

SELF-CARE: Selective Fusion with Context-Aware Low-Power Edge Computing for Stress Detection

Nafiu Rashid^{*1}, Trier Mortlock^{*2}, Mohammad Abdullah Al Faruque^{1,2}

¹*Department of Electrical Engineering and Computer Science*

²*Department of Mechanical and Aerospace Engineering*

University of California, Irvine, California, United States

{nafiu, tmortloc, alfaruqu}@uci.edu

Abstract—Detecting human stress levels and emotional states with physiological body-worn sensors is a complex task, but one with many health-related benefits. Robustness to sensor measurement noise and energy efficiency of low-power devices remain key challenges in stress detection. We propose SELF-CARE, a fully wrist-based method for stress detection that employs context-aware selective sensor fusion that dynamically adapts based on data from the sensors. Our method uses motion to determine the context of the system and learns to adjust the fused sensors accordingly, improving performance while maintaining energy efficiency. SELF-CARE obtains state-of-the-art performance across the publicly available WESAD dataset, achieving 86.34% and 94.12% accuracy for the 3-class and 2-class classification problems, respectively. Evaluation on real hardware shows that our approach achieves up to 2.2× (3-class) and 2.7× (2-class) energy efficiency compared to traditional sensor fusion.

Index Terms—Stress detection, edge computing, energy efficiency, sensor fusion

I. INTRODUCTION

The future of smart healthcare requires dependable sensor systems that can operate at increased levels of autonomy under energy constraints while providing valuable health-related information. One area gaining significant attention is *affective computing*, or the ability for machines to understand human emotional states. Stress detection is one example of affective computing that allows machines to detect stress levels within humans, which has a myriad of implications for healthcare science [1]. Stress can be interpreted as a physiological state that is triggered by chemical or hormonal surges during moments of physical, cognitive, emotional, or acute challenges [2].

Stress detection via physiological sensor data has been widely studied [3]–[6]. Physics-based models cannot relate this sensor data to explicit stress states, so classical machine learning models (random forests, decision trees, etc.) or deep learning models (convolutional neural networks, long short-term memory, etc.) are often used to perform stress classification as the models learn over labeled datasets [7]–[10]. However, classical machine learning models are more commonly used than deep learning approaches due to computational complexity and explainability [11], [12].

Methods using sensor fusion across multi-modal physiological data have been commonly used to increase performance

of emotion recognition [12]. *Early*, or feature-level, fusion focuses on combining data at the raw-data level compared to *late*, or decision-level, fusion that combines the final outputs of a system. Even methods that employ both early and late fusion are noticeably limited, since they have static architectures that cannot adapt to changing contexts [13]. Another key challenge in using these physiological signals is that they are susceptible to large amounts of noise during physical motion. Fusing such noisy measurements can subsequently degrade the classification performance [11]. Lastly, there is a lack of focus in stress detection to evaluate feasibility for edge (on-device) computing [14] as solutions should be energy-efficient and capable of running on resource-constrained devices [15].

Key research challenges arise from the current methods for stress detection, notably: (i) how to develop an adaptive architecture that alters the fusion schema depending on the current context; (ii) how to utilize measurements from the motion sensors to model the context; (iii) how to achieve comparable results to higher fidelity chest-worn devices while using more energy-efficient wrist-worn sensors; and (iv) how to incorporate temporal aspects into the stress classification problem that can further improve accuracy.

To address these challenges, we propose SELF-CARE, a fully wrist-based solution that models context as a function of motion and proposes a selective sensor fusion that adapts based on this learned context. SELF-CARE outperforms existing solutions for stress detection, while also providing energy efficiency suitable for edge computing on the wrist. We present the following key contributions:

- 1) We introduce a selective sensor fusion method that learns context based on motion, and dynamically adjusts the sensor fusion performed to maximize classification performance while ensuring energy efficiency.
- 2) We propose a late fusion technique for classification using a Kalman filter that incorporates temporal dynamics.
- 3) We validate our methodology on the WESAD dataset, showing that SELF-CARE achieves state-of-the-art performance for the 3-class and 2-class stress detection problems while using only wrist-worn sensors.
- 4) Experimental evaluation on real hardware shows that our SELF-CARE methodology is feasible for on-device, energy-efficient computing.

