. Types of Anomalies in rPPG Signals [1-26]

Anomaly type	Cause of anomaly
Physiological Irregularities	Arrhythmias or stress
Motion-Induced Anomalies	head movement, facial expressions, or body motion
Occlusion-Related Anomalies	Facial occlusions
Lighting-Induced Artifacts	Variations in ambient light or screen flicker
Sensor-Related Anomalies	hardware limitations

2. RELATED WORKS

Many researchers have concentrated on improving the quality of rPPG signals by identifying and reducing noise and irregularities. Initial methods mainly utilized signal processing techniques like bandpass filtering [20] and ICA [21] to distinguish the physiological signal from noise. De Haan and Jeanne [14] introduced the CHROM method, which enhances resistance to motion by projecting the color signal into a chrominance subspace. Poh et al. [15] also showed that blind source separation techniques can successfully extract rPPG signals from facial videos, even when significant artifacts are present.

Recent studies have increasingly focused on solutions based on machine learning. Speth et al. [4] presented a deep learning model that is anomaly-aware and discourages the prediction of unrealistic periodic signals, thereby enhancing the robustness of rPPG in unusual conditions. Botina-Monsalve et al. [5] investigated the use of LSTM networks for deep temporal filtering of rPPG signals, achieving better denoising results than traditional filters. Furthermore, Liu and Yuen [9] developed a deep disentanglement model that can separate environmental noise from physiological signals, thus improving the signal-to-noise ratio in various conditions.

There have been initiatives to evaluate and eliminate unreliable video segments before extracting signals. Kim et al. [3] introduced a screening technique based on noise assessment, which greatly enhances the quality of rPPG signals by removing frames of poor quality. Additionally, the recent DiffPhys model by Chen et al. [2] utilizes diffusion models to improve rPPG signals by learning the natural periodicity of blood flow, showing promising outcomes in boosting signal accuracy.

3. PROPOSED METHODOLOGY

Figure 1 illustrates the system diagram. This section outlines the methodology employed to address the challenges mentioned earlier. A hybrid model combining CNN and BiLSTM Autoencoder was introduced to learn the reconstruction of only normal patterns. By comparing the input and output and setting a threshold on the error, anomalies such as motion artifacts in the rPPG signal extraction process are identified. The performance was evaluated based on specificity and sensitivity.

3.1. Video Processing and Initial Remote Photoplethysmography Signal Estimation

Initially, video frames undergo processing to identify and monitor the region of interest (ROI) by employing the technique outlined in [24].

3.1. Time Series

A time series is a set of observations documented in a chronological order. Time series medical data pertains to information gathered over time from patients or medical devices, encompassing vital signs, laboratory results, and sensor data from wearable technology [26]. This type of data is utilized to monitor the development of a patient's condition, assess the effectiveness of treatments, or detect potential health concerns. In the realm of photoplethysmography signals, a time series denotes a sequence of signal values recorded at regular intervals, usually derived from facial video frames. Each value in the series represents the minute changes in skin color over time, resulting from fluctuations in blood volume due to the cardiac cycle.

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 13, Issue 4 | Pages 704-711 | e-ISSN: 2347-470X

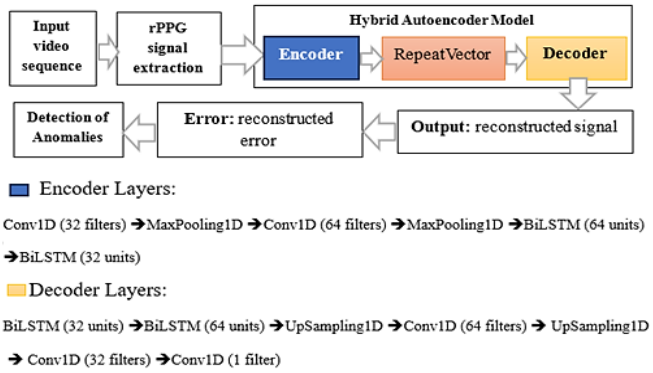

Figure 1. General methodology for detecting anomalies in the remote photoplethysmography signal

3.2. Autoencoder

An autoencoder (AE) is a type of unsupervised neural network designed to discover the optimal encoding and decoding strategy from data. Typically, it includes an input layer, an output layer, an encoder neural network, a decoder neural network, and a latent space. When data is introduced to the network, the encoder compresses it into the latent space, while the decoder reconstructs the encoded data back into the output layer. The resulting output is then compared to the original data, and any errors are backpropagated through the network to adjust the weights accordingly [26].

