BST-GATNet: Bidirectional Spatio-Temporal Graph Attention Network for Driver Stress Detection

Chenpei Xie, Yiting Wei*, *Student Member, IEEE*,
Mostafa Haghi, *Member, IEEE*, Nima TaheriNejad, *Member, IEEE*

***Abstract*—Accurate, real-time driver stress detection under on-device computational constraints is a critical capability for intelligent vehicles and human-centric safety systems. While multimodal physiological signals provide rich information for stress recognition, robustly modeling their time-varying cross-signal interactions remains challenging in resource-constrained deployment scenarios. To address this challenge, we propose Bidirectional Spatio-Temporal Graph Attention Network (BST-GATNet), a lightweight spatio-temporal learning framework that explicitly models cross-modal physiological dependencies while being designed for real-time, on-device execution. BST-GATNet integrates a Transformer-enhanced graph attention network (GAT) module to capture instantaneous inter-signal relationships with a bidirectional long short-term memory network (Bi-LSTM) module to model temporal dynamics. Evaluated on the real-world AffectiveROAD dataset, BST-GATNet achieves 88.43% accuracy and consistently outperforms existing methods. Under a controlled CPU benchmark setting, the proposed model operates at millisecond-level latency with a P99 latency of 7.31 ms, validating its suitability for real-time in-vehicle deployment and demonstrating a favorable accuracy–efficiency trade-off. Overall, BST-GATNet enables scalable, low-cost, and privacy-preserving driver stress monitoring through efficient on-device inference, supporting objective continuous assessment and contributing toward sustainable and equitable health monitoring technologies.**

***Index Terms*—Graph Attention Network, Multimodal signals, Driver Stress Detection, Spatio-Temporal Modeling**

I. Introduction

The complexity of modern traffic has made driver stress a pervasive threat to road safety. This threat is primarily cognitive, as stress impairs executive functions: reducing hazard perception, delaying reaction times, and inducing attentional narrowing [1]–[3]. Since such impairments often occur without the driver's awareness, the objective identification of stress is essential for enhancing driving safety.

Existing solutions, primarily integrated within driver monitoring systems (DMS) [4], have made initial attempts to address this challenge. These systems often rely on cameras to analyze visual cues or use Controller Area Network (CAN) bus data to interpret control behaviors such as steering and braking [5]. However, their reliability is inherently limited, as they depend on indirect behavioral inferences [6]. These proxies can easily be confounded by external factors. For instance, a sudden manoeuvre may reflect an adaptive response rather than stress, and interpreting visual cues such as facial expressions or eye movements becomes unreliable when key features are obscured by poor lighting or occlusion [7]. This underscores the need for more direct and reliable measurement approaches.

Physiological signals provide a direct means of assessing the driver's internal state by capturing stress responses regulated by the autonomic nervous system (ANS). Core markers include heightened electrodermal activity (EDA) resulting from sympathetic activation of sweat glands, elevated heart rate (HR) and blood volume pulse (BVP) driven by sympathetic arousal, and changes in skin temperature (TEMP) caused by altered blood flow distribution [8]. These physiological responses are further complemented by accelerometer (ACC) data, which reflects stress-induced modifications in body dynamics such as shifts in torso and head movements [9]. With the increasing availability of wearable sensors, the continuous acquisition of such multimodal data in real-world driving conditions is becoming increasingly practical [10].

Recent years have seen a growing body of research devoted to stress assessment through physiological signals. Early studies, such as Bui et al. [11], demonstrated the feasibility of multimodal approaches by training classifiers on statistical features extracted from signals including HR, EDA, and TEMP. However, the computational complexity of these methods poses significant challenges for the increasing demand for on-device deployment. Subsequent efforts have shifted toward the design of lightweight architectures, often at the expense of informational richness or inter-modal dynamics. For example, Jaiswal et al. [12] proposed TinyStressNet, which achieves high efficiency on resource-constrained platforms by restricting input to the EDA signal alone. Similarly, Cvetkovic et al. [13] introduced BioEdgeNet, which employs photoplethysmogram and ACC signals but processes them in parallel streams to minimize computational cost and thereby omitting explicit cross-modal fusion. Although these models represent substantial progress in on-device stress detection, their design strategies — either reducing to a single modality or isolating multiple modalities — overlook the dynamic interdependencies inherent in physiological processes. This limitation is critical, as the interpretation of individual signals is often ambiguous and sensitive to inter-subject variability. Consequently, there remains a pressing need for a framework that is both computationally efficient and capable of explicitly modeling cross-signal dependencies, enabling more robust and

Chenpei Xie, Yiting Wei, Mostafa Haghi and Nima TaheriNejad are with the EclectX Team, Institute of Computer Engineering, Heidelberg University, Heidelberg, Germany (*Corresponding author: Yiting Wei, email: yiting.wei@uni-heidelberg.de)

personalized stress detection.

