v.7. n.5. 2026

SCIENTIFIC JOURNAL - RECIMA21 ISSN 2675-6218

TOWARD ROBUST STRESS DETECTION: A 1D-CNN APPROACH FOR MULTICLASS ECG CLASSIFICATION UNDER LIMITED DATA CONDITIONS

DETECÇÃO ROBUSTA DE ESTRESSE: UMA ABORDAGEM 1D-CNN PARA CLASSIFICAÇÃO MULTICLASSE DE ECG EM CONDIÇÕES DE DADOS LIMITADOS

DETECCIÓN ROBUSTA DEL ESTRÉS: UN ENFOQUE BASADO EN CNN 1D PARA LA CLASIFICACIÓN DE ECG MULTICATEGORÍA EN CONDICIONES DE DATOS LIMITADOS

Clarissa Rodrigues[1], Sandro Rigo[2], Kauã Mark[1]

e757932
https://doi.org/10.47820/recima21.v7i5.7932

PUBLISHED: 05/2026

ABSTRACT
Automated stress detection through physiological signals holds considerable promise for transforming healthcare delivery, reducing associated costs, and enabling timely clinical intervention in stress-related conditions. This paper presents a review of recent developments in stress classification based on physiological data, with emphasis on the most relevant computational methods, prevailing methodological challenges, and emergent research directions in the field. Particular attention is devoted to the constraints imposed by small-scale datasets, the critical role of subject-specific personalized models, and the practical barriers to real-time deployment in naturalistic, uncontrolled settings. Concurrently, we introduce and assess a novel one-dimensional convolutional neural network (CNN) architecture tailored to classify electrocardiogram (ECG) signals across four distinct stress-phase categories. Under data-constrained experimental conditions, the model exhibits robust learning dynamics and adequate generalization, attaining 72.81% accuracy on an independent held-out test set. These outcomes underscore the viability of deep learning for stress classification tasks and highlight the need for future research oriented toward personalized, multimodal, and real-time physiological monitoring frameworks.

KEYWORDS: *Stress. Classification. Physiological signals. Deep learning. Convolutional neural networks. Electrocardiogram (ECG).*

RESUMO
O reconhecimento automatizado do estresse por meio de sinais fisiológicos constitui uma área de crescente relevância clínica e tecnológica, com impacto direto na melhoria dos desfechos em saúde, na redução de custos assistenciais e na viabilização da intervenção precoce em transtornos relacionados ao estresse. O presente trabalho desenvolve uma revisão dos avanços no campo da classificação do estresse baseado em dados fisiológicos, identificando os métodos mais representativos, os principais desafios metodológicos e as tendências que orientam a pesquisa emergente na área. Destaque especial é conferido às limitações impostas por conjuntos de dados de pequena escala, à relevância dos modelos personalizados por sujeito e às dificuldades inerentes à aplicação em tempo real em contextos não controlados. Em paralelo, este trabalho propõe e avalia uma nova arquitetura de rede neural convolucional (CNN) unidimensional, desenvolvida para classificar sinais de eletrocardiograma (ECG) em quatro categorias distintas, correspondentes às fases do estresse. O modelo revelou capacidade de aprendizado robusto e de generalização adequada, mesmo sob condições de escassez de dados, atingindo 72,81% de acurácia em um conjunto de teste independente. Esses resultados evidenciam o potencial do aprendizado profundo para a classificação do estresse e sublinham a necessidade de abordagens futuras que incorporem personalização, processamento em tempo real e fusão multimodal de sinais fisiológicos.

PALAVRAS-CHAVE: Estresse. Classificação. Sinais fisiológicos. Aprendizado profundo. Redes neurais convolucionais. Eletrocardiograma (ECG).

[1] UNISINOS
[2] PhD in Computer Science from the Universidade Federal do Rio Grande do Sul.

