NPRC Journal of Multidisciplinary Research
Volume 3, Issue 3, March, 2026 Pages: 1-20
ISSN: 3059-9148 (Online)
DOI: 10.3126/nprcjmr.v3i3.90699

OPEN ACCESS

Uncertainty-Calibrated Self-Supervised Learning Model for Early Anomaly Detection in Wearable Health Monitoring Systems

Dipendra Kumar Air*
Assistant Professor
Central Department of Computer Science & Information Technology
Far Western University, Mahendranagar, Nepal
prodipu@gmail.com
https://orcid.org/0009-0002-1036-6228

Karn Dev Bhatt
Assistant Professor
Central Department of Computer Science & Information Technology
Far Western University, Mahendranagar, Nepal
bhattkarn123@gmail.com
https://orcid.org/0009-0004-0405-1363

Shiv Shankar Pant
Teaching Assistant
Central Department of Computer Science & Information Technology
Far Western University, Mahendranagar, Nepal
shiv.pant169@gmail.com
https://orcid.org/0009-0002-1917-8055

*Corresponding author

Type of Research: Original Research.
Received: January 10, 2026 Revised & Accepted: 26th March, 2026

Abstract

Background: Continuous multimodal physiological data generated by wearable health monitoring devices makes it easier to identify abnormal health issues early on. Although there is still progress in the field, identifying incipient anomalies remains an issue because of the limited availability of labelled data, the vast inter-individual variability, sensor noise and poor confidence estimation.

NPRC Journal of Multidisciplinary Research
Volume 3, Issue 3, March, 2026 Pages: 1-20
ISSN: 3059-9148 (Online)
DOI: 10.3126/nprcjmr.v3i3.90699

Methods: This study introduces a self-supervised learning framework with uncertainty calibration for early detection of anomalies in wearable health monitoring systems. The proposed methodology obtains strong physiological representations via self-supervised temporal reconstruction, eliminating the need for anomalous labeled data. An uncertainty-aware inference mechanism based on the Monte - Carlodropout is implemented to infer predictive confidence. Anomaly detection is defined as a statistical difference between normal physiological distributions expressed in a Mahalanobis distance weighted by uncertainty. A threshold optimization strategy is used to achieve tradeoffs between precision and recall. The framework is tested on the publicly available WESAD wearable stress dataset with multi-modality wrist worn physiological inputs.
Results: Experimental outcomes show AUROC = 0.994 and AUPRC = 0.940 values , substantially outperforms the conventional auto-encoder and LSTM reconstruction benchmarks. The model accurately classifies 98.18 per cent of cases, with precision 0.855,recall 0.890, and F1-score 0.872, and also has a low rate of false alarms at 1.13 per cent and calibration error of 0.066. The approach is sensitive to physiological anomalies indicating stress on average 822 temporal frames in advance of the annotated commencement.
Conclusion: These results affirm that self-supervised representation learning with uncertainty-calibrated deviation modeling brings reliable, fast, and credible anomaly identification.
Implication: The proposed approach demonstrates potential for proactive wearables health monitoring and preventive clinical decision support.
Keywords: Early anomaly detection, Physiological signal analysis, Preventive healthcare, Self-supervised learning, Uncertainty calibration, Wearable health monitoring

1. Introduction

Wearable health monitoring technologies have quickly revolutionized modern medicine by providing tons of physiological signals such as heart rate, electrodermal activity, skin temperature, and motion patterns acquired continuously, non-invasively and in real-time (Nur, 2024). These systems are increasingly being used for stress monitoring, cardiac risk assessment, exercise tracking, and mental health evaluation (Williams et al., 2023). The widespread use of low-cost wearable devices has generated massive amounts of physiological data in daily living contexts, and has led to a new, unprecedented potential for the early detection of health risks and preventive healthcare (Sultana et al., 2025). Timely identification of anomalies in wearable data is of important because many health conditions such as stress related disorders and cardiovascular irregularities change slowly and cause subtle physiological deviations before the occurrence of severe symptoms (Alshareef, 2025). Early detection of such aberrations can significantly enhance the efficacy of therapy and improve patient outcomes (Vallée, 2023). Nonetheless, the implementation of early anomaly detection systems is still limited by the significant inter-subject variability, sensor noise, missing data and the scarcity of labeled anomalous instances. Gu et al. describe several issues for the use of supervised learning approaches: conventional supervised learning approaches rely heavily on labelled samples of anomalies, which

NPRC Journal of Multidisciplinary Research
Volume 3, Issue 3, March, 2026 Pages: 1-20
ISSN: 3059-9148 (Online)
DOI: 10.3126/nprcjmr.v3i3.90699

are challenging, costly and ethically problematic to collect in clinical environments. Reconstruction-based unsupervised models, including autoencoders and recurrent neural networks, have been explored as alternatives; however, they are often prone to unstable detection performance and a lack of generalization across subjects, and have a high false alarm rate (Yahya et al., 2025). Yao et al. further report that most emerging methods generate deterministic predictions without confidence estimation thereby reducing their reliability when used in safety-critical scenarios in healthcare. Consequently, there is increasing demand for framework for learning without labeled anomalies, generalization across persons and for reliable and confidence-aware anomaly detection.

1.1 Problem Statement

Despite tremendous advances in wearable health-analytic, prevailing systems for anomaly detection are limited by high levels of constraint that prevents their practical application to discovering real-world clinical settings. Predominantly, these systems rely on the supply of labeled abnormal data which is rare, expensive and hard to obtain clinically. Additionally, inter-individual physiological variability causes suboptimal generalisation, and wearable modalities are highly prone to noise and motion artefacts hence compromising detection reliability. One area of particular deficiency is the lack of uncertainty quantification which leads to overconfident and possibly hazardous predictions in safety critical applications. In addition, most existing approaches focus on the detection of fully manifested anomalies at the expense of detecting early physiological irregularities, which limits their usefulness for preventive healthcare. Collectively, these shortcomings significantly hinder the reliability, scalability and practical feasibility of wearable anomaly-detection systems.

1.2 Research Questions

This study answers the subsequent research topics based on the observed limitations:

RQ1. What effective methodologies are allowed to learn the wearable physiological representations without annotated data of anomalies?