3.3. Overview of the Proposed Model's Structure

A hybrid model architecture is introduced, as illustrated in Figure 1. This model integrates Convolutional Neural Networks (CNNs) for extracting features and Bidirectional Long Short-Term Memory (BiLSTM) layers for learning temporal patterns, all within an Autoencoder framework. Its purpose is to identify and reconstruct sensor noise anomalies in rPPG time-series signals. The model is characterized as an Autoencoder, consisting of a hybrid encoder (CNN-BiLSTM), a latent bottleneck depicted by RepeatVector, and a hybrid decoder (CNN-BiLSTM). The following sections provide a detailed layer-by-layer description of the model's features.

3.3.1. Hybrid Encoder

The initial layer is a one-dimensional convolutional (Conv1D) layer, configured with thirty-two (32) filters and a kernel size of three (3), which determines the filter dimensions. The input shape is defined as (300, 1), signifying a time-series window of 300 frames that has been normalized and reshaped into a single channel, representing the rPPG signal amplitude. This layer utilizes the 'ReLU' activation function. Following this, a MaxPooling1D layer is incorporated, with a pool size of two (2), to downsample the temporal dimension, thereby reducing computational load while preserving key features.

Subsequently, another Conv1D layer is employed to identify more intricate patterns over a broader receptive field. This layer is set with sixty-four (64) filters and a kernel size of three (3), and also uses the 'ReLU' activation function. Another MaxPooling1D layer is added to this second Conv1D layer to further downsample, effectively halving the time steps again for a more compact representation, with a pool size of two (2).

Next, two BiLSTM layers are introduced. The first BiLSTM layer, with sixty-four (64) units or memory cells, is designed to capture temporal dependencies in both forward and backward directions, with return sequences set to True. The second BiLSTM layer, containing thirty-two (32) units or memory cells, aims to produce a fixed-size encoded vector (bottleneck) that encapsulates the temporal dynamics of the window, with return sequences set to False.

Following the second BiLSTM, a RepeatVector is utilized to prepare the bottleneck vector for reintroduction into the decoder for sequence reconstruction.

3.3.2. Hybrid Decoder

To initiate the process, two BiLSTM layers were employed. The first layer, consisting of thirty-two (32) units or memory cells, is responsible for initiating the reconstruction of the time series from its compressed form. The return sequences were configured to True. The second layer, with sixty-four (64) units or memory cells, is designed to further decode the sequence by leveraging long-term dependencies, with return sequences also set to True.

Following this, an UpSampling1D Layer is incorporated into the preceding layer, with the size parameter set to two (2), aiming to restore temporal resolution.

A one-dimensional convolution (Conv1D) layer is then applied to enhance the features of the upsampled signal. This layer is configured with sixty-four (64) filters and a kernel size of three (3), utilizing the 'ReLU' activation function.

Another UpSampling1D Layer is added to the previous Conv1D layer, with the size parameter again set to two (2), to recover the original sequence length completely.

An additional Conv1D layer is applied to smooth the intermediate features, featuring thirty-two (32) filters and a kernel size of three (3), with the 'ReLU' activation function.

The final output layer is a one-dimensional convolution (Conv1D) that aims to produce the reconstructed signal in the same shape as the input. This layer is configured with one filter and a kernel size of one, using the 'Linear' activation function.

3.4. Segmenting the rPPG Signal

Segmenting remote photoplethysmography signals into brief windows is a crucial step in preparing time-series data for models such as CNNs or BiLSTMs. This task involves breaking down a lengthy, continuous signal into smaller, overlapping or non-overlapping segments, making them more manageable for analysis and modeling. Such segmentation helps in capturing local temporal patterns. Considering a raw rPPG signal $X = \{x_1, x_2, x_3, \ldots, x_{K-1}\}$, where $X \in \mathbb{R}^K$ represents the complete time series rPPG signal, the total number of frames K, and the scalar value of the rPPG x_t at time t. The number of segments can then be determined as [23]:

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 13, Issue 4 | Pages 704-711 | e-ISSN: 2347-470X

$$N = \frac{K-W}{S} + 1 \quad (1)$$

Where W is the window length, and S is the stride. And for each window $X^{(i)} \in \mathbb{R}^W$ is given by $X^{(i)} = \{x_{iS}, x_{iS+1}, .., x_{iS+W-1}\}$, with $i \in \{0, 1, 2, \dots, N-1\}$