v.7. n.5. 2026

SCIENTIFIC JOURNAL - RECIMA21 ISSN 2675-6218

RESUMEN
La detección automatizada del estrés mediante señales fisiológicas representa una línea de investigación de creciente relevancia, con impacto directo en la mejora de resultados clínicos, la reducción de costos en salud y la posibilidad de intervención temprana en trastornos vinculados al estrés. El presente trabajo realiza una revisión exhaustiva de los avances recientes en la clasificación del estrés a partir de datos fisiológicos, destacando los métodos computacionales más relevantes, los desafíos metodológicos predominantes y las tendencias emergentes en el campo. Se hace especial énfasis en las restricciones impuestas por conjuntos de datos de pequeña escala, la importancia de los modelos personalizados por sujeto y las dificultades para la implementación en tiempo real en entornos no controlados. Paralelamente, se propone y evalúa una nueva arquitectura de red neuronal convolucional (CNN) unidimensional diseñada para clasificar señales de electrocardiograma (ECG) en cuatro categorías distintas correspondientes a fases del estrés. Bajo condiciones experimentales con datos limitados, el modelo exhibe un aprendizaje robusto y una generalización adecuada, alcanzando una precisión del 72,81% en un conjunto de prueba independiente. Estos resultados respaldan la viabilidad del aprendizaje profundo para la clasificación del estrés y destacan la necesidad de investigaciones futuras orientadas hacia marcos personalizados, multimodales y de monitoreo en tiempo real.

PALABRAS CLAVE: *Estrés. Clasificación. Señales fisiológicas. Aprendizaje profundo. Redes neuronales convolucionales. Electrocardiograma (ECG).*

1. INTRODUCTION

The automatic analysis and interpretation of physiological signals has consolidated itself as one of the most productive and strategically relevant research areas in contemporary biomedical engineering. The growing societal demand for continuous, non-invasive health monitoring solutions has positioned this field at the convergence of signal processing, machine learning, and clinical science, with tangible applications spanning remote healthcare monitoring, affective computing, and automated stress detection. Among the various physiological modalities accessible for such purposes, the electrocardiogram (ECG) occupies a privileged position by virtue of its established clinical validity, ease of non-invasive acquisition, and demonstrated capacity to reflect the functional dynamics of the autonomic nervous system in response to physical, cognitive, and emotional stressors. The rapid maturation of deep learning methodologies has substantially accelerated progress in this domain, enabling models to autonomously extract hierarchical and discriminative feature representations from minimally preprocessed or raw signal data, consistently surpassing conventional machine learning approaches in time-series classification benchmarks (ANSARI *et al.*, 2023).

Notwithstanding these advances, developing robust, deployable stress classification systems from physiological signals remains a formidable challenge. Data scarcity remains one of the most critical structural limitations: physiological stress datasets collected under standardized experimental protocols are inherently constrained in scale, frequently exhibiting severe class imbalances and limited demographic diversity. These characteristics substantially limit the generalization capacity of trained models, particularly when evaluation conditions differ from those

SCIENTIFIC JOURNAL - RECIMA21 ISSN 2675-6218

during training (FINSETH *et al.*, 2023). Equally significant is the pronounced inter-individual variability in physiological responses to standardized stressors, which results in systematic performance degradation when generalized models are applied across subjects, underscoring the need for architectures that operate reliably under constrained data conditions.

A considerable share of recent research efforts has been directed toward hybrid architectures that integrate convolutional processing with recurrent networks or self-attention mechanisms, aiming to simultaneously capture local morphological patterns and long-range temporal dependencies embedded in physiological signals. While such designs have demonstrated meaningful accuracy gains across benchmark datasets, they demand substantially greater computational resources and larger annotated corpora. These requirements are difficult to satisfy in physiological stress research, where data scarcity is a structural constraint. These constraints critically limit the deployability of such models in resource-restricted environments, including wearable sensor platforms and edge computing systems for real-time stress monitoring (MODI *et al.*, 2025). This mismatch between architectural complexity and operational feasibility provides strong motivation for investigating lightweight, regularized models that prioritize robustness and deployability without sacrificing predictive reliability.

One-dimensional convolutional neural networks (1D-CNNs) have established themselves as particularly effective baseline architectures for ECG and physiological time-series classification tasks. Their inductive bias toward local temporal pattern detection, combined with relatively modest computational requirements, makes them well-suited for extracting morphological and rhythmic features from ECG signals while preserving a favorable balance between representational power and practical deployability (GUHDAR; MOHAMMED; MSTAFA, 2025). Critically, when supported by systematic preprocessing, class-weighted optimization, and structured regularization, 1D-CNNs have demonstrated stable convergence and competitive generalization even under conditions of limited data availability and distributional imbalance — attributes central to the experimental objectives of the present work.