To address this limitation, we proposed a solution for driver stress detection that captures complex spatio-temporal dependencies among physiological signals. By integrating graph-based relational modeling with attention-driven temporal learning, the proposed method enhances both accuracy and efficiency. The main contributions are as follows:

1) **Bidirectional Spatio-Temporal Graph Attention Network (BST-GATNet) Architecture.** BST-GATNet integrates a Transformer-enhanced graph attention network (GAT) module with a bidirectional long short-term memory network (Bi-LSTM) temporal model to explicitly capture cross-signal dependencies and their temporal evolution from multimodal physiological inputs. This design enables joint modeling of inter-signal relationships and time-varying dynamics for driver stress recognition.
2) **Accuracy–Efficiency Trade-off for On-device Execution.** The proposed architecture is designed to balance predictive performance with practical runtime constraints. Early temporal downsampling and compact embeddings reduce computational cost, while attention-based graph modeling preserves informative cross-signal structure. As a result, BST-GATNet maintains strong stress detection accuracy while operating at millisecond-level latency under a controlled CPU benchmark, demonstrating feasibility for on-device deployment.

II. Dataset

In this study, we selected the AffectiveROAD dataset [14], which was collected during a real-world driving protocol designed to mirror a typical daily commute. Participants drove a 31km metropolitan route in their own vehicles to ensure familiarity and minimize confounding variables. A key strength of this dataset lies in its meticulous, driver-in-the-loop annotation process. During each session, an experimenter provided continuous stress ratings on a scale of 0 to 1. Afterward, each driver reviewed synchronized videos of their session and the experimenter's ratings, validating or correcting the stress labels to align with their own perceived experience. This validation step ensures a high-quality, fine-grained ground truth that reflects the driver's subjective state.

The dataset comprises 13 distinct driving sessions from 9 unique drivers, as shown in Fig. 1. The physiological signals used in this work include EDA and TEMP (4 Hz), BVP (64 Hz), ACC (32 Hz), and HR (1 Hz). Together, these characteristics provide a realistic and well-annotated setting for studying multimodal stress detection, while the repeated-session structure enriches the dataset with additional variation under comparable driving conditions.

III. Our Proposed Method

A. Data Preprocessing

Our preprocessing pipeline first transforms the raw physiological data into a structured format. All raw signals from the E4 wristband (EDA, TEMP, BVP, HR, ACC) were resampled

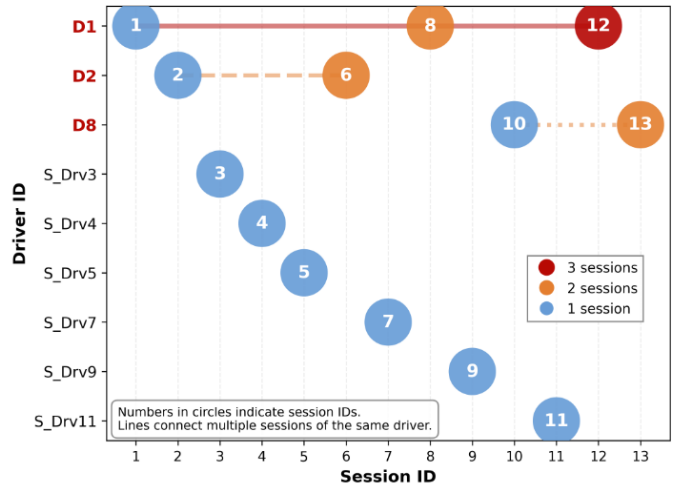

Fig. 1. Overview of the repeated-measures experimental design in the AffectiveROAD dataset.

to a uniform frequency of 64 Hz using cubic spline interpolation to ensure temporal alignment with the continuous stress annotations. No explicit filtering was applied; instead, high-frequency noise and minor signal fluctuations were implicitly attenuated through the model's multi-scale convolutional layers. Finally, this continuous data was segmented into 30-second windows with a 15-second step (50% overlap) to capture dynamic changes in the driver's state.