3.5. Measurement of Reconstruction Error

The reconstruction model aims to identify distorted parts in rPPG time series by understanding the behavior of a clean signal. It flags segments that deviate from this learned pattern. Each window $X^{(i)} \in \mathbb{R}^W$ is taken as input, and the model outputs the reconstructed segment $\tilde{X}^{(i)} \in \mathbb{R}^W$ for each window. The difference between the input and the reconstructed output is known as the reconstruction error. The model is trained to minimize this error between the input and output. During training, the model is exposed only to clean segments, allowing it to learn their typical structure. During inference, if the input deviates from this structure, the model struggles to reconstruct it accurately, resulting in a high error. The reconstruction error is measured as the Mean Absolute Error (MAE) per segment, which can be determined as [23]:

$$e = \frac{1}{W}\sum_{t=1}^{W}\left|X_t - \tilde{X}_t\right| \quad (2)$$

To determine an anomaly, we first compute the error e_i for each segment $X^{(i)}$, which can be derived from *equation (2)* as:

$$e_i = \frac{1}{W}\sum_{t=1}^{W}\left|X_t^{(i)} - \tilde{X}_t^{(i)}\right| \quad (3)$$

A threshold was determined based on the reconstruction errors from the training set. This threshold is specified in [23] as follows:

$$\Omega = \mu_e - \beta * \sigma_e \quad (4)$$

Where μ_e and σ_e respectively represent the mean and the standard deviation of the errors on the training data. The coefficient β acts as the sensitivity factor, with a value of 2 as stated in the literature [23]. The threshold Ω is used to flag the anomaly by comparing it to the reconstruction error for each segment. If $e_i^{test} > \Omega$, then the segment $X^{(i)}$ is classified as "anomalous"; otherwise, it is deemed "normal".

3.6. Performance Analysis

The proposed system model was evaluated using established metrics from the literature, such as precision (PR), recall (R), F1-score, and false positive rate (FPR). Precision denotes the positive predictive value, recall (also referred to as sensitivity) indicates the True positive rate (TPR), and the F1-score is derived as the harmonic mean of precision and recall. The formulas for each of these metrics are specified in [26] as follows:

$$PR = \frac{TP}{(TP + FP)} \quad (5)$$

$$recall = \frac{TP}{(TP + FN)} \quad (6)$$

$$FPR = \frac{FP}{(FP + TN)} \quad (7)$$

$$F1 - score = 2 * \frac{(PR * Recall)}{(PR + Recall)} \quad (8)$$

TP represents True Positives, which are the correctly predicted positive cases. TN stands for True Negatives, indicating the correctly predicted negative cases. FP refers to False Positives, where positive cases are incorrectly predicted. FN denotes False Negatives, where negative cases are incorrectly predicted.

4. EXPERIMENTS AND RESULTS

4.1. Dataset

The experiment utilizes public datasets for both training and testing. It involves three databases, including the PURE dataset [28], the COHFACE dataset [27], and the UBFC-RPPG database [22]. Each of these three databases consists of synchronized videos and physiological signals. Only PURE was used for training. All three databases are used for testing. *Table 2* summarizes the specific parameters for each.

Table 2. Summary of the Datasets

Datasets	Subjects	Frame rate	Resolution	Videos
COHFACE	40	20 fps	640 x 480	160
UBFC-RPPG	50	20-30 fps	640 x 480	50
PURE	10	30 fps	640 x 480	60

4.2. Data Preparation and Detection of Anomalous Segments

The remote photoplethysmography signal presented in *figure 2* was extracted from the PURE dataset. Using a window length of 300 frames and a stride of 50 frames between windows, the raw rPPG signal was divided into N overlapping segments. the value of N is determined using *equation (1)*. A low-pass filter was computed to remove high-frequency noise, simulating a "clean" reference signal. These segments are the noisy inputs used to train the model. The same segmentation method was applied to the filtered signal to use as the denoising target.

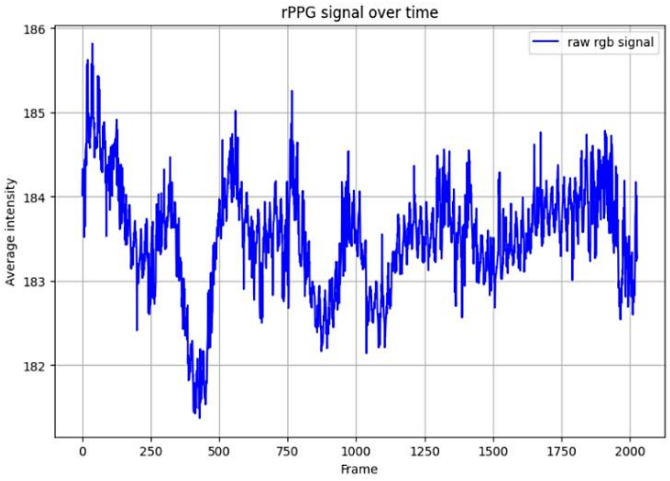

Figure 2. Remote photoplethysmography signal extracted

Figure 3 illustrates the reconstruction loss from the test data alongside the static reconstruction threshold derived from the