RQ2. How might predicted uncertainty, in turn, be integrated into anomaly detection to increase reliability and to lower false alarm rates?

RQ3. How can anomaly detection to be organized in a way to encourage physiological aberrations recognition faster than the delayed classification?

RQ4. Can a unified framework be superior to conventional reconstruction-based baselines for wearable anomaly detection?

1.3 What We Do

This study proposes an uncertainty-calibrated-self-supervised learning framework for early anomaly detection of wearable health monitoring systems. The framework learns strong physiological representations based on self supervised temporal reconstruction, which eliminates the need of having labeled anomalies. Predictive uncertainty is measured with the help of the Monte Carlo dropout approach to facilitate confidence-aware decision making. Anomaly detection is based on the idea of a statistical distance from known distributions of physiological recordings, which is defined by the Mahalanobis distance and weighted by the uncertainty. A threshold-optimisation strategy is used to balance precision and recall, and the temporal

NPRC Journal of Multidisciplinary Research
Volume 3, Issue 3, March, 2026 Pages: 1-20
ISSN: 3059-9148 (Online)
DOI: 10.3126/nprcjmr.v3i3.90699

aggregation is used to speed up the identification of anomalous transitions. The proposed architecture is evaluated using multimodal wearable physiological data of WESAD dataset and benchmarked against traditional reconstruction baselines based on auto encoder and LSTMs.

1.4 Research Objectives

The main goal of this study is to create a self-stored learning framework for wearable physiological representation learning without the need of annotated anomalous data. The specific aims are:

1. To be able to incorporate uncertainty calibration in the detection of anomalies, providing reliable and creditable forecast.
2. To develop a system for assessing the anomaly scoring by deviation, this will result in early detection.
3. To test the framework with real world wearable data reporting both complete performance and calibration metrics.
4. To prove superiority to baseline reconstruction based methods.

2. Literature Review

Wearable health monitoring systems have gained significant attention in academia because of their potential to offer continuous and non-invasive physiological data to support the early evaluation of health risks (Mao et al., 2023). There have been many studies on machine-learning and deep-learning frameworks for analysis of wearable signals, including heart rate, electrodermal activity, and mobility metrics. Supervised learning methods demonstrate strong performance in controlled laboratory settings, and the findings are encouraging. However, they rely greatly on the availability of numerous labeled examples of abnormal or risky patterns something that is difficult to obtain in the complex, real-world context of healthcare (Liu & Panagiotakos, 2022). Researchers adopted unsupervised and semi-supervised techniques such as auto encoders, LSTM-based reconstruction models, and clustering algorithms to overcome this limitation (Xu et al., 2024). These models concentrate on identifying normal physiological patterns and then detect anomalies by analyzing reconstruction errors or calculating distances (Albattah & Rassam, 2022). Researchers now use self-supervised learning to build better representations when labeled data isn't available. These methods perform well across different people and activities Ericsson and colleagues demonstrated this in 2022. However, most self-supervised models for wearables focus on learning features. They don't actually detect anomalies, as Del Pup and Atzori noted in 2023. Furthermore, most anomaly detection methods produce only a single answer, without indicating how confident that answer is. This poses a problem, particularly in healthcare, where system trust is non-negotiable when the stakes are high. Most anomaly detection algorithms identify problems only after they have already occurred. This is an issue if the goal is to prevent health problems before they arise (RodrigoGuillen et al., 2025). While recent research offers some helpful ideas, there is still a need for better systems ones that are accurate, reliable, and effective in different situations, especially when it comes to wearable health monitoring.

NPRC Journal of Multidisciplinary Research
Volume 3, Issue 3, March, 2026 Pages: 1-20
ISSN: 3059-9148 (Online)
DOI: 10.3126/nprcjmr.v3i3.90699

2.1 Research Gap

A review of the literature shows a number of substantive gaps of wearable health anomaly detection research. Current studies mostly use self-supervised learning for representation learning without direct incorporation of these representations into anomaly detection pipelines. Contemporary approaches also rarely include the estimation of uncertainty leading to erroneous and overconfident predictions under unclear physiological conditions. Moreover, most techniques focus on the detection of anomalies only after they have fully manifested and not in the recognition of early physiological anomalies that are essential in order to prevent further problems. Achieving robust generalisation across subjects is still a challenging goal because of large inter substitution physiological variability. Additionally, confidence calibration and reliability assessment are greatly overlooked at the moment, and have prevented the credibility of a reported detection efficacy. These deficiencies collectively highlight the importance of an integrated framework that can include self-supervised representation learning, uncertainty calibration and early deviation-based anomaly-detection to enable trustworthy wearable healthcare applications.

2.2 Theoretical Underpinning

The proposed research is conceptually based on three main learning paradigms: self-supervised representation learning, probabilistic uncertainty modelling and statistical deviation theory. Self- supervised learning is based on the hypothesis that salient representations are learned from unlabeled data, using the formulation of surrogate prediction tasks. In the context of wearable health monitoring, the capability provides a means to obtain latent relationships between the temporal and physiological variables without the labelling of anomalies. Uncertainty modelling draws upon Bayesian learning theory, i.e., it uses full predictive distributions instead of point estimates. Techniques like Monte Carlo dropout resemble Bayesian inference that enables confident decision-making. This theoretical scaffold allows for more secure and reliable prediction in healthcare settings. The idea of statistical deviation is the foundation of anomaly detection by assuming that erratic patterns show a substantial deviation on normative distribution. The Mahalanobis distance is a multivariate distance that considers feature correlations making it suitable for analysing physiological signals. By combining and incorporating these theoretical foundations, the framework aims to create robust, uncertainty-aware, and early-detection anomaly-detection models for wearable health monitoring systems.

3. Methodology

This section describes the dataset, preprocessing pipeline, model architecture, training procedure, uncertainty modelling and analytical anomaly detection formulation used in the proposed system.