Against this background, the present study examines the behavior of a compact 1D-CNN architecture applied to multiclass temporal signal classification, with a deliberate emphasis on robustness under experimentally realistic constraints. The model is evaluated under conditions of limited sample availability, signal contamination, and class imbalance, with the primary research objective of determining whether a carefully regularized, architecturally concise convolutional network can achieve reliable generalization without resorting to complex hybrid designs or large-scale data augmentation pipelines. The developed experiments used a new Dataset, called WearHealth. The WearHealth dataset was created as part of an interdisciplinary research initiative focused on the assessment of psychophysiological stress through wearable sensing technologies, conducted between 2021 and 2025 in compliance with institutional ethical standards and with informed consent

SCIENTIFIC JOURNAL - RECIMA21 ISSN 2675-6218

obtained from all participants. The dataset includes recordings from 74 individuals drawn from the general population, providing a larger sample than several commonly used public stress datasets. Stress induction was carried out using the Trier Social Stress Test (TSST), a well-established, experimentally validated protocol that reliably elicits physiological and neuroendocrine stress responses in controlled laboratory environments.

Thus, the present study aims to: (i) review the state of the art in automated stress recognition using physiological signals, identifying the most representative methods, datasets, and open challenges; and (ii) propose and evaluate a compact 1D-CNN architecture for multiclass ECG classification across four stress phases, assessing its generalization capacity under limited data, class imbalance, and signal noise conditions.

2. LITERATURE REVIEW

We conducted a non-systematic literature review using a snowball sampling strategy, starting with a curated set of foundational studies and progressively extending the search through their bibliographic networks. The electronic databases consulted comprised ACM Digital Library, IEEE Xplore, Elsevier, Springer, and ScienceDirect. Search terms encompassing "stress," "physiological signals," "classification," and "customization" were employed to retrieve publications indexed from 2021 onwards. An initial corpus of 27 publications was identified, subsequently reduced to five articles following a structured relevance screening procedure.

The review was structured around the following research questions, framed around the central challenge of automated stress recognition through physiological signals:

- RQ1: Which physiological signals — used individually or in combination — are most commonly employed in automated stress detection systems, and what is their clinical and technological relevance?
- RQ2: Which machine learning and deep learning algorithms have been applied to physiological signal-based stress classification, and which demonstrate the most competitive performance in this context?
- RQ3: Which techniques have been proposed to address inter-individual variability and personalization challenges in automated stress recognition, and how effective are they in improving model generalization?

Studies not sufficiently aligned with the defined research questions were excluded, yielding a final set of five articles for in-depth analysis. The analysis of the selected studies considered several key dimensions to ensure a comprehensive evaluation of the literature. First, the main research objectives of each work were examined. Next, the types of physiological signals employed were analyzed, including their individual use and possible combinations. The datasets used in the studies were also evaluated, considering whether they were based on real experimental data or synthetic data,

SCIENTIFIC JOURNAL - RECIMA21 ISSN 2675-6218

and their availability for public access. In addition, the classification methods used in the studies were investigated, focusing on the algorithms employed for data classification. The classes defined in each study were analyzed in relation to the employed signals and classification approaches. Finally, classification accuracy was assessed to evaluate the effectiveness of the proposed methods, particularly across different combinations of physiological signals and algorithms.

Recent advances in stress classification using physiological signals have been significantly driven by the evolution of machine learning and deep learning techniques. Physiological signals provide objective and continuous measurements of human responses, making them more reliable than subjective assessments. Among these signals, electrocardiogram (ECG) has emerged as one of the most relevant modalities due to its ability to capture heart rate variability (HRV), which reflects the activity of the autonomic nervous system under stress conditions. A comprehensive review by Yildirim *et al.* (2023) highlights the growing importance of ECG-based approaches, emphasizing their robustness and strong physiological interpretability.

In addition to ECG, other physiological signals such as electrodermal activity (EDA), photoplethysmography (PPG), respiration, and skin temperature are widely used in stress detection systems. Early surveys, such as the work by Can *et al.* (2019), established the foundation for multimodal stress detection by demonstrating that combining multiple physiological signals improves classification performance. More recent studies reinforce this trend. For instance, Xiang *et al.* (2025) proposed a multimodal framework integrating ECG, EDA, temperature, and accelerometry signals using parallel convolutional neural networks (CNNs), showing that combining time-domain and frequency-domain features leads to improved classification accuracy. Similarly, Yang *et al.* (2025) explored the transformation of physiological signals into two-dimensional representations using Gramian Angular Fields, enabling the application of computer vision-based CNN architectures to stress detection tasks.

From a data perspective, the literature continues to rely heavily on benchmark datasets, with WESAD remaining one of the most widely used resources. Introduced by Schmidt *et al.* (2018), WESAD provides multimodal physiological recordings under controlled stress conditions and has become a standard benchmark for evaluating stress classification models. However, despite its widespread use, limitations related to small sample sizes and limited diversity persist. Gjoreski *et al.* (2017) further demonstrated the challenges of deploying stress detection systems in real-world environments, highlighting the variability introduced by daily-life conditions. Recent work by Zhou *et al.* (2024) reinforces this issue, showing that models trained on specific datasets often fail to generalize across different datasets, suggesting that many approaches capture general arousal rather than stress-specific patterns.

