

Golden Jackal Optimization-Driven Progressive Wasserstein Generative Adversarial Network for Wearable Cortisol-Based Stress Severity Classification

Jeya Daisy I¹, Dinesh Kumar V²

^{1,2}*Department of Electronics and Instrumentation, Kumaraguru college of Technology, Coimbatore, India*

Abstract—Chronic dysregulation of the hypothalamic-pituitary-adrenal axis is increasingly recognized as a driver of cognitive decline, cardiovascular disease and reduced quality of life, yet clinically meaningful, continuous, non-invasive tracking of stress remains an open problem. Sweat cortisol offers a biochemically direct proxy for circulating cortisol, but wearable sweat-sensing platforms still struggle with noisy, low-resolution readings, discontinuous flow and site-dependent variability, and the software pipelines built on top of them commonly rely on shallow classifiers whose hyper-parameters are chosen manually. This paper proposes SLM-CS-PWGAN, a stress-level monitoring pipeline that couples multi-site cortisol sensing with a Progressive Wasserstein Generative Adversarial Network (PWGAN) classifier whose hyper-parameters are tuned automatically by the Golden Jackal Optimization Algorithm (GJOA). Raw multi-site cortisol streams collected from eight participants are first denoised with an Unscented Trainable Kalman Filter, after which time-domain statistical descriptors are extracted through a Dual-Tree Complex Wavelet Transform. The resulting feature vectors are classified into Normal, Acute and Chronic stress categories by the PWGAN, whose generator and discriminator are grown progressively across resolutions to stabilize adversarial training, while GJOA searches the learning-rate, batch-size and architectural-depth space that conventional PWGAN training leaves to manual tuning. The paper details the complete architecture, the mathematical formulation of every stage, the evaluation protocol against three literature baselines, and the metrics on which the framework is to be benchmarked. The design is intended to reduce manual tuning effort and improve the stability of adversarial training relative to fixed-hyper-parameter PWGAN and the classical machine-learning baselines reviewed in this work.

Index Terms—cortisol sensor, wearable stress monitoring, Golden Jackal Optimization, Progressive

Wasserstein GAN, Kalman filtering, dual-tree complex wavelet transforms, hyper-parameter optimization

I. INTRODUCTION

The physiological stress response is, in modest and time-limited doses, an adaptive mechanism that helps the body respond to challenge. When the response is triggered too often, too strongly, or fails to switch off, the resulting allostatic load is associated with measurable harm to the brain and cardiovascular system, and it has been linked to the progression of neurodegenerative disease and cancer [1], [2]. Because stress is fundamentally an internal, subjective state, tracking it from outside the body is inherently difficult, and clinical instruments such as the Trier Social Stress Test or self-report questionnaires do not scale to continuous, everyday monitoring [3].

Consumer wearables already estimate proxies such as heart-rate variability, electrodermal activity and skin temperature, but these signals are downstream consequences of the autonomic nervous system rather than direct measurements of the hormonal stress cascade, so their correspondence with true internal stress state is loose [4]. A more direct route is to measure the stress hormone itself. Circulating cortisol rises during a stress episode, and this rise is mirrored in the cortisol secreted by eccrine sweat glands, which makes sweat an attractive, continuously replenished biofluid for wearable biosensing [5], [6]. Electrochemical cortisol sensors have consequently received sustained research attention, yet the field remains comparatively young, and several hardware-level obstacles continue to limit the temporal fidelity of the resulting measurements [7].

Most reported wearable cortisol platforms place electrodes directly against the skin [8]. This design is mechanically simple, but because sweat is sensed rather than collected and isolated, evaporation, contamination, bubble formation and the mixing of fresh and stale sweat all corrupt the signal, and the device loses the ability to resolve how cortisol concentration changes over short time-scales [9], [10]. Skin topography, local temperature and pH further confound the reading, and without an active mechanism to clear the sensing volume, analytes diffuse and blur the measurement [11]. Microfluidic sweat-collection layers built from silicone elastomers address some of these issues, but their hydrophilic surface treatment is often temporary, several candidate biomarkers adsorb non-specifically onto the channel walls, and the multistep lithography used to pattern them is slow and expensive [12], [13]. Dual-layer designs that combine an inner polyurethane layer with an outer silicone layer improve wettability, but at the cost of additional material layers, processing steps and design complexity [14]. To circumvent these constraints, our broader project uses a disposable, laser-cut, skin-interfaced microfluidic device (MicroSweat) whose channel and valve geometry was optimised with resistive-network modelling to sustain bubble-free, unidirectional sweat flow [15]; the present paper focuses on the downstream computational pipeline that turns the multi-site cortisol readings this device provides into a stress-severity decision.