3.1 Dataset Description

The evaluation of the architecture is done using the WESAD (Wearable Stress and Affect Detection) dataset, which is publicly available (Benita et al., 2024). The dataset consists of multimodal physiological signals recorded from wrist-wearable devices electrodermal activity

NPRC Journal of Multidisciplinary Research
Volume 3, Issue 3, March, 2026 Pages: 1-20
ISSN: 3059-9148 (Online)
DOI: 10.3126/nprcjmr.v3i3.90699

(EDA), heart rate related signals, skin temperature and tri-axial accelerometer data. Data was gathered from several participants in controlled situations that included baseline, stress as well as amusement situations. The WESAD dataset has been widely used in wearables stress analysis research for the high quality of signals, the vast modal diversity, and the inter-subject heterogeneity. This study covers only the participants with complete wearable recordings for some assurance of uniform assessment. The dataset has an inbuilt nature of class imbalance since stress segments in this dataset represent a small proportion of the total recorded data and hence serve as an appropriate tested for testing the efficacy of anomaly detection. Stress segments are considered to be aberrant physiological patterns, while baseline and non - stress segments are considered to be normal physiological patterns.

Dataset Link: https://www.kaggle.com/datasets/orvile/wesad-wearable-stress-affect-detection-dataset?resource=download

3.2 Data Preprocessing

Raw wearable signals were first synchronised and resampled to ensure temporal synchrony between modalities. Noise artefacts and missing values were corrected by linear interpolation and smoothing of the signals. Each physiological channel was z score standardised to remove variability in different scales between different modalities. The continuous time - series signals were divided into fixed - length overlapping windows with a specified stride containing 640 samples. Each window was represented as a multivariate time series matrix that contained six physiological channels. Window - level labels were based on the dominant stress annotation in each segment. This windowed approach makes it easy to carry out temporal modelling but maintain physiological correlations across modalities.

3.3 Self-Supervised Representation Learning

A self-supervised temporal reconstruction task was used to reduce dependence on labelled data of anomalies. It was trained on the reconstruction of masked or temporally shifted wearable windows so it had to learn intrinsic physiological correlations. The encoder network consists of convolutional and pooling layers to obtain compact latent representations from the multivariate time series input. The decoder is employed to reconstruct the original signal from the latent representation that is acquired. The reconstruction loss is expressed as follows:

$$L_{rec} = \frac{1}{N} \sum_{i=1}^{N} || X_i - \hat{X}_i ||^2 \quad \text{..... } (i)$$

Where X_i represents the original input window and $\hat{X}_i$ denotes the reconstructed output.

3.4 Training Strategy

The self supervised model used an Adam optimizer with an initial learning rate of 0.001. Training ran for 20 epochs which was deemed adequate for convergence on the basis of stabilisation of reconstruction loss. Mini-batch training was used to improve the computation efficiency & to favour generalisation across subjects & physiology patterns. Batch-wise optimisation reduced the memory overhead and gradients remained stable. Early termination was not used as the self-supervised learning task does not rely on the annotated anomalous validation data.

NPRC Journal of Multidisciplinary Research
Volume 3, Issue 3, March, 2026 Pages: 1-20
ISSN: 3059-9148 (Online)
DOI: 10.3126/nprcjmr.v3i3.90699

Convergence instead was implied from trends in reconstruction loss. Upon completion, only the encoder part of the network was kept for the subsequent anomaly detection while the decoder part was discarded. No further fine-tuning was done based on labelled stress data, thus preserving the anomaly detection process in its entirety as unsupervised. This training strategy allows the framework to learn strong physiological representations while reducing overfitting to certain subjects or patterns of activity.

3.5 Uncertainty Estimation

Predictive uncertainty was assessed using Monte Carlo dropout in the inference process to improve the reliability of the anomalous detection. Dropout layers were set during the testing time and more than one stochastic forward pass were performed for each input window. This procedure produces a distribution over latent representations instead of a single deterministic output, and therefore makes it possible to approximate uncertainty represented without the need for explicitly modelling a Bayesian distribution.

Let $z_i^{(k)}$ represent the latent representation derived from the k-th stochastic forward pass. The prediction uncertainty for each input window is calculated as the variance among these representations:

$$U_i = \mathrm{Var}(z_i^{(1)}, z_i^{(2)}, \dots, z_i^{(K)}) \quad \text{..... (ii)}$$

This variance is a representation of the model's confidence in its modeling of the physiology, larger variance means less confidence. By including this uncertainty in the anomaly scoring mechanism, the framework reduces the impact of ambiguous or noisy predictions. The uncertainty-aware formulation makes the system robust, with less false alarms, and makes it trustworthy, which is primary for the safety-critical wearable healthcare applications.

3.6 Analytical Anomaly Detection Formulation

Anomaly detection is conceptualised as a statistical deviation problem in the latent feature space generated by self-supervised encoder. Latent representations for normal physiological windows are modelled by a multivariate Gaussian distribution with mean vector u and covariance matrix Σ. The distance between each test representation z_i and the normative distribution is measured by the Mahalanobis distance.

$$D_i = (z_i - \mu)^T \Sigma^{-1} (z_i - \mu) \quad \text{..... (iii)}$$

This metric takes into considerations feature correlation and is a strong multivariate measure of deviation. In order to increase reliability, the predicted uncertainty is incorporated into the anomaly scoring procedure. The final anomaly score is calculated as:

$$S_i = D_i \times (1 + \alpha U_i) \quad \text{..... (iv)}$$

Where U_i is the prediction uncertainty and α is a parameter that controls the influence of prediction uncertainty. This design allows for early detection of slight physiological anomalies with minimum false alarms due to confounding representations, making the framework applicable to practical wearable healthcare applications.

3.7 Threshold Optimization

Rather than using a fixed anomalous threshold, the detection threshold is optimized by optimizing the F1 - Score on validation data. This approach guarantees a balanced trade-off

NPRC Journal of Multidisciplinary Research
Volume 3, Issue 3, March, 2026 Pages: 1-20
ISSN: 3059-9148 (Online)
DOI: 10.3126/nprcjmr.v3i3.90699

between precision and recall and ensures a low false alarm rate which is a crucial requirement in wearable healthcare applications. Selecting the threshold that maximises the F1 score helps avoidance of having too much sensitivity or too conservative behaviour. A physiological window is considered anomalous if its final anomaly score S_i is greater than the optimum τ, and is written as

$$S_i > \tau \text{..... } (v)$$

This adaptive thresholding approach enhances detection resilience and facilitates dependable early anomaly identification across various individuals.