Methodologically, both traditional machine learning and deep learning approaches have been explored. Classical methods such as Support Vector Machines (SVM), Random Forests (RF),

SCIENTIFIC JOURNAL - RECIMA21 ISSN 2675-6218

and k-Nearest Neighbors (k-NN) remain relevant, particularly when applied to handcrafted features derived from ECG signals. However, deep learning approaches have become dominant due to their ability to learn representations directly from raw data. Reiss and Stricker (2019) discuss the advantages of deep learning for sensor-based analysis, emphasizing its ability to model complex patterns in time-series data.

Recent studies have also explored more advanced architectures, particularly transformer-based models. Kyrou *et al.* (2025) conducted a comprehensive survey of deep learning approaches, highlighting the growing adoption of transformers alongside CNN and recurrent neural network (RNN) architectures. These findings suggest a shift toward more expressive models capable of capturing both temporal and cross-modal dependencies.

Despite these advances, several challenges remain. Inter-individual variability remains one of the most critical limitations, as physiological responses to stress differ significantly across subjects. Liu *et al*. (2025) emphasize that current models often struggle to generalize due to limited and homogeneous datasets. Additionally, noise, motion artifacts, and missing data in real-world environments further degrade model performance. Another important limitation concerns the labeling of stress, which often relies on experimental protocols or self-reported measures, potentially introducing bias into the training process.

Yildirim *et al*. (2023) present a comprehensive review of deep learning techniques applied to ECG-based stress detection. The study synthesizes recent developments in CNN, RNN, and hybrid architectures, emphasizing the relevance of HRV-derived features and raw signal processing. The authors highlight that ECG signals, due to their strong physiological grounding, enable robust stress detection even in single-modality settings. The review does not rely on a single dataset but rather aggregates findings from multiple benchmarks, including WESAD and similar repositories. A key contribution of this work is the identification of trends toward end-to-end deep learning models that reduce dependence on handcrafted features.

Table 1. Selected Studies

Study	Signals	Dataset	Method	Contribution	Accuracy
Yildirim *et al.* (2023)	ECG	Review	DL survey	ECG-based DL overview	—
Xiang *et al.* (2025)	ECG + multimodal	Experimental	CNN	Time + frequency fusion	>90%
Yang *et al.* (2025)	Multimodal	WESAD	CNN (2D transform)	Signal-to-image encoding	High
Schmidt *et al.* (2018)	Multimodal	WESAD	Dataset	Benchmark dataset	—
Gjoreski *et al.* (2017)	ECG + EDA	Real-world	ML	Continuous stress detection	—

SCIENTIFIC JOURNAL - RECIMA21 ISSN 2675-6218

Xiang *et al.* (2025) propose a multimodal stress detection framework that integrates ECG, EDA, temperature, and accelerometer data collected in experimental conditions. The main contribution lies in the use of parallel CNN architectures to process both time-domain and frequency-domain representations of physiological signals. By applying transformations such as Fast Fourier Transform (FFT), the model captures complementary information from the signals, leading to improved classification performance. The study demonstrates that multimodal fusion significantly enhances robustness compared to single-signal approaches.

Yang *et al.* (2025) introduce an innovative approach that transforms physiological time-series signals into two-dimensional representations using Gramian Angular Fields (GAF). These representations are then processed using CNN architectures traditionally applied in computer vision. Using datasets such as WESAD, the study shows that encoding temporal signals as images enables more effective feature extraction, particularly for capturing complex temporal dependencies. The main contribution is the bridge between time-series analysis and image-based deep learning techniques.

Schmidt *et al.* (2018) introduce the WESAD dataset, one of the most widely used benchmarks in stress detection research. The dataset includes multimodal physiological signals such as ECG, EDA, respiration, and body temperature, collected under controlled experimental conditions. Its main contribution is providing a standardized benchmark that enables reproducibility and comparison across studies. Despite its importance, the dataset is limited in scale, which impacts model generalization. Gjoreski *et al.* (2017) investigate continuous stress detection in real-world environments using wearable devices. The study utilizes ECG and EDA signals collected in both laboratory and real-life conditions, applying traditional machine learning techniques for classification. The main contribution is the demonstration of the challenges associated with real-world deployment, including noise, missing data, and variability in user behavior.