^{*}Both authors contributed equally to this research.

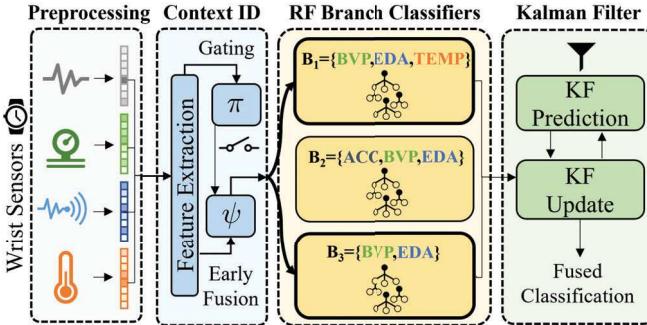


Fig. 1. Proposed SELF-CARE Architecture. In this depiction four types of wrist-worn sensors are used, the gating model selects two branches given the context, a Random Forest classifier is used for the branch models, and a Kalman filter is used for the late fusion over the two selected branches.

II. METHODOLOGY

In this section we detail SELF-CARE, depicted in Fig. 1. Our proposed method performs stress classification given sensor measurements from four wrist sensing modalities: tri-axis accelerometer (ACC), blood volume pulse (BVP), electrodermal activity (EDA), skin temperature (TEMP). SELF-CARE uses the four main blocks: (i) preprocessing, (ii) context identification, (iii) branch classifiers, and (iv) late fusion.

A. Preprocessing Step

SELF-CARE takes in as inputs data from any number of heterogeneous physiological sensors. Preprocessing is used over the raw, unfiltered sensor data by applying various filters (e.g., band-pass filters or lowpass filters) to the input data to reduce sensor noises and more easily extract important features. The preprocessing performed over each sensing modality follows that performed in [9].

B. Context Identification

1) *Feature Extraction*: The purpose of the context identification block is to select the branch classifier(s) based on the context of the motion. It first extracts only ACC features as they are directly related to the relative motion of the test subject. These features are then processed by the gating model to select the best performing branch. The feature extraction of the three other modalities takes place after the gating model has selected which branch(es) will be executed. We refer readers to [6] for the full list of features per sensor.

2) *Gating Model (π)*: The gating model trains a classifier that uses the ACC features as inputs to select one of the available branch classifiers for branches $B_1=\{\text{BVP, EDA, TEMP}\}$; $B_2=\{\text{ACC, BVP, EDA}\}$; $B_3=\{\text{BVP, EDA}\}$. A Decision Tree (DT) classifier is used for our gating model, as it is lightweight and adds minimum overhead for our architecture. Note that, for each round of leave-one-subject-out (LOSO) validation, only training data is used to generate gating labels. Additionally, one, two or all the final classifiers may be selected for final classification depending on the value of δ , detailed next.

3) *Performance-Energy Trade-off (δ)*: An important feature of SELF-CARE is its ability to balance constraints between performance and energy. We introduce the term δ that aids the gating decision in considering this trade-off. The gating model outputs prediction probabilities for the available branches with \bar{b} representing the maximum probability branch. δ has a range between 0 and 1, representing the range in which non-maximum branches are selected by allowing branches with probabilities greater than $\bar{b} - \delta$ to be also selected. Lower δ values indicate tighter energy constraints, with $\delta = 0$ indicating that only the highest probability branch from the gating classifier is selected, while higher δ values allow more branches to be selected, with $\delta = 1$ indicating that all possible branches are selected. For our 3-class (2-class) classification problem we set $\delta = 0.4$ ($\delta = 0.1$).