B. BST-GATNet Architecture

We propose BST-GATNet, a lightweight two-stream architecture for driver stress detection (Fig. 2). The multimodal physiological input is first organized into a graph representation, on top of which cross-modal spatio-temporal interactions are learned. The model then combines a dynamic spatio-temporal stream operating on raw signals with a static feature stream built from low-frequency statistical descriptors to jointly capture fast transients and slow trends; their outputs are fused for the final stress classification. To further model bidirectional temporal dependencies, a Bi-LSTM is introduced as a temporal refinement layer after the attention-based interaction modeling.

1) Graph Signal Modeling.: We represent the multimodal physiological input as a *5-node graph*, where each node corresponds to one modality channel (EDA, TEMP, BVP, HR, ACC). Rather than specifying a predefined adjacency matrix derived from physiological priors or statistical correlations, we assume an implicit fully-connected topology. Inter-signal coupling is learned via multi-head self-attention and is allowed to vary across samples and over time; in implementation, we use nn.MultiheadAttention without an attention mask, such that each node can attend to all others.

2) Dynamic Spatio-Temporal Stream.: This stream models inter-signal relationships between dynamically coupled physiological signals like electrodermal activity (EDA) and heart rate (HR). These signals are treated as nodes in a graph, with a GAT used to learn spatial correlations across

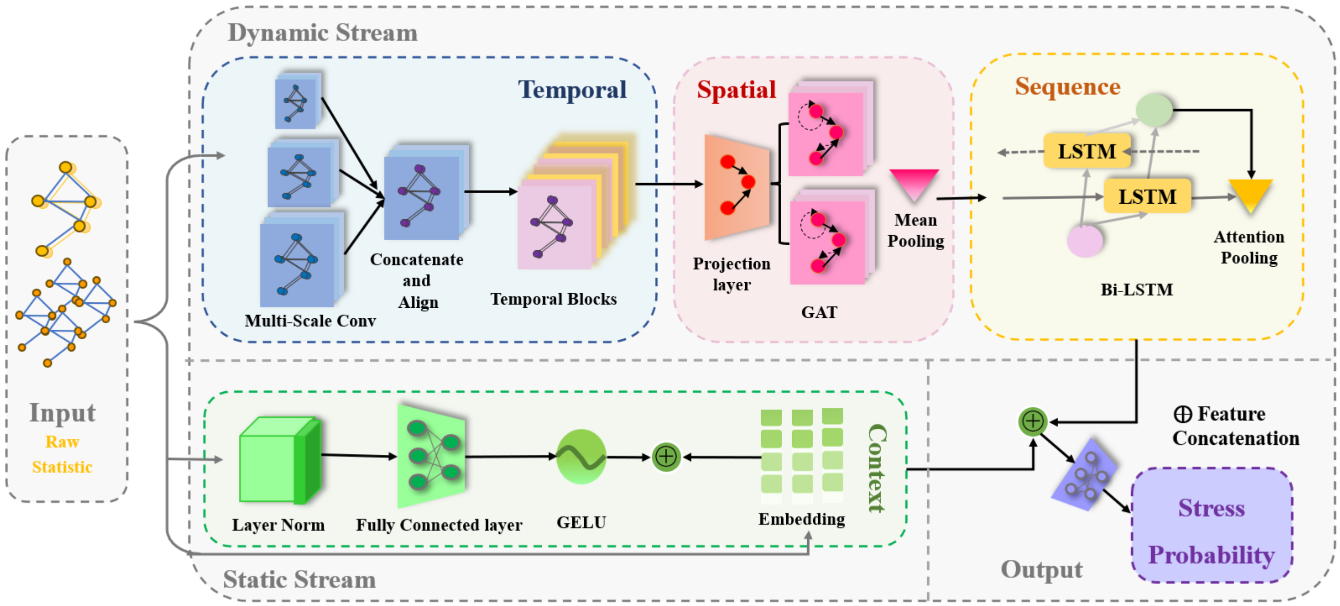

Fig. 2. Overview of the proposed BST-GATNet.

modalities This approach captures the instantaneous, non-static coupling between signals, enabling efficient modeling of their interactions over time and across modalities. The GAT computes self-attention coefficients as shown in (1), which are used to update node features via multi-head aggregation as shown in (2):

$$\alpha_{ij} = \exp\left(\text{LeakyReLU}\left(\mathbf{a}^T\left[\mathbf{W}_i\mathbf{h}_i \parallel \mathbf{W}_j\mathbf{h}_j\right]\right)\right) \quad (1)$$

where $\mathcal{N}(i)$ denotes the neighborhood of node i (fully connected over modalities, including self-loops), K is the number of attention heads, and $\sigma(\cdot)$ is the activation function. The attention weights $\alpha_{ij}^{(k)}$ are derived from (1), and the resulting node representations are updated through multi-head aggregation as defined in (2).