Even with a well-behaved sensor, the signal-processing and classification layer built on top of it has its own weaknesses. Several published stress-detection pipelines pair physiological wearable signals with a cortisol-based ground truth and report a consistent trade-off: models that are fast to run tend to be inaccurate, and models that are accurate tend to be slow or brittle across subjects [16]–[21]. Part of the difficulty is that these pipelines mostly use fixed-architecture or fixed-hyper-parameter classifiers, so the model that performs best on one cohort rarely transfers cleanly to another without re-tuning by hand. Generative adversarial approaches, and in particular Wasserstein GANs trained progressively across resolutions, have shown in adjacent signal domains that they can learn a more discriminative feature-to-label mapping than a purely feed-forward classifier while remaining more stable to train than a standard

GAN [25]. Their main practical drawback is that they introduce a larger hyper-parameter space learning rate, batch sizes, growth schedule, network depth that is conventionally tuned manually, which is exactly the kind of search problem that nature-inspired metaheuristics are suited to. The Golden Jackal Optimization Algorithm, which models the cooperative pursuit behaviour of male and female jackal pairs, has been shown to balance exploration and exploitation efficiently on multimodal engineering problems [26], which motivates its use here to automate PWGAN tuning rather than leaving it to trial and error.

This paper brings these two threads together. Section II reviews prior wearable stress-classification and cortisol-sensing work in more detail. Section III presents the proposed SLM-CS-PWGAN pipeline: multi-site real-time acquisition, Kalman-filter denoising, dual-tree wavelet feature extraction, and GJOA-tuned progressive Wasserstein GAN classification. Section IV describes the experimental protocol, baselines and evaluation metrics against which the framework is to be benchmarked. Section V discusses the novelty of the design relative to the literature, and Section VI concludes the paper and outlines future work.

A. Contributions

- A stress-level monitoring framework, SLM-CS-PWGAN, that couples real-time multi-site cortisol sensing with a metaheuristically-tuned generative classifier, is proposed for three-class stress severity grading (Normal, Acute, Chronic).
- Cortisol time-series from eight participants, sensed simultaneously at the armpit, lower back, upper back and hand, are denoised using an Unscented Trainable Kalman Filter (UTKF) that combines the sigma-point unscented transform with a trainable process-noise estimator.
- Time-domain descriptors (mean, median, standard deviation, skewness, kurtosis and coefficient of variation) are extracted from the denoised signal through a Dual-Tree Complex Wavelet Transform (DT-CWT), which gives near shift-invariant sub-bands compared with the conventional discrete wavelet transform.
- A Progressive Wasserstein GAN (PWGAN) is adopted as the classifier, growing its generator and

discriminator across resolutions to improve training stability relative to a fixed-depth adversarial network.

- The Golden Jackal Optimization Algorithm (GJOA) is used to search the PWGAN hyper-parameter space (learning rate, batch size, generator depth) automatically, replacing manual tuning.
- The complete pipeline is positioned against three literature baselines SLM-CS-LSTM [16], SLM-CS-RF [17] and SLM-CS-LOOCV [18] and an evaluation protocol covering accuracy, precision, sensitivity, specificity, F1-score, error rate, ROC-AUC and computational time is defined for the benchmarking study.

II. RELATED WORK

Related work is organised into two strands: sensing-hardware studies that address how cortisol is physically collected and measured from sweat, and machine-learning studies that address how a stress/no-stress or stress-severity decision is derived once physiological or biochemical signals are available.

A. Sweat Cortisol Sensing Hardware

Karuppaiah et al. surveyed electrochemical cortisol sensor designs and characterisation protocols, and concluded that fabrication repeatability and long-term stability of the sensing membrane remain the dominant open problems in the field [12]. Pali et al. proposed a wearable aptamer-based cortisol watch (CATCH) intended to act as a biomimetic anxiety alarm, illustrating the trend toward affinity-based rather than purely electrochemical sensing [13]. Pérez and Orozco reviewed wearable electrochemical biosensors for analytes, such as hormones and proteins, that partition unevenly between blood and sweat, highlighting that partition-coefficient variability is a confound that any downstream classifier must implicitly absorb [6]. Zea et al. reviewed the broader landscape of cortisol-sensing electrochemistry and observed that most reported devices still fall short of the sensitivity and selectivity required for unsupervised, at-home deployment [15]. Mohammadi et al. reviewed aptamer-based cortisol monitoring platforms and reported similar concerns about long-term biofouling of the sensing surface [5]. Collectively, this body of work motivates the multi-site, filtered acquisition strategy adopted in Section III, since no single sensing

site or raw signal is assumed to be clean enough to classify directly.

B. Machine-Learning-Based Stress Classification

Nath and Thapliyal, together with Caban-Holt, examined electrodermal activity (EDA) and blood-volume-pulse (BVP) signals from 19 older adults during the Trier Social Stress Test, using salivary cortisol as the ground-truth label. Thirty-nine statistical descriptors of the EDA and BVP peaks were reduced by supervised feature selection before four classical classifiers were compared; the reported pipeline was computationally light but comparatively inaccurate [16]. In a follow-up study, Nath and Thapliyal extended the same cortisol-referenced protocol to a wrist-worn platform combining EDA, BVP, inter-beat interval and skin-temperature sensors across a year-long, 40-participant data-collection effort; classification precision improved, but F1-score remained comparatively low, indicating a class-imbalance-sensitive model [17]. Iqbal et al. carried out a sensitivity analysis of heart rate, respiratory rate, skin conductance, RR interval, heart-rate variability and electromyography on the public WESAD dataset, and found respiratory rate and heart rate to be the two most discriminative indicators between stressed and unstressed states, although the reported specificity of the resulting models was modest [18].