4) *Early Fusion (ψ)*: Once the branches are selected after applying δ on the gating model decision, the features for those branches will be extracted and concatenated together to be passed to the corresponding classifiers (branches). In the example in Fig. 1, B_1 and B_3 are the selected branches from the gating model. The features from BVP, EDA, and TEMP signals are concatenated together using early fusion and fed to the branch classifier for B_1 , with B_3 operating in similar fashion for its sensor modalities.

C. Branch Classifiers

Next, the corresponding branch classifier(s) is (are) used to perform classification of the segment. To train the individual branch classifiers within SELF-CARE we train using different combinations of input sensor data. For our analysis, we use five different early fusion combinations of wrist sensors as input branches - $B_1=\{\text{BVP, EDA, TEMP}\}$; $B_2=\{\text{ACC, BVP, EDA}\}$; $B_3=\{\text{BVP, EDA}\}$; $B_4=\{\text{ACC, BVP}\}$; $B_5=\{\text{ACC, EDA}\}$. Each branch is evaluated on five different machine learning classifiers — Decision Tree (DT), Random Forest (RF), AdaBoost (AB), Linear Discriminant Analysis (LDA), K-Nearest Neighbor (KNN). The classifiers are chosen to ensure a fair comparison with the original WESAD work [6]. Additionally, the low complexity of the classifiers makes SELF-CARE suitable for wearable devices. Following the work in [6], we use same configurations for the classifiers. Out of the 25 (5 branches x 5 classifiers per branch) possible branch classifiers, the branches with the minimum training loss are selected to be used within SELF-CARE. Each selected branch outputs a classification prediction to be fused by the late fusion method.

D. Late Fusion Method

Here we present our *Kalman filter-based* method for classification over an ensemble of classifiers, although we claim that any applicable late fusion method is supported within SELF-CARE. In the context of our problem, we consider a Kalman filter approach towards the multi-class classification problem like in [16], however, we additionally model the temporal dynamics in the stress classification problem for each sample. The unknown state our filter is attempting to estimate is the probability of each class during each segment. Thus,

TABLE I
OVERALL PERFORMANCE COMPARISON OF RELATED WORKS USING LOSO VALIDATION

Modality Used	3-Class			2-Class			Wrist Only	Wrist Computing
	Best Model	Macro F1	Accuracy	Best Model	Macro F1	Accuracy		
Related Works								
All (Wrist+Chest) [6]	AB	68.85	79.57	LDA	90.74	92.28	No	No
All Wrist [6]	AB	64.12	75.21	RF	84.11	87.12	Yes	Yes
All Wrist + Trans. Chest [7]	GAN-RF	74.5	81.4	GAN-RF	89.7	92.1	No	No
All Wrist [8]	DNN	-	83.43	DNN	-	93.14	Yes	No
BVP (Wrist) [9]	HCNN	64.15	75.21	HCNN	86.18	88.56	Yes	No
Selected Branch Classifiers [Ours]								
$B_1 = \{\text{BVP, EDA, TEMP}\}(\text{Wrist})$	RF	62.73	76.62	RF	84.66	89.01	Yes	Yes
$B_2 = \{\text{ACC, BVP, EDA}\}(\text{Wrist})$	RF	62.88	77.71	RF	85.08	88.76	Yes	Yes
$B_3 = \{\text{BVP, EDA}\}(\text{Wrist})$	RF	61.02	73.96	RF	86.37	89.33	Yes	Yes
Traditional Late Fusion [Ours]								
Soft-voting (B_1, B_2, B_3)	RF	63.75	78.79	RF	87.09	90.00	Yes	Yes
Hard-voting (B_1, B_2, B_3)	RF	64.02	78.70	RF	87.17	89.89	Yes	Yes
SELF-CARE [Ours]								
Kalman (B_1, B_2, B_3)	RF	71.97	86.34	RF	92.93	94.12	Yes	Yes