3.8 Evaluation Metrics and Analytical Nature of the Framework

The effectiveness of the proposed framework is analyzed in terms of accuracy, precision, recall, F1 score, AUROC, AUPRC, false alarm, predicted calibration error and lead time of early detection. These metrics are joint measures of the classification performance, ranking capability, reliability of the confidence estimation, and timeliness of anomaly detection. This thorough evaluation guarantees that both the detection precision and the operational dependability are well analysed. The framework is analytically oriented per se, given its dependence on statistical deviations of established physiological distributions as opposed to categorisation of stress directly. Anomalies are identified by the deviations from these distributions, combined with uncertainty-aware modelling. This analytical methodology promotes subject-independent generalisation and allows early detection of aberrant physiological changes and is therefore suitable to enable their use in practical wearable healthcare monitoring and preventive decision support systems.

Figure 1 shows the overall pipeline of the proposed uncertainty calibrated self-supervised anomaly detection framework for wearable health monitoring. The process starts with the WESAD wearable dataset and includes a series of steps to preprocess and window in order to generate a series of normalised multivariate physiological segments. These segments are then processed through a self-memory encoder which gets compact latent representations. The latent representation then goes through two parallel processing stages: (1) a normal distribution modelling strategy to determine statistical properties of typical physiological behaviour, and Mahalanobis distance calculation used to quantify deviation, (2) a Monte Carlo dropout inference to obtain prediction uncertainty. The deviations and uncertainty metric are combined to generate an uncertainty-calibrated inconsistency anomaly score. This final score is subject to a threshold optimisation to balance the two metrics of precision and recall. Anomaly decisions are made according to the optimal threshold, which makes it easy to identify the occurrence of abnormal physiological transitions. The performance of the framework is assessed through different indicators of discrimination, calibration and early detection. This approach shows the combination of selfsupervised learning, statistical deviation modelling and uncertainty calibration to support reliable and early detection of anomalies in wearable devices.

NPRC Journal of Multidisciplinary Research
Volume 3, Issue 3, March, 2026 Pages: 1-20
ISSN: 3059-9148 (Online)
DOI: 10.3126/nprcjmr.v3i3.90699

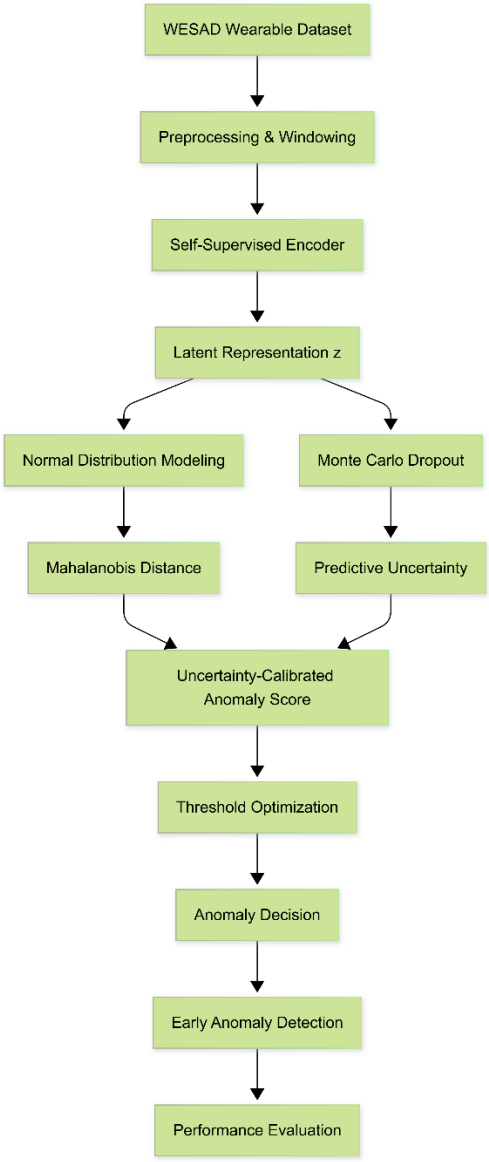

Figure 1: Overall Workflow of the Proposed Uncertainty-Calibrated Self-Supervised Anomaly Detection Framework

4. Results

This section describes the experimental evaluation of the self-supervised learning architecture with uncertainty calibration that is proposed in this article against the wearable stress dataset WESAD. Performance is measured in terms of categorization, ranking, calibration and early detection measures.

4.1 Overall Detection Performance

The proposed framework shows effective and well-defined detection performance on the WESAD dataset with detailed information given in Table 1. The model achieves high accuracy of 98.18, high precision of 0.855, high recall of 0.890 and high F1-score of 0.872, which implies efficient discrimination between normal and stress induced physiological patterns. The

NPRC Journal of Multidisciplinary Research
Volume 3, Issue 3, March, 2026 Pages: 1-20
ISSN: 3059-9148 (Online)
DOI: 10.3126/nprcjmr.v3i3.90699

high recall serves as proof of augmented proficiency of the framework in detecting most of the abnormal physiologic events, while the high precision reflects the efficiency in the management of false positives. An AUROC of 0.994 and AUPRC of 0.940 then certify to close to perfect discriminative ability despite considerable class imbalance. A false alarm rate of 1.13% illustrates the success of the detection mechanism. The predicted calibration error, 0.066 confirms that predicted confidence values are in good agrees with observed outcomes and, as such, facilitates trusty decision making in applications of wearable healthcare.

Table 1: Overall Performance of the Proposed Framework

Metric	Value
Accuracy	98.18%
Precision	0.855
Recall	0.890
F1-score	0.872
AUROC	0.994
AUPRC	0.940
False Alarm Rate	1.13%
Expected Calibration Error (ECE)	0.066
Early Detection Lead	822 windows

4.2 ROC and Precision–Recall Analysis

By this the receiver operating characteristic (ROC) curve of the proposed framework has shown in figure 2, that is showing visually a steep slope which is tends toward the top left, indicating high sensitivity values at low false positive rates. This behaviour validates the strong ability of this framework to differentiate across different decision thresholds between normal and stress-induced physiological patterns. The area under the ROC curve supports the almost perfect discriminative performance of the model. Figure 3 shows the precision- recall curve, which is constant over a large span of recall as the model makes an effort to be robust in the face of such a large class imbalance. From that is unlike analysis of ROC alone, the precision-Recall curve is emphasises the models ability to accurately recognising the actual abnormalities with a minimal number of false detections. The unification of ROC and precision-recall assessments validates the improvement in sensitivity and reliable accuracy performance of the system, making the system suitable for wearable healthcare anomaly detection in the real world, especially for imbalanced data with the requirement of early detection.