Although these benchmark repositories have made physiological signal data accessible to the broader research community, they are constrained by small participant cohorts and prioritize methodological comparisons over systematic data collection. Recognizing meaningful physiological changes attributable to stress typically demands large-scale data, and while windowing techniques help manage data volume, the challenge of preserving sufficient signal content and temporal context remains an open methodological problem.

The reviewed literature demonstrates a clear evolution toward deep learning-based approaches, particularly hybrid and transformer-based architectures. While multimodal approaches tend to achieve higher accuracy, ECG remains a central signal due to its physiological relevance and compatibility with wearable devices. However, persistent challenges such as dataset limitations, lack of generalization, and inter-subject variability continue to hinder the development of robust stress detection systems.

Moreover, the literature indicates that current models may capture general physiological

SCIENTIFIC JOURNAL - RECIMA21 ISSN 2675-6218

arousal rather than stress-specific responses, particularly when evaluated across datasets. This limitation highlights the need for improved validation protocols and more diverse datasets. Additionally, personalization strategies and adaptive models remain underexplored, representing a promising direction for future research. This review identifies a persistent gap in the systematic exploration of diverse physiological signal combinations and in innovation in preprocessing methodologies, underscoring an underexplored dimension with meaningful potential for advancing stress pattern identification. Notably absent from the reviewed literature is the integration of hormonal biomarkers — particularly salivary cortisol, a well-established objective marker of HPA axis activation and physiological stress — into classification and annotation pipelines. Incorporating such endocrine indicators is recommended to strengthen the validity of stress labels and enhance the accuracy of experimental datasets.

A cross-study comparison reveals a consistent tension between model complexity and practical deployability. While transformer-based and hybrid CNN-RNN architectures consistently report higher accuracy on benchmark datasets, they depend on large annotated corpora and substantial computational resources — conditions rarely available in applied physiological monitoring contexts. In contrast, compact convolutional models, when supported by principled regularization and preprocessing, demonstrate competitive generalization under constrained conditions. This trade-off has been insufficiently addressed in the reviewed literature, where architectural complexity is frequently pursued at the expense of operational feasibility.

Furthermore, the absence of hormonal biomarkers such as salivary cortisol in annotation pipelines represents a critical gap: without biochemical validation, stress labels derived solely from experimental protocols or self-report measures remain susceptible to systematic bias. The present work directly addresses these two gaps — limited-data robustness and biomarker-informed dataset design — through the WearHealth dataset and the proposed 1D-CNN framework.

In summary, although significant progress has been made, stress classification using physiological signals — especially ECG — still faces important challenges. Addressing these limitations requires developing models that generalize across individuals, handle noisy real-world data, and incorporate more reliable labeling strategies. These gaps motivate the methodological contributions proposed in this work. Given the confluence of challenges identified above — notably limited data availability, multiclass distributional imbalance, and the persistent difficulty of generalizing across subjects — the following section presents the experimental component of this work: a CNN-based classification model specifically designed to operate under these conditions.

3. EXPERIMENT: CNN MODEL USING ECG

This section details the experimental design, implementation, and evaluation of a CNN model for multiclass classification of one-dimensional temporal physiological signals. The experiment is

SCIENTIFIC JOURNAL - RECIMA21 ISSN 2675-6218

framed around three interrelated goals: assessing architectural robustness under data constraints, evaluating generalization on an independent held-out test set, and examining the suitability of the proposed model for practical deployment. The following subsections describe the model architecture, the dataset and its preprocessing pipeline, the training configuration, and the resulting performance metrics.

3.1. The dataset

The WearHealth dataset was developed within an interdisciplinary research program dedicated to psychophysiological stress measurement using wearable sensor technologies, conducted between 2021 and 2025 under institutional ethical review and with informed consent obtained from all enrolled participants. The dataset comprises recordings from 74 individuals recruited from the general population, a sample size exceeding that of several widely referenced public stress datasets. Stress induction was performed using the Trier Social Stress Test (TSST), a standardized and empirically validated protocol that reliably elicits both physiological and neuroendocrine stress responses under controlled laboratory conditions.

The multimodal physiological recording infrastructure integrated continuous ECG, electrodermal activity (EDA), and electromyography (EMG) streams, supplemented by heart rate and heart rate variability (HRV) measurements from a validated reference device, all recorded across sessions lasting approximately 60 to 70 minutes per participant. Complementing these autonomic signals, four serial salivary cortisol samples were collected from each participant to characterize the delayed hormonal activation of the hypothalamic-pituitary-adrenal (HPA) axis, providing an objective biochemical biomarker that enables independent validation of stress labeling. The joint capture of fast autonomic responses and slow endocrine dynamics represents a key distinguishing feature of this dataset relative to existing public repositories.