Kim et al. built a thermal-comfort prediction model that additionally incorporated blood glucose and salivary cortisol alongside conventional bio-signals collected from 30 participants across cold, neutral and warm conditions, comparing distributed random forest, gradient boosting and neural-network learners; sensitivity improved at the cost of a substantially longer training time [19]. Aristizabal et al. evaluated the feasibility of a deep-learning pipeline on a small dataset of EDA, skin-temperature and heart-rate measurements combined with self-reported anxiety and stress in a semi-controlled laboratory setting, and reported a favourable error rate together with a still-limited sensitivity, which the authors attributed to the small sample size [20]. Clark et al. addressed a related but distinct problem predicting a driver's stress state up to one minute in advance from galvanic skin response, respiration and ECG signals collected from the hand, foot and chest using a Random Forest classifier validated with a leave-one-subject-out

protocol, obtaining high precision but a comparatively higher error rate than the in-lab studies above [21]. Across [16]– [21], a recurring pattern emerges: pipelines that rely on cortisol only as a ground-truth label rather than as an input signal, and that use conventional supervised classifiers with hand-set hyper-parameters, trade accuracy against computation time rather than improving both simultaneously. This observation is the direct motivation for the problem statement in Section III-B.

C. Generative and Metaheuristic Modelling of Physiological Signals

A separate strand of work, outside the stress-monitoring literature specifically, has explored generative adversarial modelling as a way of learning richer feature representations from noisy, low-sample-size biomedical signals than a discriminative classifier can learn on its own. Wang and Hu showed that a progressively-grown Wasserstein GAN can denoise low-dose CT scans more effectively than fixed-depth convolutional networks, because growing the network resolution stage-by-stage avoids the training instability that a full-depth Wasserstein critic exhibits when trained from scratch on high-resolution inputs [25]. The same progressive-growth idea transfers naturally to the present setting, where the input is a low-dimensional statistical feature vector rather than an image, but the underlying instability problem a critic that dominates the generator early in training is the same.

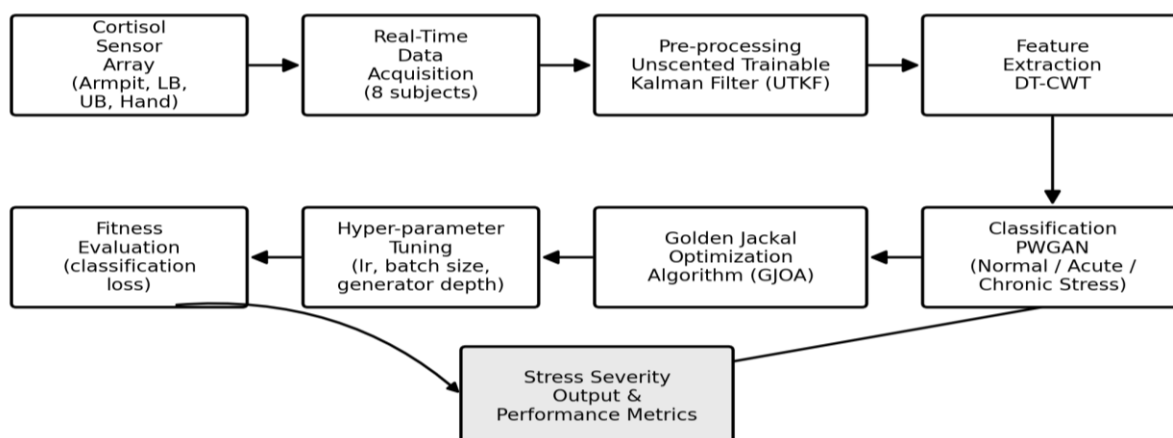
A parallel strand addresses the hyper-parameter sensitivity of adversarial and deep networks through nature-inspired metaheuristics rather than grid or random search. Chopra and Ansari introduced the

Golden Jackal Optimization Algorithm as a population-based optimiser for constrained engineering design problems, showing that its paired exploration-exploitation mechanism converges faster than particle-swarm and grey-wolf optimisers on a standard multimodal benchmark suite [26]. Yu et al. separately showed that a Kalman filter whose noise covariances are themselves learned rather than fixed the Unscented Trainable Kalman Filter adopted in Section III-D outperforms a fixed-covariance unscented Kalman filter when the underlying process is only partially observed, which is the case for a cortisol signal sampled at a handful of body sites [23]. None of these three techniques (progressive Wasserstein GANs, Golden Jackal optimisation, trainable Kalman filtering) has previously been combined into a single cortisol-based stress-classification pipeline, which is the gap this paper addresses.

III. PROPOSED METHODOLOGY

Stress leaves a biochemical trace in sweat cortisol before it is reliably visible in autonomic proxies such as heart rate or skin conductance, but that trace is small, site-dependent and easily buried in sensor noise. The framework proposed in this paper, SLM-CS-PWGAN, is designed around that observation: rather than trying to classify stress from a single noisy channel, it fuses multi-site cortisol readings, filters them explicitly, extracts descriptors that are robust to small time shifts, and hands the result to a classifier whose own hyper-parameters are optimised automatically instead of hand-tuned. Fig. 1 shows the resulting end-to-end architecture.