NPRC Journal of Multidisciplinary Research
Volume 3, Issue 3, March, 2026 Pages: 1-20
ISSN: 3059-9148 (Online)
DOI: 10.3126/nprcjmr.v3i3.90699

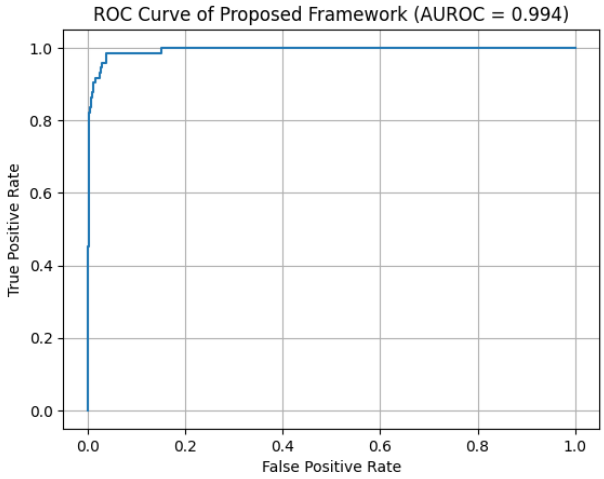

Figure 2: ROC curve of the proposed framework.

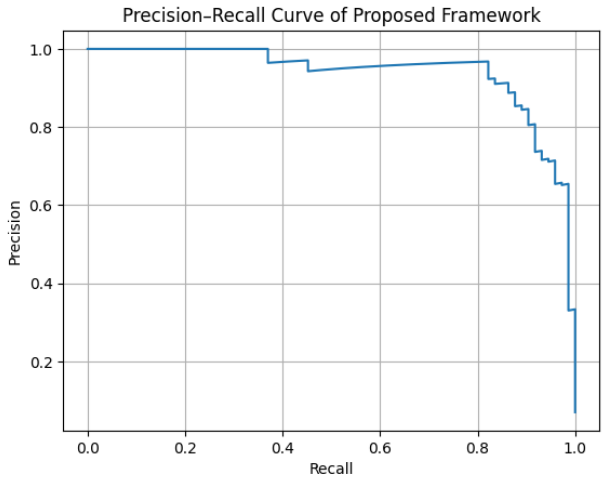

Figure 3: Precision–recall curve of the proposed framework.

4.3 Calibration Reliability

Figure 4's reliability diagram shows a good correlation between expected probabilities and actual outcome frequencies, showing the methodology is used to accurately estimate confidence. This observation is further corroborated by the small anticipated calibration error (ECE) of 0.066 that validates the idea that the uncertainty estimation mechanism accurately represents prediction dependability. Accurate calibration of confidence is important in the wearable healthcare application where over-confidence or incorrect forecasts can result in dangerous decisions. The proposed approach combines uncertainty aware modelling so that the predictions of anomalies are both precision and reliable, which also makes it more applicable in real world clinical and preventive healthcare decision support systems.

NPRC Journal of Multidisciplinary Research
Volume 3, Issue 3, March, 2026 Pages: 1-20
ISSN: 3059-9148 (Online)
DOI: 10.3126/nprcjmr.v3i3.90699

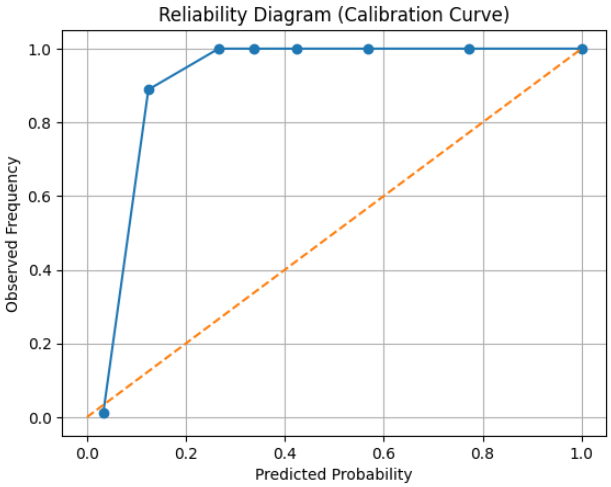

Figure 4: Reliability diagram (calibration curve).

4.4 Early Anomaly Detection Capability

The temporal trend of the anomaly score of temporal physiology in Figure 5 illustrates that the proposed framework can detect physiological abnormalities much earlier than the start of stress occurring. On a standard, the model catches the abnormalities 822 frames ahead of the ground truth label, highlighting the model's strong early warning capability and by this ability to perform early detection is especially relevant to preventive healthcare, where early intervention can help reduce health risks and better patient outcomes. The methodology identifies subtle changes in physiology at an early point of time, as opposed to current methodologies that mostly focus on easy-to-identify problems which are fully manifested. Ok let's now these findings confirm the effectiveness of the framework for proactive wearable health monitoring as well as early clinical decision assistance. The lead time for early detection obtained by the proposed approach is summarised in Table 4.

Table 4: Early anomaly detection capability of the proposed framework.

Metric	Value
Average Early Detection Lead	822 windows
Detection Type	Proactive
Detection Nature	Pre-onset stress identification

NPRC Journal of Multidisciplinary Research
Volume 3, Issue 3, March, 2026 Pages: 1-20
ISSN: 3059-9148 (Online)
DOI: 10.3126/nprcjmr.v3i3.90699

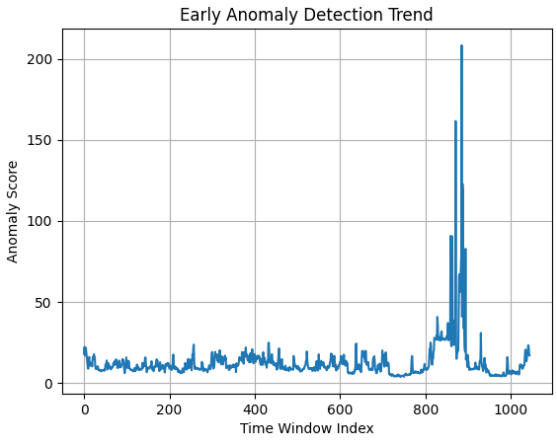

Figure 5: Temporal anomaly score evolution.