In addition to the core physiological streams, the WearHealth dataset includes standardized psychological questionnaires, anthropometric measurements, and detailed session-level annotations. Signal data is labeled across four sequential experimental phases — baseline, anticipatory stress, active stress, and post-stress recovery — enabling fine-grained temporal analysis of stress dynamics beyond conventional binary classification frameworks. This structured, multimodal design is intended to support diverse research applications spanning stress detection, analysis of individual physiological variability, and model generalization studies, while also providing the methodological groundwork for future label refinement informed by biochemical validation.

3.2. Proposed architecture

The proposed architecture is a one-dimensional convolutional neural network (1D-CNN) tailored for multiclass classification of temporal physiological signals. The architectural philosophy prioritizes a balance between sufficient representational capacity for hierarchical feature extraction and computational parsimony, ensuring stable and consistent learning behavior under the data-constrained and class-imbalanced conditions characteristic of the experimental setting.

The feature extraction backbone comprises three successive convolutional blocks with progressively increasing filter depths of 32, 64, and 128. Each convolutional operation applies a kernel of size 3 with symmetric padding to preserve the temporal dimensionality of the input signal. Batch normalization and ReLU activation are applied after each convolution to stabilize gradient flow, accelerate convergence, and mitigate the adverse effects of internal covariate shift during training. Max-pooling operations are interposed after the first and second convolutional blocks to reduce temporal resolution, compress the feature representation, and introduce a degree of translational invariance that contributes to overfitting control.

Following the final convolutional block, dropout regularization at a rate of 0.3 is applied to the feature maps to stochastically deactivate activation pathways and improve generalization. The flattening step is replaced by an adaptive average pooling layer, which compresses the temporal feature maps into a fixed-length global descriptor regardless of input length, enabling graceful handling of variable-duration signal windows while retaining the most salient temporal patterns.

The classification head consists of two fully connected layers: the first projects the pooled feature vector into a 128-dimensional latent representation space using ReLU activation, while the second maps this representation to the four output neurons corresponding to the target stress phase classes. This hierarchical architecture enables effective multi-level feature abstraction while maintaining structural simplicity, signal noise robustness, and resilience to class imbalance.

3.3. Data preprocessing

Input data comprised fixed-length one-dimensional temporal signal windows stored in comma-separated value (CSV) format, with each window assigned to one of four target class labels. Feature matrices and label vectors were loaded from separate structured files to ensure modularity and reproducibility of the data ingestion pipeline.

Raw input values were cast to numeric data types, and missing entries were resolved through column-wise mean imputation to ensure numerical completeness while avoiding sample loss. Class labels were integer-encoded to support supervised multiclass training.

To reflect realistic limitations in data availability, only 75% of the complete dataset was included in the experimental evaluation. Within this subset, stratified sampling was employed throughout all data partitioning steps to maintain proportional class representation across training, validation, and test splits. The overall dataset was divided such that 80% of the retained samples were allocated to the combined training and validation partition, with the remaining 20% reserved as a fully independent test set.

Input feature normalization was implemented via z-score standardization, with the normalization parameters (mean and standard deviation) estimated exclusively from the training partition and subsequently applied to both the validation and test sets, thereby preventing information leakage across splits. A channel dimension was appended to each input tensor to conform to the input format expected by 1D convolutional layers.

This preprocessing pipeline enforces numerical stability and temporal structural integrity, while providing a principled mechanism for mitigating class imbalance effects across splits. The deliberate use of a constrained data subset aligns the experimental conditions with the realistic data limitations prevalent in clinical and applied physiological monitoring contexts, and explicitly contrasts with the full-scale, class-balanced datasets commonly employed in the benchmarked literature, including ASCERTAIN, DEAP, and DREAMER.

3.4. Training configuration

Model training was conducted within a supervised learning framework explicitly configured to address the multiclass imbalance characteristic of the experimental dataset. The primary optimization objective was categorical cross-entropy loss with class weights inversely proportional to class frequency in the training partition, ensuring that minority classes contribute proportionally to the aggregate gradient signal.

Parameter optimization employed the Adam algorithm with a base learning rate of 1×10^{-3} and an L2 regularization weight decay of 1×10^{-5}, providing an additional implicit regularization mechanism that penalizes large parameter magnitudes. Mini-batch gradient descent was performed with a batch size of 64, balancing computational throughput with gradient estimation stability.