$$\mathbf{h}_i' = \frac{1}{K}\sum_{k=1}^{K}\sigma\left(\sum_{j\in\mathcal{N}(i)}\alpha_{ij}^{(k)}\mathbf{W}^{(k)}\mathbf{h}_j\right) \quad (2)$$

To capture the temporal dynamics of signal interactions, the spatio-temporal GAT (ST-GAT) backbone applies the GAT at each temporal step, starting with five raw physiological signals: EDA, TEMP, BVP, HR, and ACC, arranged in a tensor of size (B, 5, 1920), where B is the batch size. A multi-scale convolutional module performs early temporal downsampling, reducing the sequence length from 1920 to 61. This is followed by three residual temporal blocks using 1D convolutions and residual connections to refine temporal dependencies. The resulting features are reshaped into graphs and passed through stacked GAT layers to model modality interactions.

While ST-GAT captures instantaneous spatial relationships, it lacks memory for long-range sequential context. To address this, node representations are aggregated via mean pooling and passed through a three-layer Bi-LSTM to model bidirectional temporal dependencies [15]. An attention pooling layer aggregates the outputs, producing a fixed-length representation $\mathbf{h}_{\text{seq}}$ (B, 128) for fusion with the static feature stream.

3) Static Feature Stream.: This stream encodes non-temporal statistical descriptors that complement the dynamic stream. The 37-dimensional input vector includes five phasic skin conductance response (SCR) features derived from EDA, twelve time-domain heart rate variability (HRV) metrics derived from BVP, and basic statistics (mean, standard deviation, minimum, maximum, median, range) for TEMP, HR, and ACC. These features are normalized and passed through a fully connected layer with Gaussian error linear unit (GELU) activation, producing a 64-dimensional dense embedding $\mathbf{s}_{\text{stat}}$.

The outputs from both streams: the dynamic vector $\mathbf{h}_{\text{seq}}$ (B, 128) and the static vector $\mathbf{s}_{\text{stat}}$ (B, 64) are concatenated into a single multimodal representation. A compact multilayer perceptron (MLP) head, consisting of fully connected layers with GELU activations and dropout, produces a final logit representing the stress probability.

To ensure efficiency and real-time operation, three key design decisions were made: first, early downsampling reduces the raw sequence length from 1920 to 61, significantly cutting attention complexity from $O(1920^2)$ to $O(61^2)$ while retaining dynamics; second, compact hidden widths (e.g., $d = 64$) ensure a lightweight model with only 0.62M parameters, much smaller than typical Transformer models; finally, the GAT mechanism efficiently captures spatial dependencies by adaptively weighting signal nodes, reducing reliance on costly sequence-wise self-attention [16].

C. Model Training Strategy

In training, we formulated an extreme binary stress classification task using only samples from the lower and upper

TABLE I
PERFORMANCE COMPARISON WITH EXISTING METHODS ON AFFECTIVEROAD

Type	Method	Accuracy (%)	F1	AUC
Traditional ML	SVM (Single-task) [18]	70	0.72	–
Traditional ML	Random Forest [19]	70.4	–	–
Traditional ML	MT-MKL (T=3, RBF_L2) [18]	83	0.87	–
Deep Learning	ResNet-Conv1D [12]	71.5	0.67	–
Deep Learning	CFL (Federated) [20]	74	–	–
Deep Learning	Dense ANN-2L [18]	83.13	0.85	–
Ours	**BST-GATNet**	**88.43**	**0.88**	**0.93**

TABLE II
ABLATION BASELINES ON AFFECTIVEROAD

Model	Accuracy (%)	F1	AUC
graph neural network (GNN)	56.73	0.55	0.61
GAT	53.69	0.46	0.58
Bi-LSTM	55.33	0.53	0.61
BST-GATNet (Ours)	**88.43**	**0.88**	**0.93**

tails of each subject's stress distribution. Specifically, for each subject, windows whose stress ratings fell below the 15th percentile were labeled as low-stress, while those above the 85th percentile were labeled as high-stress; windows in the middle range were excluded to emphasize clear-cut stress states. We first randomly split the resulting dataset into 80% for training and 20% for testing. To further assess its robustness and generalization, we employed five-fold cross-validation to ensure the statistical reliability of the reported results.