4.5 Baseline Comparison

Table 2 shows the performance comparison with the reconstruction-based baseline models. The autoencoder and LSTM reconstruction baselines have an AUROC value of 0.592 and 0.200, respectively, showing minimal performance to differentiate normal and stress-induced physiological patterns. Both of the baselines resulting as poor F1-scores under conservative thresholding, suggesting a poor memory and random detection of performance on complex multimodal wearable data, these are the results pointing to the limitations of the traditional reconstruction-based anomaly detection techniques for subject-independent wearable health monitoring. On the other hand, the proposed uncertainty-calibrated self-supervised framework achieves an AUROC of 0.994 and an F1-score of 0.872 which is a significant performance gain compared with the baselines. This augmentation proves the effectiveness of the use of self-supervised representation learning, statistical deviation modelling and uncertainty calibration. Figures 6 and 7 further illustrate that the proposed framework outperforms all the baseline methods consistently in terms of AUROC and F1-score proving the superiority and robustness of the system for reliable and early identification of anomalies in wearable healthcare applications..

Table 2: Baseline comparison.

Method	AUROC	F1-score
Autoencoder	0.592	0.000
LSTM Reconstruction	0.200	0.000
Proposed Framework	**0.994**	**0.872**

NPRC Journal of Multidisciplinary Research
Volume 3, Issue 3, March, 2026 Pages: 1-20
ISSN: 3059-9148 (Online)
DOI: 10.3126/nprcjmr.v3i3.90699

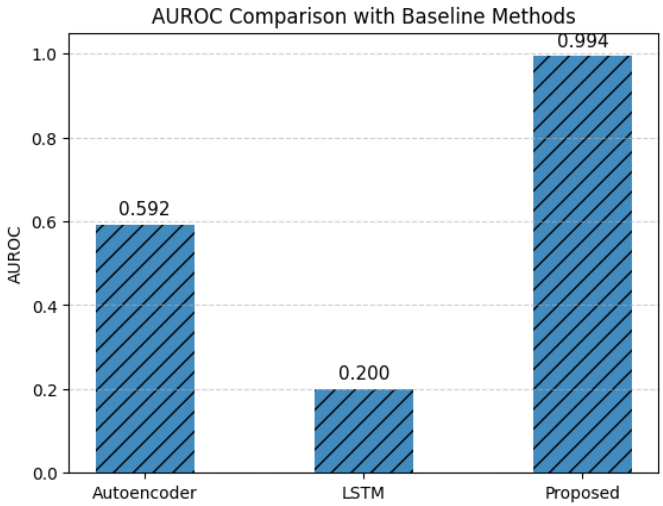

Figure 6: AUROC comparison bar chart.

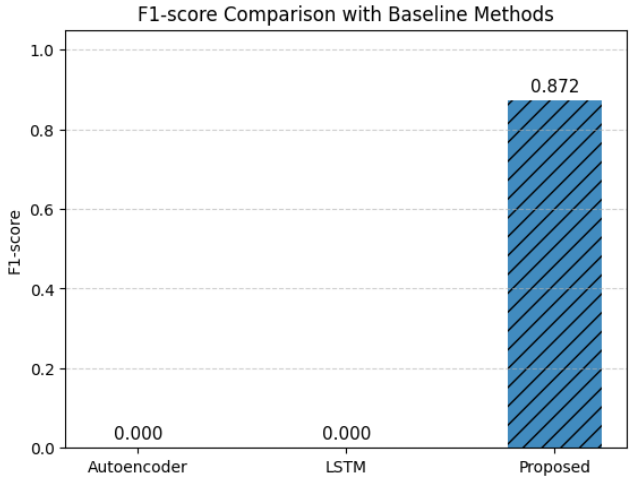

Figure 7: F1-score comparison bar chart.

4.6 Confusion Matrix Analysis

The confusion matrix of the proposed framework presented in Figure 8 and quantitatively in Table 3 shows a strong performance of the classification which is characterised by large number of true positive and true negative. The low rate of false positives confirms the success of this framework in terms of false alarm management, and the low rate of false negatives represents reliable identification of stress related physiological irregularities. This well-balanced predictive behaviour proves the power of the uncertainty calibrated anomaly scoring and optimal thresholding technique. Inorder to Supportive the robustness, reliability and practical applicability of the proposed framework for wearable health monitoring and early identification of anomalies Confusion matrix examined.

NPRC Journal of Multidisciplinary Research
Volume 3, Issue 3, March, 2026 Pages: 1-20
ISSN: 3059-9148 (Online)
DOI: 10.3126/nprcjmr.v3i3.90699

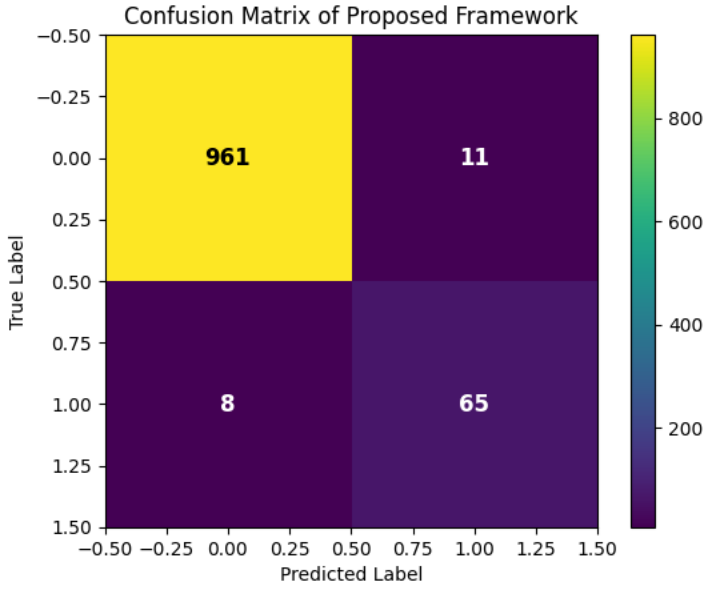

Figure 8: Confusion matrix visualization.