A ReduceLROnPlateau learning rate scheduling policy monitored the validation loss trajectory and dynamically reduced the learning rate upon detection of performance plateaus, allowing finer parameter refinement as the model approached local convergence. Early stopping with a patience of 7 epochs was additionally applied, halting training when no improvement in validation loss was observed across consecutive epochs and restoring the best-validated model checkpoint.

Training was bounded at a maximum of 50 epochs, though early stopping typically intervened before this limit was reached. The combination of adaptive learning rate reduction, early stopping, and weight decay collectively produced a stable and reproducible training regime that effectively counteracted overfitting tendencies under constrained data conditions.

3.5. Obtained results

The model was assessed on the independent test partition using accuracy, precision, recall, and F1-score as the primary performance indicators. The proposed 1D-CNN architecture achieved an overall test accuracy of 72.81%, representing competitive classification performance for a four-class temporal signal problem under conditions of data scarcity, imbalance, and noise.

Analysis of per-class performance revealed that Classes 0 and 1 were classified most reliably, with F1-scores exceeding 0.79, reflecting the model's capacity to capture discriminative temporal features for these stress phases. Classes 2 and 3 exhibited reduced recall values, consistent with the greater degree of temporal pattern overlap between adjacent stress phases — a well-documented source of confusion in multiclass physiological signal classification.

The macro-averaged F1-score of 0.7265 and the weighted F1-score of 0.7263 jointly indicate balanced classification performance across the class distribution, with the close correspondence between macro and weighted metrics suggesting robustness to the dataset's imbalance

SCIENTIFIC JOURNAL - RECIMA21 ISSN 2675-6218

characteristics. The training and validation loss trajectories exhibited smooth and monotonic convergence behavior throughout training, with no evidence of significant overfitting, validating the effectiveness of the adopted regularization, scheduling, and early stopping configuration.

Table 2 presents a comparative overview of the proposed model's performance against representative methods from the reviewed literature, contextualized by dataset, signal modality, and classification approach. It is important to note that the comparison presented in Table 2 is intended as an illustrative context for the proposed approach, since the evaluated studies employ different datasets, signal acquisition protocols, modalities, preprocessing strategies, and classification settings, which limit the possibility of a direct performance comparison.

Table 2. Comparative Table

Reference	Dataset	Signal	Model	Classes	Accuracy
Xiang *et al.* (2025)	Experimental	ECG + multimodal	CNN (parallel)	Binary	>90%
Yang *et al.* (2025)	WESAD	Multimodal	CNN (2D transform)	CNN (2D-GAF)	High
Present study	**WearHealth**	**ECG**	**1D-CNN**	**4 classes**	**72.81%**

Collectively, the experimental results demonstrate that the proposed 1D-CNN architecture achieves robust and reproducible generalization in a data-limited, noisy, multiclass scenario, supporting its candidacy as a practical and scalable baseline for physiological temporal signal classification. It is important to note that direct numerical comparisons should be interpreted with caution, as the studies differ in class granularity, dataset characteristics, and evaluation protocols. Notably, the proposed model operates under a four-class framework — a substantially harder task than binary classification — under limited data conditions, which contextualizes its 72.81% accuracy as a competitive result.

3.6. Ethical aspects

As this study constitutes a systematic literature review and computational experiment conducted exclusively on a previously collected dataset obtained under ethical oversight, no additional ethics committee review was required for the present work.

4. CONCLUSION

This study investigated the classification performance of a one-dimensional convolutional neural network (CNN) architecture applied to multiclass temporal physiological signal analysis under conditions reflecting real-world operational constraints, including restricted dataset size, class distribution imbalance, and ambient signal noise. By emphasizing architectural parsimony, principled regularization, and systematic preprocessing, the proposed model achieved stable convergence

SCIENTIFIC JOURNAL - RECIMA21 ISSN 2675-6218

dynamics and consistent generalization performance, reinforcing the continued relevance of compact CNN architectures as practical and reproducible baselines for ECG-based physiological classification.

The experimental evidence indicates that the proposed model successfully learns discriminative temporal representations despite the imposed data limitations, achieving a test accuracy of 72.81% and balanced macro-F1 performance. This behavior aligns with converging evidence in the recent literature demonstrating that appropriately regularized CNN architectures maintain competitive performance on ECG and physiological time-series classification tasks without necessitating hybrid architectural complexity (ANSARI *et al.*, 2023). Rather than maximizing performance through increasingly intricate model designs, this work foregrounds robustness, computational reproducibility, and operational feasibility as primary design criteria.