A comprehensive data augmentation strategy was adopted to simulate real-world signal variability. At the signal level, augmentations included random noise injection, time warping, channel mixing, and frequency masking. At the batch level, advanced regularization techniques such as Mixup and CutMix [17] were progressively introduced during training to enhance stability. The model was optimized using the AdamW optimizer with a progressive dropout schedule and a cosine annealing learning rate policy that incorporated an initial warm-up phase to ensure smooth convergence.

IV. RESULTS AND DISCUSSION

To evaluate whether the proposed BST-GATNet can reliably support stress state detection under realistic deployment constraints, we conducted a comprehensive experimental analysis focusing on both predictive performance and practical feasibility. In addition to quantitative classification results, particular attention is given to the robustness of the model under compact representations and its ability to capture meaningful physiological interactions.

A. Performance Comparison with Existing Methods

The classification performance of the proposed BST-GATNet was evaluated on the AffectiveROAD dataset using five-fold cross-validation. The model achieves a mean accuracy of 88.43% $\pm$ 2.89, a mean F1-score of 0.88 $\pm$ 0.03, and a mean AUC of 0.93 $\pm$ 0.02, indicating consistently strong predictive performance across validation folds. These results demonstrate that the proposed method provides reliable stress state detection under the evaluated experimental setting.

TABLE III
COMPARISON OF INFERENCE LATENCIES, MODEL SIZE, AND ACCURACY FOR DIFFERENT MODELS

Model	Size (MB)	Mean (ms)	P99 (ms)	Accuracy (%)
MT-MKL (T=3, RBF L2) [18]	–	18.29	33.36	83.00
Transformer-Base (6L)	4.77	19.12	25.01	74.87
ResNet Conv1D [12]	14.71	3.84	5.11	71.50
BST-GATNet (Ours)	**2.35**	5.50	7.31	**88.43**

Compared with conventional machine learning approaches, including SVM, Random Forest, and MT-MKL, BST-GATNet achieves substantially higher classification accuracy, as summarized in Table I. These traditional methods rely primarily on handcrafted features and fixed statistical representations, which limits their ability to capture the complex temporal dynamics inherent in affective-related physiological signals. The observed performance gap highlights the advantage of end-to-end representation learning for this task.

When compared with existing deep learning–based baselines, including ResNet-Conv1D, CFL, and Dense ANN-2L, the proposed BST-GATNet consistently achieves superior average performance across the reported metrics. As shown in Table I, the proposed method attains higher mean accuracy and F1-score than all evaluated deep learning baselines under the same experimental protocol. These results indicate that BST-GATNet provides a more effective modeling framework for affective-related physiological signals within the evaluated setting.

To further investigate the source of the observed performance gains, an ablation study was conducted by evaluating individual components of the proposed architecture in isolation. As reported in Table II, standalone GNN, GAT, and Bi-LSTM models yield classification accuracies close to chance level, indicating limited discriminative capability when used independently. In contrast, the full BST-GATNet architecture achieves a substantial improvement, confirming that the superior performance arises from the effective integration of spatial physiological dependency modeling and bidirectional temporal sequence learning.

B. Real-Time Inference Performance Across Heterogeneous Platforms

To evaluate whether BST-GATNet can meet real-time inference requirements in practical deployment scenarios, we benchmark its inference latency and model footprint under unified and controlled experimental conditions, and compare it against representative baseline methods.

1) Experimental Setup and Fairness: All models are evaluated using standard implementations without deployment-specific optimizations, as this study focuses on method-level efficiency rather than engineering acceleration techniques. All experiments are conducted on the same dataset with batch