Fig. 1. System architecture of the proposed SLM-CS-PWGAN framework



A. Problem Statement

Long-term, unmanaged stress accelerates physical and cognitive decline, and this is particularly consequential for older adults, for whom the ability to self-monitor and self-regulate stress is closely tied to independence and perceived well-being. The pipelines reviewed in Section II [16]–[21] each address part of this need, but none combines direct cortisol sensing with an automatically-tuned generative classifier: existing systems either accept lower accuracy for the sake of speed, or accept longer computation time for the sake of accuracy. Both trade-offs are limiting for a device intended to run continuously on the body.

B. Motivation

The trade-off identified above is not fundamental it is largely an artefact of hand-tuned, fixed-hyper-parameter models applied to a noisy input signal. If the input signal is cleaned and represented well, and if the classifier's own configuration is searched automatically rather than fixed a priori, both accuracy and computation time can, in principle, improve together. This is the working hypothesis behind coupling a progressively-grown Wasserstein GAN with the Golden Jackal Optimization Algorithm.

C. Data Acquisition

Cortisol concentration is sensed in real time from eight human participants using flexible cortisol sensors placed simultaneously at four body locations armpit, lower back (LB), upper back (UB) and hand so that the pipeline can exploit inter-site agreement rather than relying on a single, possibly compromised, site. Table I summarises the cortisol concentration ranges associated with severe stress that are reported for the reference dataset used in this study [22].

TABLE I. CORTISOL CONCENTRATION RANGE ASSOCIATED WITH SEVERE STRESS

Body Region	Cortisol Range, Severe Stress
Armpit / LB / UB / Hand (multi-region avg.)	25 – 125 ng·mL ⁻¹
Axilla	100 – 1000 ng·mL ⁻¹

D. Pre-processing Unscented Trainable Kalman Filter

The raw, multi-site cortisol stream is corrupted by sensor drift, motion artefacts and occasional dropouts,

so it is first passed through an Unscented Trainable Kalman Filter (UTKF) [23]. Unlike the classical Kalman filter, which linearises the process model, UTKF propagates a small set of deterministically chosen sigma points through the true nonlinear observation model and reconstructs the posterior mean and covariance from their weighted statistics; the process- and measurement-noise covariances are themselves treated as trainable parameters and adapted from the observed innovation sequence rather than fixed by hand. For a state vector x_k capturing the underlying cortisol level and its rate of change, the filter alternates a prediction step,

$$\hat{x}_{k|k-1} = f(\hat{x}_{k-1|k-1}), P_{k|k-1} = \chi W \chi^T + Q_k$$

with a measurement-update step that folds in the noisy sensor reading z_k through the Kalman gain K_k , and Q_k is re-estimated online rather than fixed, which lets the filter tighten its trust in the model during stable periods and loosen it during sensor artefacts. The output of this stage is a smoothed, outlier-corrected cortisol trajectory per body site.

E. Feature Extraction Dual-Tree Complex Wavelet Transform

From the filtered signal, six time-domain statistical descriptors are extracted using a Dual-Tree Complex Wavelet Transform (DT-CWT) [24]. DT-CWT decomposes the signal through two parallel real discrete wavelet trees whose filters are designed so that the combined transform is approximately shift-invariant and directionally selective, which conventional real-valued DWT is not; these matters here because the exact temporal alignment of a cortisol pulse relative to the sampling grid is not controllable. The six descriptors computed per sub-band are listed in Table II.

TABLE II. TIME-DOMAIN STATISTICAL FEATURES EXTRACTED PER DT-CWT SUB-BAND

Feature	Definition
Mean	Average of sub-band coefficients
Median	Middle value of sorted coefficients
Std. deviation	Spread about the mean
Skewness	Asymmetry of the distribution
Kurtosis	Peakedness / tail weight
Coeff. of variation	Std. deviation ÷ mean

F. Classification Progressive Wasserstein GAN

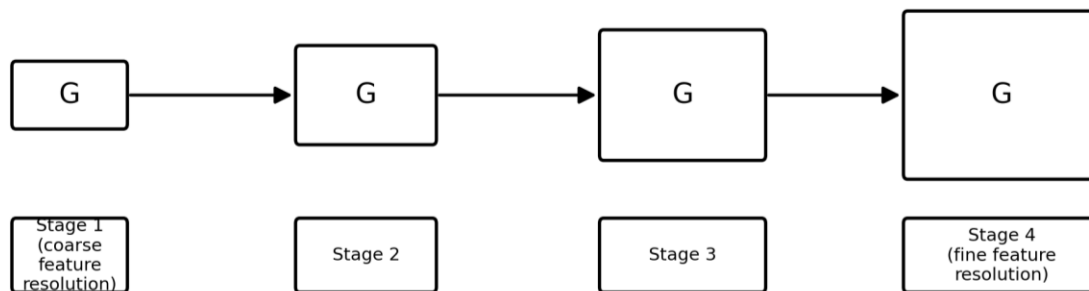
The extracted feature vectors are classified into Normal, Acute and Chronic stress by a Progressive Wasserstein Generative Adversarial Network (PWGAN) [25]. The generator G and discriminator (critic) D are grown together across resolution stages starting from a coarse representation of the feature vector and progressively adding layers that resolve finer structure which has been shown in other signal domains to stabilise adversarial training relative to training a full-depth network from the first iteration. Fig. 3 illustrates the growth schedule. The critic is trained to approximate the Wasserstein-1 distance between the real feature distribution and the