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 13, Issue 4 | Pages 704-711 | e-ISSN: 2347-470X

training loss. The signal in *figure 2* was divided into N = 35 segments. The model identifies an anomaly segment by computing the Mean Absolute Error (MAE) of the training data, with the highest MAE loss value serving as the Reconstruction Error Threshold. The model processes the test data, and due to the presence of an anomalous segment in the sequence, it is anticipated to reconstruct poorly, resulting in a high reconstruction loss at these anomalous points. Any reconstruction test loss value exceeding the reconstruction error threshold is marked as an anomaly.

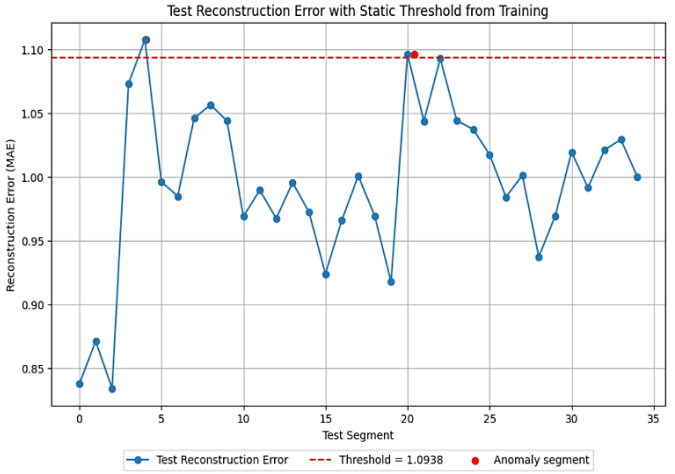

Figure 3. Test reconstruction error with static threshold

In *figure 4*, the motion artifacts present in the signal during video recording have been detected clearly. Our CNN-BiLSTM model has detected two (2) anomalous segments during the reconstruction loss.

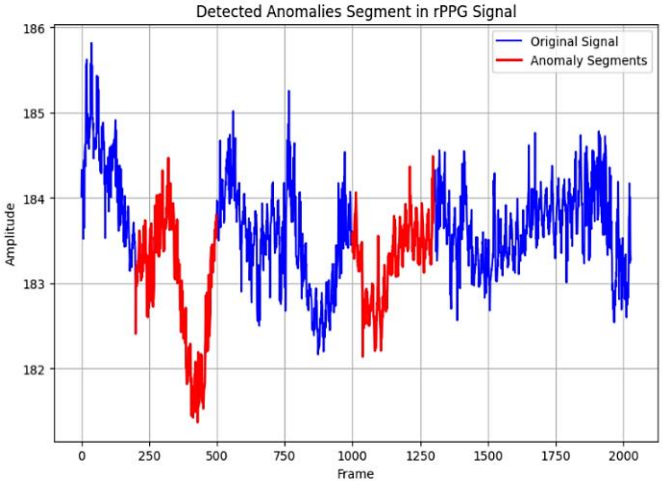

Figure 4. Motion artifacts detected on the rPPG signal

Figure 5 depicts the reconstructed rPPG signal following the removal of anomalies using the proposed CNN–BiLSTM Autoencoder. The blue waveform indicates the purified rPPG signal, while the green sections represent the parts that were reconstructed to substitute anomalous intervals. As illustrated, the model effectively reconstructs smooth, physiologically consistent waveforms in areas previously impacted by noise or motion artifacts. This highlights the model's capability to accurately identify and rectify distorted segments without excessive smoothing or losing the periodic structure of the rPPG waveform, which is crucial for the reliable downstream estimation of physiological parameters such as heart rate and respiration.

Figure 5. The final rPPG signal after reconstructing the two anomaly segments

4.3. Model Performance Evaluation

The confusion matrix presented in *figure 6* is used to represent the matrix of predicted labels. It illustrates that the autoencoder demonstrated strong differentiation, with True Positives (TP) = 1, True Negatives (TN) = 33, False Positives (FP) = 1, and False Negatives (FN) = 0. This suggests that all noisy samples were accurately identified (recall = 1.000), while only one normal sample was incorrectly classified as noisy (false positive rate = 2.94%). Consequently, the model prioritizes high sensitivity, ensuring no anomalies are overlooked, which is crucial for physiological monitoring applications where missing anomalies are more critical than occasional false alarms.