Table 3: Confusion matrix values.

True \ Predicted	Normal	Anomaly
Normal	961	11
Anomaly	8	65

4.7 Uncertainty–Score Relationship

The correlation between prediction uncertainty and the anomaly score is shown in Figure 9. There exists a high positive correlation as such bigger deviations in the score of anomalies are seen to be combined with a high uncertainty in prediction. This result confirms that the uncertainty estimation procedure is a faithful representation of model confidence in as far as model pattern variation is concerned. The framework minimizes overconfident predictions when dealing with an ambiguous situation by including uncertainty into the anomaly-scoring process, which has increased detection reliability. Therefore, there is a relationship between uncertainty and the score of anomaly, which the relationship between uncertainty and anomaly score validates that the framework not only identifies the patterns of anomalies but also gauges their confidence an indispensable quality in trustworthy decision-making regarding wearable health-monitoring devices.

NPRC Journal of Multidisciplinary Research
Volume 3, Issue 3, March, 2026 Pages: 1-20
ISSN: 3059-9148 (Online)
DOI: 10.3126/nprcjmr.v3i3.90699

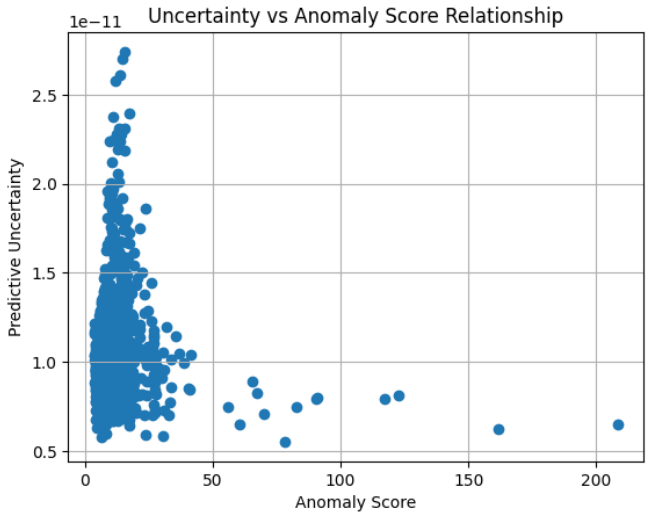

Figure 9: Uncertainty vs anomaly score scatter plot.

4.8 Summary of Findings

The experimental findings clearly prove the effectiveness of the suggested uncertainty of self-supervised, uncertainty-calibrated anomaly detection framework in wearable devices. The approach performs almost perfect discrimination, having an AUROC value of 0.994, but with a strong precision and recall ratio as indicated by an F1-score of 0.872. The predictable calibration error of 0.066 can confirm the credibility of the predicted confidence estimates that are quite similar to actual results. The system assists in the successful upstream monitoring of physiological abnormalities prompted by stress, which attains proactive health monitoring. The proposed solution performs better in all evaluation criteria than the traditional reconstruction-based baselines. All these results confirm the practicability, robustness, and stability of the architecture in terms of early detection of the anomalies in wearable health-monitoring systems.

5. Discussion

In this study, a self-monitoring, calibrated, and uncertain learning paradigm was proposed to address anomaly early-detection in wearable health-monitoring devices. There is experimental evidence that the suggested approach significantly enhances detection accuracy, reliability and early-warnings compared to conventional reconstruction-based baselines. The latent representations that were learned by the framework have an AUROC value of 0.994, which confirms that it is efficient in encoding discriminative physiological signals that indicate stress. The framework, unlike standard auto-encoder and LSTM reconstruction models, are not as separable, and thus the latent space deviation-based modeling supports the criticalness of explicit statistical distribution modeling, as opposed to just using reconstruction error as a sole anomaly detection method in wearables. The high F1 -score of 0.872 indicates a balanced trade-off between recall and precision. This balance is especially relevant to the medical field where false alarming too often reduces the credibility of the system, and undetected anomalies raise the risk of medical professionals. To balance this, an optimization approach based on threshold helped conclude this balance of the decision boundary depending on validation performance.

NPRC Journal of Multidisciplinary Research
Volume 3, Issue 3, March, 2026 Pages: 1-20
ISSN: 3059-9148 (Online)
DOI: 10.3126/nprcjmr.v3i3.90699

Uncertainty calibration has a powerfully positive impact on detectability: the small value of Expected Calibration Error (ECE) indicates that the confidence levels predicted by the model are quite close to the empirical values of detection. Clinical decision-support systems need both this alignment where false optimism of predictions can be disastrous. In addition, the positive correlation between uncertainty and anomaly deviation is also observed and this also proves the efficacy of uncertainty-aware scoring. The ability to identify early is an important contribution of the work. The framework identifies stress anomalies an average of 822 windows prior to the reported onset, which means that the instances of subtle physiological changes can be identified long before they can be completely realized. Preventative care and real time intervention systems cannot do without this foresight. Basic comparisons support in the originality of the suggested methodology; the poor performance of reconstruction-based auto-encoders or a LSTM-based model demonstrates that they cannot process multimodal wearable data that is both typified by high inter-subject variability. The benefit of using self-supervised methods of representation learning and statistical deviation modeling allows the system to achieve resilience in heterogeneous subjects. With these positive outcomes, there are some limitations. The strategy has only been tested with one wearable dataset, and further research ought to test its applicability to the wide range of different datasets and patient groups. The possibility of deploying wearable hardware in real-time efficiently and energy-sensitive model optimization is a promising direction of research. The discussion indicates that incorporation of self-supervised learning, uncertainty calibration as well as analytical modeled deviation will result in a sound and reliable methodology of detecting early anomaly on wearable devices.