generator's distribution, and the adversarial objective is

$$\min_G \max_D E[D(x)] - E[D(G(z))] + \lambda \cdot E[(\|D(\hat{x})\|_2 - 1)^2]$$

where x is a real feature vector, z is latent noise, $\hat{x}$ is a random interpolate between real and generated samples, and the gradient-penalty term enforces the 1-Lipschitz constraint that the Wasserstein formulation requires. A lightweight auxiliary classification head attached to the discriminator's penultimate layer is trained jointly with a cross-entropy loss over the three severity classes, so that the adversarially-learned representation is also directly discriminative for the downstream three-class decision.

Fig. 3. Progressive growing strategy of the generator (G) in the proposed PWGAN, adding higher-resolution feature layers stage-wise



G. Optimization Golden Jackal Optimization Algorithm

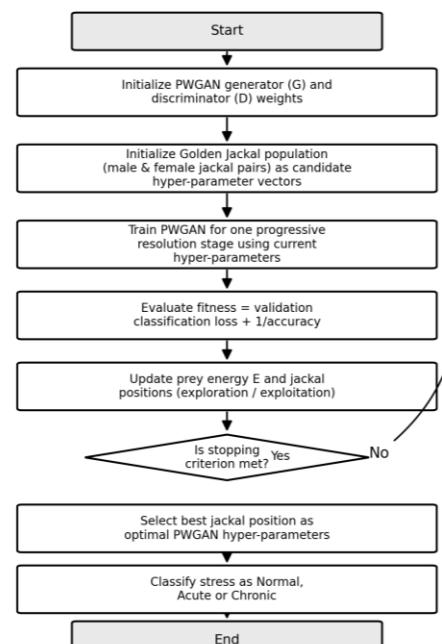
A known weakness of PWGAN in the form described above is that it offers no built-in mechanism for choosing its own learning rate, batch size or generator depth; these are ordinarily set by hand or by exhaustive grid search, both of which scale poorly. This paper uses the Golden Jackal Optimization Algorithm (GJOA) [26] to search this space automatically. GJOA models the cooperative hunting behaviour of male and female jackal pairs: each candidate hyper-parameter vector is a jackal position, the population alternates between an exploration phase driven by a prey energy term E that decays across iterations and an exploitation phase that refines positions around the current best solution, and male and female positions are updated jointly toward the prey (the best-found configuration) at each iteration:

$$E = E_0 (1 - t / T), \quad Y_i = Y_M - E \cdot |Y_M - r \cdot Y_prey|$$

where Y_M denotes the male jackal's position, Y_prey the current best solution, t the iteration index, T the

maximum number of iterations, and r a random vector in $[0, 1]$;

Fig. 2. Flowchart of GJOA-based hyper-parameter optimization for the progressive PWGAN



an analogous update governs the female jackal position Y_2 , and the pair's positions are averaged after each iteration. The fitness function minimised by GJOA is the PWGAN's validation classification loss augmented with an inverse-accuracy penalty, so that the search is pulled toward configurations that are both stable to train and accurate to classify with. Fig. 2 summarises the resulting optimisation loop, and Table V lists the search space explored for each hyper-parameter.

TABLE V. GJOA SEARCH SPACE FOR PWGAN HYPER-PARAMETERS

Hyper-parameter	Search Space	Encoding
Learning rate	$1e-5 - 1e-2$	Log-uniform
Batch size	$8 - 128$	Integer, 2^n
Generator depth	$2 - 6$ stages	Integer
Critic steps	$1 - 5$	Integer
Grad. -penalty λ	$1 - 20$	Uniform
Jackal population	$10 - 40$	Integer

H. Overall Algorithm

```

Algorithm 1 SLM-CS-PWGAN
Input: multi-site cortisol streams from  $N$  participants
Output: stress class  $\in \{\text{Normal, Acute, Chronic}\}$ 
1: Acquire cortisol signal at armpit, LB, UB, hand
2: Denoise each channel with UTKF (Sec. III-D)
3: Extract mean, median,  $\sigma$ , skewness, kurtosis, CoV per DT-CWT sub-band (Sec. III-E)
4: Initialise PWGAN weights and GJOA jackal population
5: while stopping criterion not met do
6:   Train PWGAN one progressive stage with current hyper-parameters
7:   Evaluate fitness on validation split
8:   Update jackal positions  $Y_1$ ,  $Y_2$  and prey energy  $E$ 
9: end while
10: Fix hyper-parameters at best jackal position
11: return arg-max class probability from trained PWGAN

```

I. Computational Complexity

Table VII summarises the per-window computational complexity of each pipeline stage, where n is the number of samples in an acquisition window, L the number of DT-CWT decomposition levels, d the PWGAN feature dimensionality, s the number of

progressive growth stages, and P the GJOA population size evaluated over T iterations. The UTKF and DT-CWT stages are linear to log-linear in the window length, so the acquisition and feature-extraction path is not the bottleneck; the dominant offline cost is the GJOA search, which is amortised because it is run once to fix the hyper-parameters and is not repeated at inference time, consistent with the implementation detail in Section IV-D.