we define this \mathbf{x} as a c dimensional vector of estimated class probabilities. Additionally, the predictions from each separate classifier are the measurements \mathbf{z} , which are processed sequentially per time step. For the 3-class (2-class) problem, we initialize $\mathbf{x}_0 = [0.8, 0.1, 0.1]^\top$ ($\mathbf{x}_0 = [0.8, 0.2]^\top$) with estimation error covariance matrix $\mathbf{P}_0 = 0.01 \cdot \mathbf{I}_{3 \times 3}$ ($\mathbf{P}_0 = 0.01 \cdot \mathbf{I}_{2 \times 2}$). The state transition matrix and measurement matrix are identity matrices for the respective problems. The process noise for both problems is modeled as a discrete time white noise with variance set at $5e-4$. The measurement noise is modeled as a function of each measurement to allow the filter to adjust the confidence of the measurements according to each reported class probability: $\mathbf{R} = ((1 - \mathbf{z}) \cdot 2 \cdot \mathbf{I}_{3 \times 3})^2$ ($\mathbf{R} = ((1 - \mathbf{z})/2 \cdot \mathbf{I}_{2 \times 2})^2$). Lastly, a tunable threshold technique was used to process the measurements which involved (i) an ϵ parameter to select measurements which had a maximum predicted probability above the threshold and (ii) a γ factor to scale the measurements to account for the imbalanced class distribution in the dataset. This thresholding process allows for the filter to weight each measurement it receives with a differing degree of noise while also attempting to resolve issues that arise from imbalanced datasets. For the 3-class (2-class) problem, we set $\epsilon = 0.4$ ($\epsilon = 0.7$) and $\gamma = [.278, 1, 1]^\top$ ($\gamma = [.667, 1.1]^\top$). During 3-class classification, the prediction probabilities are generally lower as they are distributed across an additional class when compared to 2-class classification, thus calling for a lower ϵ threshold. To validate our Kalman-filter based method, we benchmark its performance against commonly used voting mechanisms for late fusion: *hard-voting* and *soft-voting* [17]. The method of hard-voting assigns the final class based on the class most commonly voted by each classifier, whereas soft-voting selects the class with the highest average value across all the classifiers.

III. EXPERIMENTAL ANALYSIS

A. Dataset Evaluation Metrics

SELF-CARE is validated on the publicly available WESAD dataset [6]. The dataset contains data for a total of 15 subjects

from both chest (RespiBAN) and wrist (Empatica E4) worn sensors. Our work focuses on stress detection using only wrist-based data, as we use the following sensors from the Empatica E4: ACC BVP, EDA, TEMP. The dataset has three types of classes related to emotional states, namely — baseline (neutral), amusement, and stress. For the 2-class problem, baseline and amusement are considered as the non-stress class. The filtered signals are segmented by a window of 60 seconds of data with a sliding length of 5 seconds following [7]. This gives a total of 6458 segments for each signal across all subjects of the WESAD dataset. The WESAD dataset is highly imbalanced in terms of the number of segments per class. For this reason, F1 score is also used along with accuracy to measure the classification performance. To ensure a fair comparison with other works, we use the macro F1 score.

B. Experimental Results

This section presents the performance of SELF-CARE for stress detection in 3-class and 2-class classification. We also demonstrate the energy efficiency of our approach in a ultra-low-power 32-bit microcontroller EFM32 Giant Gecko (EFM32GG-STK3700A) [18] representing a wearable device operating on the edge. The microcontroller has an ARM Cortex-M3 processor with a maximum clock rate of 48 MHz. It has 128 KB of RAM and 1 MB of flash memory.

1) *Performance Evaluation*: Table I shows the overall performance comparison of the related works against our proposed method. Authors in [6] explored different combinations of chest and wrist sensors across a variety of models. The results for three deep learning methods are also shown [7]–[9]. For our three selected branch classifiers, the soft- and hard-voting methods are applied, showing performance improvements compared to the individual branch classifiers for both 3-class and 2-class classification. Lastly, SELF-CARE using Kalman filter-based late fusion further improves the performance for 3-class and 2-class classification compared to these traditional late fusion methods. Despite using only wrist signals, SELF-CARE outperforms all other state-of-the-art works that use either wrist, chest, or both sensors for 3-class