6. Conclusion

This paper has suggested an uncertainty-tuned, self-monitored learning model in the early detection of anomalies in wearable health-monitoring devices. The approach overcomes a variety of issues that are inherent to wearable analytics, such as the paucity of labeled data, inter-subject heterogeneity, noise sensitivity and poor consistency in estimating confidence. The method with uncertainty modeling, self-supervised-representation learning, and statistical deviation-based anomaly scoring is demonstrated to deliver an approach in which the discrimination is almost perfect and the accuracy and recall are balanced. WESAD dataset evaluation reveals a high level of performance, which is indicated by high values of the AUROC and AUPRC scores, the low level of false-alarm, the high level of reliable calibration, and the high level of early-detection. The new framework outperforms the traditional reconstruction-based baselines by a substantial amount, hence justifying its robustness and reasonable usefulness in wearable healthcare. Analytic qualities of the framework improve subject-neutral generalization of subjects and allow early identification of physiological abnormalities. However, extensive amounts of research opportunities exist. The framework should be confirmed at a larger level on other wearable data and a wider range of clinical groups. Privacy preserving distributed training can be facilitated through federated learning integration. The real-time deployment landscape of low-power wearable devices and extending the framework to monitor several

NPRC Journal of Multidisciplinary Research
Volume 3, Issue 3, March, 2026 Pages: 1-20
ISSN: 3059-9148 (Online)
DOI: 10.3126/nprcjmr.v3i3.90699

health conditions, such as cardiac and respiratory anomalies would be promising in the future. Those recommendations will make the suggested framework easier to use in real-life preventive healthcare facilities.

7. Author Contribution

Dipendra Kumar Air contributed to conceptualization, methodology development, model design, implementation, data analysis, visualization, writing original draft preparation, and manuscript review and editing. Karn Dev Bhatt contributed to methodology validation, data analysis support, and manuscript review. Shiv Shankar Pant contributed to supervision, conceptual guidance, and final manuscript review and approval.

8. Conflict of Interest

The authors declares that there is no conflict of interest regarding the publication of this paper.

NPRC Journal of Multidisciplinary Research
Volume 3, Issue 3, March, 2026 Pages: 1-20
ISSN: 3059-9148 (Online)
DOI: 10.3126/nprcjmr.v3i3.90699

References

Albattah, A., & Rassam, M. A. (2022). A correlation-based anomaly detection model for wireless body area networks using convolutional long short-term memory neural network. *Sensors*, *22*(5), 1951. https://doi.org/10.3390/s22051951

Alshareef, M. (2025). *Stress Detection: Leveraging IoMT Data and Machine Learning for Enhanced Well-being* (Doctoral dissertation, Queen Mary University of London). https://qmro.qmul.ac.uk/xmlui/handle/123456789/110191

Benita, D. S., Ebenezer, A. S., Susmitha, L., Subathra, M. S. P., & Priya, S. J. (2024, February). Stress detection using cnn on the wesad dataset. In *2024 International Conference on Emerging Systems and Intelligent Computing (ESIC)* (pp. 308-313). IEEE. https://doi.org/10.1109/ESIC61586.2024.10467978

Del Pup, F., & Atzori, M. (2023). Applications of self-supervised learning to biomedical signals: A survey. *IEEE Access*, *11*, 144180-144203. https://doi.org/10.1109/ACCESS.2023.3343853

Ericsson, L., Gouk, H., Loy, C. C., & Hospedales, T. M. (2022). Self-supervised representation learning: Introduction, advances, and challenges. *IEEE Signal Processing Magazine*, *39*(3), 42-62. https://doi.org/10.1109/MSP.2021.3137030

Gu, X., Deligianni, F., Han, J., Liu, X., Chen, W., Yang, G. Z., & Lo, B. (2023). Beyond supervised learning for pervasive healthcare. *IEEE Reviews in Biomedical Engineering*, *17*, 42-62. https://doi.org/10.1109/RBME.2023.3296938

Mao, P., Li, H., & Yu, Z. (2023). A review of skin-wearable sensors for non-invasive health monitoring applications. *Sensors*, *23*(7), 3673. https://doi.org/10.3390/s23073673

Liu, F., & Panagiotakos, D. (2022). Real-world data: a brief review of the methods, applications, challenges and opportunities. *BMC Medical Research Methodology*, *22*(1), 287. https://doi.org/10.1186/s12874-022-01768-6

Nur, S. (2024). The role of digital health technologies and sensors in revolutionizing Wearable health monitoring systems. *International Journal of Innovative Research in Computer Science and Technology*, *12*(6), 69-80. https://doi.org/10.55524/ijircst.2024.12.6.10

Rodrigo-Guillen, R., Garcia-D'Urso, N., Mora-Mora, H., & Azorin-Lopez, J. (2025). Detecting Abnormal Behavior Events and Gatherings in Public Spaces Using Deep Learning: A Review. *Journal of Sensor and Actuator Networks*, *14*(4), 69. https://doi.org/10.3390/jsan14040069

Sultana, S., Hriday, M. S. H., Haque, R., & Das, M. (2025). Health Monitoring Through Wearables: A Systematic Review of Innovations In Cardiovascular Disease Detection And Prevention. *Strategic Data Management and Innovation*, *2*(01), 96-115. https://doi.org/10.71292/sdmi.v2i01.13

Vallée, A. (2023). Digital twin for healthcare systems. *Frontiers in Digital Health*, *5*, 1253050. https://doi.org/10.3389/fdgth.2023.1253050

Williams, G. J., Al-Baraikan, A., Rademakers, F. E., Ciravegna, F., van de Vosse, F. N., Lawrie, A., ... & Morris, P. D. (2023). Wearable technology and the cardiovascular system:

NPRC Journal of Multidisciplinary Research
Volume 3, Issue 3, March, 2026 Pages: 1-20
ISSN: 3059-9148 (Online)
DOI: 10.3126/nprcjmr.v3i3.90699

the future of patient assessment. *The Lancet Digital Health*, *5*(7), e467-e476.https://doi.org/10.1016/S2589-7500(23)00087-0

Xu, Q., Gu, H., & Ji, S. (2024). Text clustering based on pre-trained models and autoencoders. *Frontiers in Computational Neuroscience*, *17*, 1334436. https://doi.org/10.3389/fncom.2023.1334436

Yahya, M. A., Moya, A. R., & Ventura, S. (2025). Deep learning for multivariate time series anomaly detection: an evaluation of reconstruction-based methods. *Artificial Intelligence Review*, *58*(12), 400. https://doi.org/10.1007/s10462-025-11401-9

Yao, Y., Han, T., Yu, J., & Xie, M. (2024). Uncertainty-aware deep learning for reliable health monitoring in safety-critical energy systems. *Energy*, *291*, 130419. https://doi.org/10.1016/j.energy.2024.130419