TABLE VII. PER-WINDOW COMPUTATIONAL COMPLEXITY OF EACH STAGE

Stage	Complexity
UTKF denoising	$O(n)$ per channel
DT-CWT feature extraction	$O(n \log n)$ per window, level L
PWGAN inference	$O(d \cdot s)$ per window
PWGAN training (per epoch)	$O(B \cdot d \cdot s)$
GJOA hyper-parameter search	$O(P \cdot T \cdot d \cdot s)$ (offline, one-off)

IV. EXPERIMENTAL SETUP AND EVALUATION PROTOCOL

Because this paper reports the design of the SLM-CS-PWGAN pipeline, this section fixes the experimental protocol, baselines and metrics against which the framework is to be benchmarked once the full data-collection and training run is complete; the tables below specify the evaluation design rather than final results.

A. Dataset

The evaluation dataset consists of real-time, multi-site cortisol recordings from eight participants, sensed simultaneously at the armpit, lower back, upper back and hand, together with the reference severe-stress concentration ranges in Table I [22]. Each participant's recording session is segmented into fixed-length windows prior to DT-CWT feature extraction, and windows are labelled Normal, Acute or Chronic according to the protocol under which the data were collected.

B. Baseline Methods

The proposed SLM-CS-PWGAN is to be benchmarked against three representative literature

pipelines, re-implemented under a common evaluation split so that differences in reported performance reflect the modelling approach rather than dataset differences.

Table IV summarises the characteristics of each baseline as reported in the source studies.

TABLE IV. BASELINE METHODS USED FOR COMPARATIVE EVALUATION

Method	Input Signals	Model	Reported Strength	Reported Limitation
SLM-CS-LSTM [16]	EDA, BVP (cortisol-labelled)	Feature-selected classical ML	Low computation time	Low accuracy
SLM-CS-RF [17]	EDA, BVP, IBI, ST (cortisol-labelled)	Wrist-worn multi-signal RF	High precision	Low F1-score
SLM-CS-LOOCV [18]	ECG-derived HR, RR, EMG, skin conductance	Statistical sensitivity analysis + LOOCV	Low computation time	Low sensitivity

C. Evaluation Metrics

Eight metrics are used to characterise classification quality and computational cost, defined in Table III in terms of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) computed per class and macro-averaged across the three severity classes.

TABLE III. EVALUATION METRICS AND THEIR DEFINITIONS

Metric	Definition
Accuracy	$(TP+TN)/(TP+TN+FP+FN)$
Precision	$TP/(TP+FP)$
Sensitivity	$TP/(TP+FN)$
Specificity	$TN/(TN+FP)$
F1-score	$2PR/(P+R)$
Error rate	$(FP+FN)/All$
ROC-AUC	Area under ROC (macro, 1-vs-rest)
Comp. time	Inference time / window

D. Implementation Details

The pipeline is implemented with the PWGAN trained in mini-batches drawn from the DT-CWT feature vectors, with early stopping monitored on a held-out validation split. GJOA is run with the search space in Table V, and the best hyper-parameter vector found across the optimisation budget is frozen before the final train/test evaluation, so that the reported computational-time metric reflects inference cost

rather than search cost subject was never seen during training. To determine whether an observed difference between SLM-CS-PWGAN and a baseline is likely to reflect a genuine effect rather than fold-to-fold noise, per-fold accuracy and F1-score differences will be tested with a paired

E. Statistical Validation Plan

Because the participant cohort is small (eight subjects), a leave-one-subject-out (LOSO) cross-validation scheme is adopted for the benchmarking run rather than a single random train/test split, so that every reported metric is an average over eight folds in which the test

Wilcoxon signed-rank test at the 0.05 significance level, which does not assume normally distributed per-fold scores. A Bonferroni correction will be applied across the three pairwise baseline comparisons (against SLM-CS-LSTM, SLM-CS-RF and SLM-CS-LOOCV) to control the family-wise error rate.

F. Ethical Considerations

Because the study involves real-time biochemical sensing on human participants, data collection is conducted under informed consent, with participants free to withdraw at any point and sensor placement restricted to the four non-invasive sites described in Section III-C. Cortisol recordings and derived features are anonymised prior to storage, and only de-identified feature vectors are used for PWGAN training, so that no raw biochemical trace is retained in a form that could be linked back to an individual participant beyond the original consented dataset [22].

V. NOVELTY DISCUSSION

The design choices in Section III depart from the reviewed baselines [16]–[21] along several dimensions simultaneously rather than a single one, which is summarised qualitatively in Table VI. Three aspects are worth calling out explicitly.

First, cortisol is used here as a direct input signal to the classifier rather than solely as a ground-truth label collected via saliva assay, and it is sensed at four body sites simultaneously rather than one, which lets the pipeline exploit cross-site agreement instead of trusting a single noisy channel.