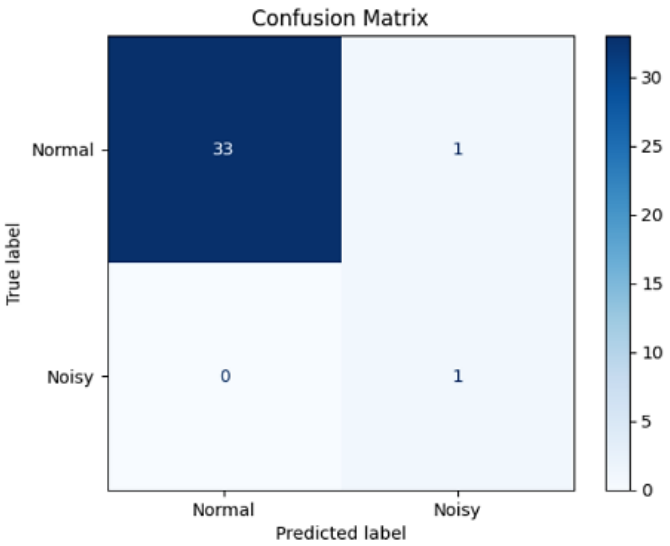

Figure 6. Confusion matrix of the proposed model

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 13, Issue 4 | Pages 704-711 | e-ISSN: 2347-470X

Figure 7 illustrates the precision and recall for different threshold values. The analysis of Precision and Recall resulted in an Average Precision (AP) score of 1.000, indicating that the model flawlessly prioritized anomalous samples over normal ones. This finding illustrates that the autoencoder achieves optimal precision-recall balance across various thresholds, further supporting the effectiveness of using reconstruction error as an anomaly score.

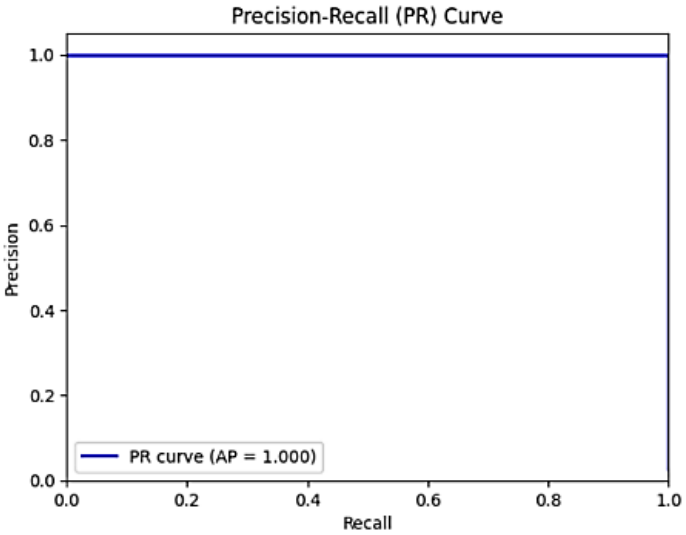

Figure 7. Precision and recall for different threshold values

The receiver operating characteristic curves (ROC) for the proposed approach are presented in *figure 8*. The ROC analysis of the autoencoder yielded an AUC of 1.000, indicating perfect differentiation between normal and abnormal segments. This performance, which is not dependent on a specific threshold, shows that the model consistently assigns higher reconstruction errors to anomalies compared to normal signals, confirming its capability to detect noise-induced deviations in the rPPG waveform accurately.