The results simultaneously expose persistent structural limitations inherent to multiclass stress classification from physiological signals. Temporal pattern overlap between adjacent stress phases and interindividual physiological variability remain the primary sources of classification ambiguity, restricting performance gains, particularly in generalized evaluation settings. These limitations are broadly acknowledged across the recent literature and reflect fundamental properties of physiological signal distributions in ecologically valid measurement contexts (XIANG *et al.*, 2025).

Prospective research directions include integrating architectures with enhanced long-range temporal modeling capabilities, such as self-attention modules and hybrid CNN-Transformer designs, which have shown meaningful progress in capturing broader contextual dynamics from physiological time series (MODI *et al.*, 2025). Expanding the training corpus, refining annotation procedures through biochemical biomarker validation — particularly salivary cortisol measurement — and exploring transfer learning from large publicly available ECG repositories represent additional pathways for improving model robustness and generalization potential.

This work contributes a transparent, reproducible, and computationally efficient CNN-based classification framework that addresses the practical data constraints that characterize physiological stress monitoring research. By deliberately targeting the low-data, noisy, and imbalanced ECG classification scenario — a condition that is underrepresented in the existing stress detection literature — this study fills a meaningful methodological gap. Future enhancements will pursue the integration of recurrent modules (LSTM, BiLSTM) for long-term temporal dependency modeling, attention-based feature weighting, multimodal fusion incorporating EDA and EEG signals, and the application of transfer learning from large-scale ECG repositories. These combined refinements hold significant potential for narrowing the performance gap between generalized and personalized stress classification systems, advancing toward accurate, adaptable, and deployable stress monitoring solutions for real-world healthcare applications.

SCIENTIFIC JOURNAL - RECIMA21 ISSN 2675-6218

REFERENCES

ANSARI, Y.; MOURAD, O.; QARAQE, K.; SERPEDIN, E. Deep learning for ECG arrhythmia detection and classification: an overview of progress for 2017–2023. Frontiers in Physiology, Lausanne, v. 14, 2023.

CAN, Y. S.; ARNRICH, B.; ERSOY, C. Stress detection from physiological signals: a survey. IEEE Access, New York, v. 7, p. 161798–161814, 2019.

FINSETH, T.; KJELDGAARD, M.; SKOV, M. B.; HANSEN, L. K. Personalized versus generalized models for physiological stress detection: challenges and opportunities. Sensors, Basel, v. 23, n. 3, 2023.

GJORESKI, M.; GJORESKI, H.; LUTREK, M.; GAMS, M. Continuous stress detection using a wrist device: in laboratory and real life. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, New York, v. 1, n. 2, p. 1–20, 2017.

GUHDAR, M.; MOHAMMED, A. O.; MSTAFA, R. J. Advanced deep learning framework for ECG arrhythmia classification using 1D-CNN with attention mechanism. Knowledge-Based Systems, Amsterdam, v. 295, 2025.

HASANPOOR REISS, A.; STRICKER, D. Deep learning for sensor-based activity recognition: a survey. Pattern Recognition Letters, Amsterdam, v. 119, p. 3–11, 2019.

KYROU, I. et al. Deep learning for physiological signal-based stress detection: a survey. IEEE Transactions on Affective Computing, New York, v. 16, n. 2, 2025.

LIU, Q.; ZHANG, X.; HUANG, M. A comprehensive review of machine learning and deep learning for stress detection using physiological signals. Applied Sciences, Basel, v. 15, n. 21, p. 11777, 2025.

MODI, N.; KUMAR, Y.; MEHTA, K.; CHAPLOT, N. Physiological signal-based mental stress detection using hybrid deep learning models. Discover Artificial Intelligence, Cham, v. 5, 2025.

SCHMIDT, P.; REISS, A.; DUERICHEN, R.; MARBERGER, C.; VAN LAERHOVEN, K. Introducing WESAD, a multimodal dataset for wearable stress and affect detection. In: ACM INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION, 20., 2018. Proceedings [...]. New York: ACM, 2018. p. 400–408.

XIANG, J.-Z.; WANG, Q.-Y.; FANG, Z.-B.; ESQUIVEL, J. A.; SU, Z.-X. A multimodal deep learning approach for stress detection using physiological signals. Frontiers in Physiology, Lausanne, v. 16, 2025.

SCIENTIFIC JOURNAL - RECIMA21 ISSN 2675-6218

YANG, H.; CHEN, L.; ZHAO, Y. Stress detection using two-dimensional transformation of physiological signals and deep learning. Scientific Reports, London, v. 15, n. 1, p. 1228, 2025.

YILDIRIM, O. et al. Deep learning for ECG-based stress detection: a review. Biomedical Signal Processing and Control, Amsterdam, v. 80, p. 104319, 2023.