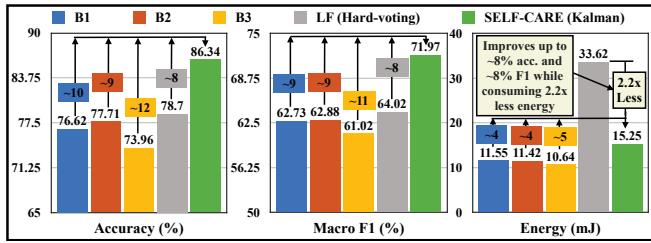


Fig. 2. Performance and Energy for 3-Class Classification

and 2-class classification. Only [7] achieves a better macro F1 score than SELF-CARE for 3-class classification. However, they use both wrist and translated chest features, and employ a computationally expensive GAN model, which is not suitable for wrist computing.

2) *Energy Evaluation*: As shown in Table I, traditional late fusion improves the performance compared to individual branch classifiers. However, it is not energy-efficient, as multiple classifiers need to be used simultaneously — unlike SELF-CARE that minimizes the number of classifiers selected for a given segment. We benchmark SELF-CARE with hard-voting late fusion, which is relatively more energy-efficient than soft-voting and shows similar performance to soft-voting. As shown in Fig. 2 for 3-class classification, SELF-CARE with $\delta = 0.4$ improves up to $\sim 8\%$ accuracy and $\sim 8\%$ F-1 score, while being $\sim 2.2\times$ energy-efficient compared to hard-voting. Similarly, for 2-class classification (Fig. 3), SELF-CARE with $\delta = 0.1$ outperforms hard-voting by up to $\sim 4\%$ accuracy and $\sim 6\%$ F-1 score while being $\sim 2.7\times$ energy-efficient. The higher energy efficiency for 2-class can be partially attributed to the lower $\delta = 0.1$, which reduces the use of multiple branches compared to $\delta = 0.4$ for 3-class. The higher δ for 3-class is chosen to prioritize performance over energy, as the 3-class problem is inherently more challenging than 2-class.

IV. CONCLUSION

In this paper we proposed SELF-CARE, a selective sensor fusion approach that uses context-aware, energy-efficient edge computing to perform stress detection. SELF-CARE models context as the motion of a subject and performs an intelligent gating mechanism to select which sensor fusion schema to use given a certain input. To the best of our knowledge, SELF-CARE achieves state-of-the-art performance on the WESAD dataset in terms of 3-class classification (**86.34%**) and 2-class classification (**94.12%**) in approaches that use LOSO validation. Furthermore, SELF-CARE achieves upto **2.2 \times** (3-class) and **2.7 \times** (2-class) energy efficiency with respect to comparable late fusion methods.

ACKNOWLEDGMENT

This work was partially supported by the National Science Foundation (NSF) under awards CMMI-1739503 and CCF-2140154. Any opinions, findings, conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the funding agencies.