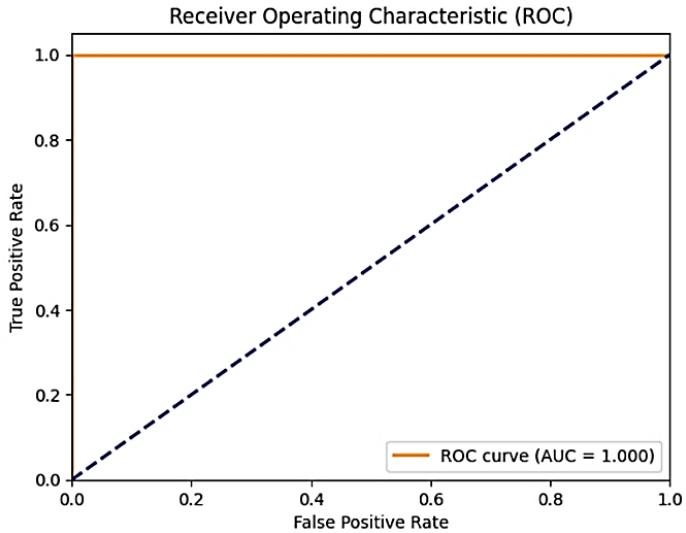

Figure 8. Receiver operating characteristic curves for the proposed approach

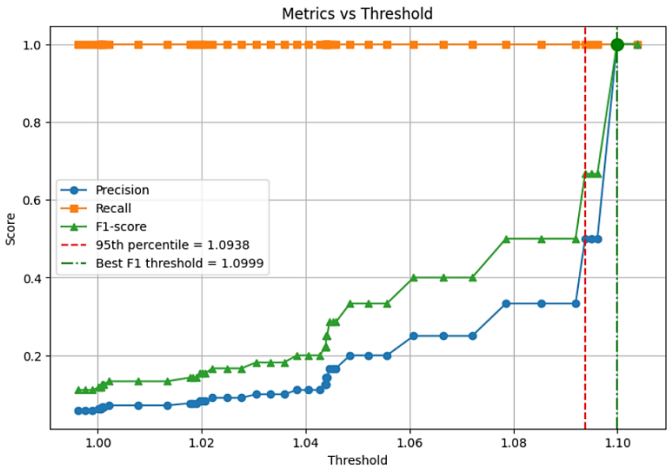

Figure 9. Sensitivity of anomaly detection to threshold choice

To evaluate the strength of the anomaly detection framework, we conducted a threshold sensitivity analysis by examining precision, recall, and F1-score over various reconstruction error thresholds in *figure 9*. The analysis showed that the model achieves its highest F1-score of 1.000 at a threshold of 1.0999, suggesting an almost perfect distinction between normal and anomalous segments. Our predetermined threshold, based on the 95th percentile of the error distribution (1.938), is slightly lower than this optimal point. This choice emphasizes recall, ensuring all anomalies are detected, while remaining very close to the F1-optimal threshold, thus preserving high precision. This closeness indicates that the model is not overly dependent on the threshold choice and that heuristic percentile-based thresholds can perform similarly to optimal data-driven cutoffs. Such consistency enhances the method's reliability for use in real-world monitoring situations where adaptive threshold adjustment may not always be possible.

4.4. Ablation Study

We performed an ablation study to compare a CNN-only model, a BiLSTM-only model, and our proposed CNN-BiLSTM Autoencoder. As illustrated in *table 3*, the hybrid model consistently outperformed the others, achieving the highest precision, recall, and F1-score, along with the lowest false positive rate. This highlights the complementary advantages of combining CNN and BiLSTM layers.

Table 3. Simulation result using testing data

DATASETS	MODELS	METRICS			
		PR (%)	Recall (%)	F1-score (%)	FPR (%)
COHFACE	CNN	60	86	79	4.7
	BiLSTM	50	95	80	2.02
	CNN-BiLSTM	**70**	**100**	**82**	**1.8**
UBFC-RPPG	CNN	70	75	80	2.06
	BiLSTM	60	89	77	3.3
	CNN-BiLSTM	**75**	**100**	**85**	**1.6**
PURE	CNN	50	86	50	3.9
	BiLSTM	50	99	59	3.05
	CNN-BiLSTM	**50**	**100**	**66.7**	**2.94**

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 13, Issue 4 | Pages 704-711 | e-ISSN: 2347-470X

4.5. Comparative Evaluation with other Methods

To assess the performance of the proposed Hybrid CNN–BiLSTM Autoencoder, we conducted a comparison with several leading rPPG anomaly detection and enhancement techniques. Despite the use of different architectures and evaluation criteria in prior research, our qualitative analysis indicates that the proposed model excels in anomaly discrimination, reconstruction accuracy, and real-time applicability.

Table 4 presents a summary of the main features and performance patterns of recent methods for anomaly detection and rPPG enhancement, in comparison to our proposed model. Unlike previous studies that primarily focus on signal enhancement or heart-rate estimation, our method prioritizes anomaly differentiation and reconstruction accuracy, providing both interpretability and suitability for real-time use.