Second, denoising and feature extraction are treated as first-class pipeline stages rather than left implicit in the classifier: the UTKF adapts its own noise model online instead of using a fixed covariance, and DT-CWT gives descriptors that do not shift under small misalignments between the sensor clock and the physiological event, unlike the raw statistical features used directly on EDA/BVP signals in [16], [17].

Third, and most significantly, the classifier's own hyper-parameters are searched automatically by

GJOA rather than fixed by the designer, which is a capability none of the six reviewed baselines has; every baseline in [16]–[21] uses a fixed-architecture classical or shallow-learning model tuned by hand.

A. Threats to Validity

Two limitations should be weighed against the design's novelty. First, the eight-participant pilot cohort is small relative to the 19–40-subject cohorts used in [16], [17], which is why Section IV-E specifies leave-one-subject-out validation with a non-parametric significance test rather than a single held-out split; a larger cohort remains necessary before the framework's accuracy claims can be considered externally valid. Second, GJOA's search budget T and population size P trade optimisation quality against wall-clock search time, and Table V's search space was chosen from ranges commonly reported for adversarial-network tuning rather than from a pilot sweep on this specific dataset; the sensitivity of the final classifier to this budget is left as part of the benchmarking study in Section IV.

TABLE VI. QUALITATIVE FEATURE COMPARISON: PROPOSED VS. REVIEWED METHODS

Capability	[16]	[17]	[18]	[19]	[20]	[21]	Proposed
Direct cortisol as model input	N	N	N	Y	N	N	Y
Multi-site sensing	N	N	N	N	N	N	Y
Adaptive (trainable) filtering	N	N	N	N	N	N	Y
Shift-invariant wavelet features	N	N	N	N	N	N	Y
Adversarial generative classifier	N	N	N	N	N	N	Y
Metaheuristic hyper-parameter search	N	N	N	N	N	N	Y
Three-class severity granularity	N	N	N	N	N	N	Y

VI. CONCLUSION AND FUTURE WORK

This paper presented SLM-CS-PWGAN, a stress-severity classification pipeline that fuses multi-site wearable cortisol sensing, Kalman-filter denoising, dual-tree wavelet feature extraction and a Golden-Jackal-tuned Progressive Wasserstein GAN classifier. The design targets the accuracy-versus-computation-time trade-off observed consistently across six reviewed baselines by treating both the input signal quality and the classifier's own configuration as things to be optimised, rather than fixed by hand. A full evaluation protocol dataset, baselines and eight quantitative metrics was defined so that the framework

can be benchmarked directly against SLM-CS-LSTM, SLM-CS-RF and SLM-CS-LOOCV under matched conditions.

Future work will (i) complete the benchmarking run defined in Section IV and report the resulting accuracy, precision, sensitivity, specificity, F1-score, error rate, ROC-AUC and computation-time figures; (ii) extend the participant cohort beyond the current eight-subject pilot to improve statistical power and assess cross-subject generalisation with a leave-one-subject-out protocol; (iii) integrate the MicroSweat microfluidic collection layer [15] directly with the sensing array to reduce the motion-artefact burden placed on the UTKF stage; and (iv) explore an on-

device, quantised version of the trained PWGAN to assess feasibility for real-time, continuous operation on a wearable platform rather than offline analysis.

Beyond these near-term steps, the modular structure of the pipeline denoising, shift-invariant feature extraction, and metaheuristically-tuned generative classification as separable stages is not intrinsically specific to cortisol or to stress. The same structure could plausibly be re-targeted at other sweat or interstitial-fluid biomarkers with similarly poor temporal resolution, provided a comparable reference dataset is available, which is a direction we intend to explore once the present benchmarking study is complete.

ACKNOWLEDGMENT

The authors thank the participants who took part in the cortisol data-collection sessions, and gratefully acknowledge the use of the publicly available reference cortisol dataset [22] in defining the evaluation ranges used in this study.