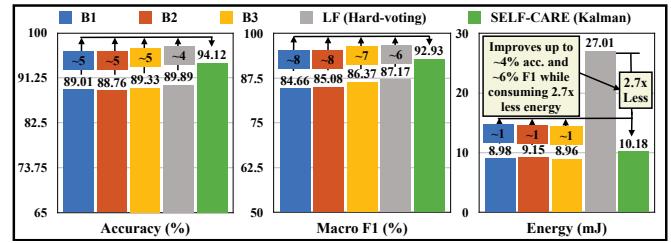


Fig. 3. Benchmarking on 2-Class Classification

REFERENCES

- [1] A. P. Association. (2020) Stress in America 2020: a national mental health crisis. [Online]. Available: <https://www.apa.org/news/press/releases/stress/2020/sia-mental-health-crisis.pdf>
- [2] D. S. Goldstein, “Adrenal responses to stress,” *Cellular and molecular neurobiology*, vol. 30, no. 8, pp. 1433–1440, 2010.
- [3] J. A. Healey and R. W. Picard, “Detecting stress during real-world driving tasks using physiological sensors,” *IEEE Trans. on intelligent transportation systems*, vol. 6, no. 2, pp. 156–166, 2005.
- [4] R. W. Picard, E. Vyzas, and J. Healey, “Toward machine emotional intelligence: Analysis of affective physiological state,” *IEEE Trans. on pattern analysis and machine intelligence*, vol. 23, no. 10, pp. 1175–1191, 2001.
- [5] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, “Deap: A database for emotion analysis; using physiological signals,” *IEEE transactions on affective computing*, vol. 3, no. 1, pp. 18–31, 2011.
- [6] P. Schmidt, A. Reiss, R. Dürichen, C. Marberger, and K. Van Laerhoven, “Introducing wesad, a multimodal dataset for wearable stress and affect detection,” in *Proceedings of the 20th ACM international conference on multimodal interaction*, 2018, pp. 400–408.
- [7] S. Samyoun, A. S. Mondol, and J. A. Stankovic, “Stress detection via sensor translation,” in *2020 16th International Conference on Distributed Computing in Sensor Systems (DCOSS)*. IEEE, 2020, pp. 19–26.
- [8] L. Huynh, T. Nguyen, T. Nguyen, S. Pirttikangas, and P. Siirtola, “Stressnas: Affect state and stress detection using neural architecture search,” in *Adj. Proc. of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2021, pp. 121–125.
- [9] N. Rashid, L. Chen, M. Dautta, A. Jimenez, P. Tseng, and M. A. Al Faruque, “Feature augmented hybrid cnn for stress recognition using wrist-based photoplethysmography sensor,” in *2021 43rd Annual Conference of the IEEE EMBC*. IEEE, 2021, pp. 2374–2377.
- [10] A. Ragav, N. H. Krishna, N. Narayanan, K. Thelly, and V. Vijayaraghavan, “Scalable deep learning for stress and affect detection on resource-constrained devices,” in *2019 18th IEEE Int. Conference On Machine Learning And Applications (ICMLA)*. IEEE, 2019, pp. 1585–1592.
- [11] P. Schmidt, R. Dürichen, A. Reiss, K. Van Laerhoven, and T. Plötz, “Multi-target affect detection in the wild: an exploratory study,” in *Proc. of the 23rd Int. Symposium on Wearable Computers*, 2019, pp. 211–219.
- [12] P. Bota, C. Wang, A. Fred, and H. Silva, “Emotion assessment using feature fusion and decision fusion classification based on physiological data: Are we there yet?” *Sensors*, vol. 20, no. 17, p. 4723, 2020.
- [13] A. V. Malawade, T. Mortlock, and M. A. A. Faruque, “HydraFusion: Context-aware selective sensor fusion for robust and efficient autonomous vehicle perception,” in *ICPPS*, 2022.
- [14] N. Rashid, B. U. Demirel, and M. A. Al Faruque, “Ahar: Adaptive cnn for energy-efficient human activity recognition in low-power edge devices,” *IEEE Internet of Things Journal*, 2022.
- [15] A. V. Malawade, T. Mortlock, and M. A. A. Faruque, “EcoFusion: Energy-aware adaptive sensor fusion for efficient autonomous vehicle perception,” in *DAC*, 2022.
- [16] A. Pakrashi and B. Mac Namee, “Kalman filter-based heuristic ensemble (kfhe): A new perspective on multi-class ensemble classification using kalman filters,” *Information Sciences*, vol. 485, pp. 456–485, 2019.
- [17] S. Oviatt, B. Schuller, P. Cohen, D. Sonntag, G. Potamianos, and A. Krüger, *The handbook of multimodal-multisensor interfaces, Volume 2: Signal processing, architectures, and detection of emotion and cognition*. Morgan & Claypool, 2018.
- [18] S. Labs. (2021) Efm32™ giant gecko 32-bit microcontroller. [Online]. Available: <https://www.silabs.com/mcu/32-bit/efm32-giant-gecko>