Table 4. Proposed model comparison with existing works

Methods	Dataset (s)	Evaluation metrics	Key findings	Comparison to the proposed model
LSTM-based filter [5]	MMSE-HR VIPL-HR COHFACE	HR-MAE, RMSE	An LSTM-based filter is a better alternative to improve the heart rate measurement. moderate robustness to motion	Our model achieves comparable efficiency with higher robustness to motion and illumination artifacts.
DiffPhys [2]	UBFC-rPPG, PURE	SNR, HR Error	Enhances SNR under noise; improves pulse estimation in low-light scenarios.	Our model complements this by performing anomaly localization and correction before HR estimation.
Optimal Digital Filtering [16]	Own data	SNR, Filter Response	Improves signal conditioning but lacks adaptive temporal modeling.	Our model captures both spatial and temporal dependencies dynamically.
Proposed: CNN–BiLSTM Autoencoder	UBFC-rPPG PURE COHFACE	AUC, AP, F1, FPR, Recall	robust real-time anomaly detection and signal reconstruction.	Demonstrates superior separability, stability across thresholds, and near-real-time inference (24 Hz).

5. DISCUSSION

The Hybrid CNN–BiLSTM Autoencoder offers a balance between precision and computational speed, making it ideal for real-time remote photoplethysmography (rPPG) monitoring. With around 1.2 million parameters, the model can handle a 300-frame sequence in about 42 milliseconds on a GPU and approximately 130 milliseconds on a CPU, allowing for nearly instantaneous inference. This level of efficiency facilitates ongoing physiological monitoring with minimal delay.

Nonetheless, certain limitations persist. The model's validation was largely conducted using controlled datasets (UBFC-rPPG, PURE, COHFACE), which might not accurately reflect real-world clinical scenarios characterized by varying lighting, movement, or skin tones. Moreover, for continuous deployment, it would be essential to ensure robust calibration, privacy protection, and validation across diverse populations. Future efforts will concentrate on adaptive thresholding, integrating multimodal data, and optimizing for edge and mobile devices to improve practical usability.

6. CONCLUSION

In this study, we presented and validated the efficacy of a hybrid CNN-BiLSTM Autoencoder model tailored for dependable anomaly detection in rPPG time-series data. This model is designed to identify anomalies such as sensor noise and motion artifacts within the remote photoplethysmography signal. The procedure starts by segmenting the signal into multiple parts. Subsequently, convolutional layers are employed to capture localized spatial and temporal patterns. Ultimately, the bidirectional LSTM's ability is harnessed to model sequential dependencies over time, enabling the model to effectively learn representations of clean physiological signals. Quantitative analyses using reconstruction error thresholds demonstrated outstanding performance with the UBFC-rPPG, COHFACE, and the PURE datasets, distinguishing clean segments from noisy ones with high precision, recall, F1-score, and low false positive rates.

7. ACKNOWLEDGMENTS

This is to acknowledge the financial support of the Pan African University under the African Union for sponsoring this research.

REFERENCES

[1] Guangkun Nie, et al. (2024). A review of deep learning methods for photoplethysmography data. arXiv preprint arXiv:2401.12783.

[2] Shutao Chen, Kwan-Long Wong, Jing-Wei Chin, Tsz-Tai Chan, and Richard H. Y. So (2024). DiffPhys: Enhancing Signal-to-Noise Ratio in Remote Photoplethysmography Signal Using a Diffusion Model Approach. Bioengineering, 11(8), 743. DOI: 10.3390/bioengineering11080743.

[3] Lee, K., Kim, S., An, B., Seo, H., Park, S., and Lee, E. C. (2023). Noise-Assessment-Based Screening Method for Remote Photoplethysmography Estimation. Applied Sciences, 13(17), 9818.

[4] J. Speth, A. D. Bagheri, D. McDuff, and S. Hershey, "Hallucinated Heartbeats: Anomaly-Aware Remote Pulse Estimation," *arXiv preprint arXiv:2303.06452*, 2023.

[5] Botina-Monsalve, Deivid, Yannick Benezeth, and Johel Miteran. "Performance analysis of remote photoplethysmography deep filtering using long short-term memory neural network." BioMedical Engineering Online 21.1 (2022): 69.

[6] Varun Chandola, Arindam Banerjee, and Vipin Kumar (2009). Anomaly Detection: A Survey. ACM Computing Surveys (CSUR), 41(3), 1–58.

[7] Massaroni, C., Nicolo, A., Sacchetti, M., and Schena, E. (2020). Contactless Methods for Measuring Respiratory Rate: A Review. IEEE Sensors Journal, 21(11), 12821–12839.

[8] Lee, J., Kim, M., Park, H. K., and Kim, I. Y. (2020). Motion Artifact Reduction in Wearable Photoplethysmography Based on Multi-Channel Sensors with Multiple Wavelengths. Sensors, 20(5), 1493. DOI: 10.3390/s20051493.