REFERENCES

- [1] R. K. Nath and H. Thapliyal, "Wearable health monitoring system for older adults in a smart home environment," in *Proc. IEEE Comput. Soc. Annu. Symp. VLSI (ISVLSI)*, 2021, pp. 390–395.
- [2] S. Shahub, S. Upasham, A. Ganguly, and S. Prasad, "Machine learning guided electrochemical sensor for passive sweat cortisol detection," *Sens. Bio-Sens. Res.*, vol. 38, p. 100527, 2022.
- [3] A. Mohammadi, M. Fakharzadeh, and B. Baraeinejad, "An integrated human stress detection sensor using supervised algorithms," *IEEE Sens. J.*, vol. 22, no. 8, pp. 8216–8223, 2022.
- [4] G. Ibáñez-Redin, O. S. Duarte, G. R. Cagnani, and O. N. Oliveira, "A machine learning approach in wearable technologies," in *Machine Learning for Advanced Functional Materials*. Singapore: Springer, 2023, pp. 43–74.
- [5] F. Mohammadi et al., "Recent advances in aptamer-based platforms for cortisol hormone monitoring," *Talanta*, p. 125010, 2023.
- [6] D. Pérez and J. Orozco, "Wearable electrochemical biosensors to measure biomarkers with complex blood-to-sweat partition such as proteins and hormones," *Microchim. Acta*, vol. 189, no. 3, p. 127, 2022.
- [7] G. Vos, K. Trinh, Z. Sarnyai, and M. R. Azghadi, "Generalizable machine learning for stress monitoring from wearable devices: A systematic literature review," *Int. J. Med. Inform.*, p. 105026, 2023.
- [8] Q. Lin, T. Li, P. M. Shakeel, and R. D. J. Samuel, "Advanced artificial intelligence in heart rate and blood pressure monitoring for stress management," *J. Ambient Intell. Humaniz. Comput.*, vol. 12, pp. 3329–3340, 2021.
- [9] E. Demir, H. S. Sas, M. Isik, E. A. Gungor, and K. Davut, "A numerical methodology for monitoring stress distributions and experimental proof of the concept on metal embedded thin polymer-matrix composites using X-ray diffraction," *Thin-Walled Struct.*, vol. 173, p. 108942, 2022.
- [10] T. Bhavani, P. VamseeKrishna, C. Chakraborty, and P. Dwivedi, "Stress classification and vital signs forecasting for IoT-health monitoring," *IEEE/ACM Trans. Comput. Biol. Bioinform.*, 2022.
- [11] C. Galvan, R. Montiel, K. Lorenz, J. Carter, N. I. Hossain, and S. Tabassum, "A microneedle-based leaf patch with IoT integration for real-time monitoring of salinity stress in plants," in *Proc. IEEE 15th Dallas Circuits Syst. Conf. (DCAS)*, 2022, pp. 1–2.
- [12] G. Karuppaiah, M. H. Lee, S. Bhansali, and P. Manickam, "Electrochemical sensors for cortisol detection: Principles, designs, fabrication, and characterisation," *Biosens. Bioelectron.*, p. 115600, 2023.
- [13] M. Pali et al., "CATCH (Cortisol Apta WATCH): 'Bio-mimic alarm' to track anxiety, stress, immunity in human sweat," *Electrochim. Acta*, vol. 390, p. 138834, 2021.
- [14] S. Koudela-Hamila, J. Smyth, P. Santangelo, and U. Ebner-Priemer, "Examination stress in academic students: A multimodal, real-time, real-life investigation of reported stress, social contact, blood pressure, and cortisol," *J. Am. Coll. Health*, vol. 70, no. 4, pp. 1047–1058, 2022.
- [15] M. Zea et al., "Electrochemical sensors for cortisol detection: Almost there," *TrAC Trends Anal. Chem.*, vol. 132, p. 116058, 2020.

- [16] R. K. Nath, H. Thapliyal, and A. Caban-Holt, "Machine learning based stress monitoring in older adults using wearable sensors and cortisol as stress biomarker," *J. Signal Process. Syst.*, pp. 1–13, 2021.
- [17] R. K. Nath and H. Thapliyal, "Smart wristband-based stress detection framework for older adults with cortisol as stress biomarker," *IEEE Trans. Consum. Electron.*, vol. 67, no. 1, pp. 30–39, 2021.
- [18] T. Iqbal, P. Redon-Lurbe, A. J. Simpkin, A. Elahi, S. Ganly, W. Wijns, and A. Shahzad, "A sensitivity analysis of biophysiological responses of stress for wearable sensors in connected health," *IEEE Access*, vol. 9, pp. 93567–93579, 2021.
- [19] H. Kim, D. Jung, H. Choi, and T. Hong, "Advanced prediction model for individual thermal comfort considering blood glucose and salivary cortisol," *Build. Environ.*, vol. 224, p. 109551, 2022.
- [20] S. Aristizabal, K. Byun, N. Wood, A. F. Mullan, P. M. Porter, C. Campanella, A. Jamrozik, I. Z. Nenadic, and B. A. Bauer, "The feasibility of wearable and self-report stress detection measures in a semi-controlled lab environment," *IEEE Access*, vol. 9, pp. 102053–102068, 2021.
- [21] J. Clark, R. K. Nath, and H. Thapliyal, "Machine learning based prediction of future stress events in a driving scenario," in *Proc. IEEE 7th World Forum Internet Things (WF-IoT)*, 2021, pp. 455–458.
- [22] Mendeley Data, "Wearable cortisol stress dataset," OnlineOnline. Available: <https://data.mendeley.com/datasets/wnshbvdxs2/1>.
- [23] Y. Yu, Q. Li, and H. Zhang, "Unscented trainable Kalman filter based on deep learning method considering incomplete information," *IEEE Access*, 2023.
- [24] R. Sharbati, F. Khoshnoudian, M. Koopialipoor, and M. M. Tahir, "Applying dual-tree complex discrete wavelet transform and gamma modulating function for simulation of ground motions," *Eng. Comput.*, vol. 37, pp. 1519–1535, 2021.
- [25] G. Wang and X. Hu, "Low-dose CT denoising using a progressive Wasserstein generative adversarial network," *Comput. Biol. Med.*, vol. 135, p. 104625, 2021.
- [26] N. Chopra and M. M. Ansari, "Golden jackal optimization: A novel nature-inspired optimizer for engineering applications," *Expert Syst. Appl.*, vol. 198, p. 116924, 2022.