[9] S.-Q. Liu and P. C. Yuen, "Robust Remote Photoplethysmography Estimation with Environmental Noise Disentanglement," *IEEE Transactions on Image Processing*, vol. 32, pp. 2570–2583,2023.

[10] Choi, K., Yi, J., Park, C., and Yoon, S. (2021), Deep Learning for Anomaly Detection in Time-Series Data: Review, Analysis, and Guidelines. IEEE Access, 9, 120043–120065. DOI: 10.1109/ACCESS.2021.3107975.

[11] Fahim, M., and Sillitti, A. (2019). Anomaly Detection, Analysis and Prediction Techniques in IoT Environment: A Systematic Literature Review. IEEE Access, 7, 81664–81681. DOI: 10.1109/ACCESS.2019.2924225.

[12] R. Yousefi, M. Nourani, Separating arterial and venous-related components of photoplethysmographic signals for accurate extraction of oxygen saturation and respiratory rate, IEEE J. Biomed. Health Inf. 19 (3) (2014) 848–857.

[13] A. Gudi, M. Bittner, J. van Gemert, Real-time webcam heart-rate and variability estimation with clean ground truth for evaluation, Appl. Sci. 10 (23) (2020) 8630.

[14] G. de Haan and V. Jeanne, "Robust Pulse Rate from Chrominance-Based rPPG," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 10, pp. 2878–2886, 2013.

[15] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation," *Optics Express*, vol. 18, no. 10, pp. 10762–10774, 2010.

[16] S. Guler, A. Golparvar, O. Ozturk, H. Dogan, and M. K. Yapici, "Optimal digital filter selection for remote photoplethysmography (rPPG) signal conditioning," *Biomedical Physics & Engineering Express*, vol. 9, no. 2, p. 025003, Feb. 2023.

[17] Stricker, R., Müller, S., Gross, H.-M. "Non-contact Video-based Pulse Rate Measurement on a Mobile Service Robot," in: Proc. 23rd IEEE Int. Symposium on Robot and Human Interactive Communication (Ro-Man 2014), Edinburgh, Scotland, UK, pp. 1056-1062, IEEE 2014.

[18] Wang, Y., Perry, M., Whitlock, D., and Sutherland, J. W. (2022). Detecting Anomalies in Time Series Data from a Manufacturing System Using Recurrent Neural Networks. Journal of Manufacturing Systems, 62, 823–834. DOI: 10.1016/j.jmsy.2022.01.011.

[19] R. Chalapathy and S. Chawla, "Deep Learning for Anomaly Detection: A Survey," Jan. 2019, Accessed: Feb. 23, 2023. [Online]. Available: http://arxiv.org/abs/1901.03407.

[20] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," *Optics Express*, vol. 16, no. 26, pp. 21434–21445, 2008.

[21] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural Networks*, vol. 13, no. 4-5, pp. 411–430, 2000.

[22] S. Bobbia, R. Macwan, Y. Benezeth, A. Mansouri, J. Dubois, "Unsupervised skin tissue segmentation for remote photoplethysmography", Pattern Recognition Letters, 2017.

[23] Pankaj, Malhotra. "LSTM-based encoder-decoder for multi-sensor anomaly detection." *CoRR* 1607 (2016).

[24] M. Mmadi, G. N. Kamucha, C. Maina, "Contactless Algorithm for Identifying the Facial Area Containing the Most Vital Sign-Related Information Using Machine Learning," Harbin Engineering Journal, vol.46, no. 05, pp. 89-99, May. 2024, https://harbinengineeringjournal.com/index.php/journal/article/view/4200/2375.

[25] Keras Docs: Anomaly Detection in Time Series Using Autoencoders, Accessed: May. 10, 2025. [Online]. Available: https://keras.io/examples/timeseries/timeseries_anomaly_detection/

[26] Nawaz, Ali, Shehroz S. Khan, and Amir Ahmad. "Ensemble of autoencoders for anomaly detection in biomedical data: A narrative review." *IEEE Access* 12 (2024): 17273-17289.

[27] Guillaume Heusch, André Anjos, Sébastien Marcel, "A reproducible study on remote heart rate measurement", arXiv, 2016. [online]. Available: http://publications.idiap.ch/index.php/publications/show/3688

[28] "PURE dataset," [Online]. Available: http://www.tu-ilmenau.de/neurob/data-sets/pulse. [Accessed: Apr. 30, 2024].

© 2025 by the Moussa Mmadi, George Kamucha, and Ciira wa Maina. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).