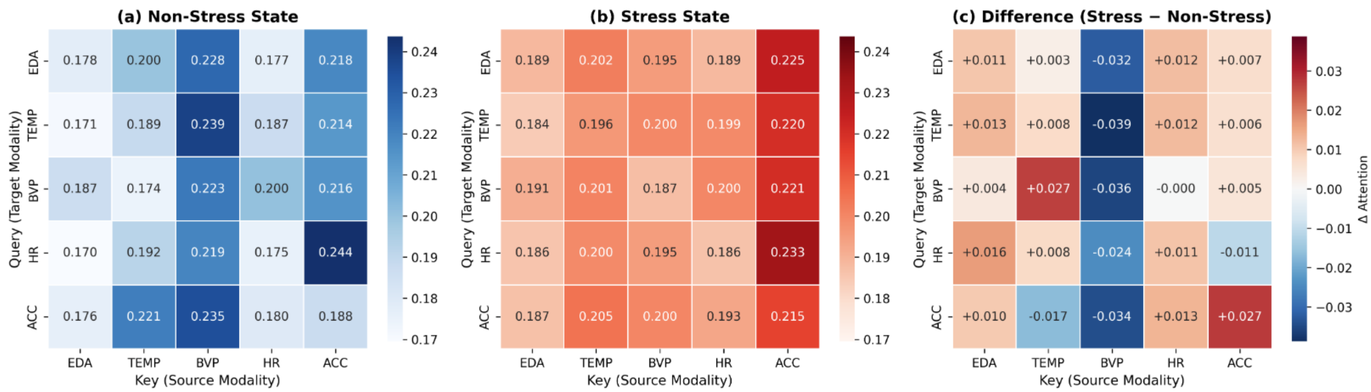

Fig. 3. Inter-modality attention patterns under non-stress and stress conditions, and the corresponding difference map (stress minus non-stress). Rows indicate the target modality being updated and columns indicate the source modality providing contextual information.

TABLE IV
CROSS-PLATFORM INFERENCE PERFORMANCE OF BST-GATNET.

Platform	Method	Mean (ms)	P99 (ms)
NVIDIA GPU	PyTorch	5.88	6.69
CPU Multi-thread (4T)	PyTorch	5.50	7.31
CPU Single-thread	PyTorch	5.32	5.95
Intel NUC i5	PyTorch	5.83	7.62
Raspberry Pi 4 (4GB)	ONNX	3.40	4.34
Raspberry Pi 5 (8GB)	ONNX	2.55	3.26
NVIDIA Jetson Nano	ONNX	4.25	5.43

size set to 1. CPU inference uses 4 threads to reflect a typical on-device multi-threaded setting. Timing covers the forward pass only, excluding data I/O and preprocessing, under PyTorch 2.4.0+cu118. We note that the original MT-MKL work does not report inference latency, indicating that runtime efficiency was not a primary design objective. Accordingly, a standardized inference procedure is adopted to ensure fair method-level comparison.

2) Accuracy–Latency–Model Size Trade-off: Table III summarizes the accuracy–latency trade-off under the controlled CPU (4-thread) benchmark. Under the same protocol, BST-GATNet achieves the highest classification accuracy (88.43%) with a mean/P99 latency of 5.50/7.31 ms and a compact model size of 2.35 MB.

In comparison, the reproduced ResNet_Conv1D baseline exhibits lower mean latency (3.84 ms) but substantially reduced accuracy (71.50%) and a significantly larger model footprint (14.71 MB), indicating a less favorable accuracy–efficiency balance for affective physiological modeling. MT-MKL shows both lower accuracy (83.00%) and markedly higher inference latency (18.29 ms). This limitation arises from its inherent inference mechanism, which requires evaluating kernel functions against all support samples, resulting in runtime complexity that scales linearly with the training set size and limiting its suitability for low-latency online inference. Transformer-based models also incur higher latency due to the quadratic complexity of self-attention with respect to sequence length, as attention is computed over the full temporal sequence. In contrast, BST-GATNet performs attention over a compact modality-node graph, combined with early temporal downsampling and localized temporal modeling, substantially reducing computational overhead.

3) Cross-Platform Inference Performance: To further assess deployment feasibility, we benchmark BST-GATNet across heterogeneous execution platforms, as reported in Table IV. Under both GPU and CPU runtimes using PyTorch, the model consistently operates at millisecond-level latency, with P99 latency below 8 ms across all tested platforms, indicating that real-time inference is achievable within common latency budgets. For resource-constrained edge devices, the model is exported to ONNX and evaluated on platforms such as Raspberry Pi 4 and NVIDIA Jetson Nano. Even under limited compute budgets, BST-GATNet maintains 3–5 ms P99 latency, demonstrating its suitability for embedded and mobile deployment scenarios.

4) Impact of Lightweight Design: Overall, BST-GATNet achieves a favorable balance between accuracy, inference latency, and model size. Its real-time performance is not the result of implementation-specific optimizations, but rather stems from a compact spatio-temporal representation and structured cross-modal attention design tailored to multimodal physiological signals. These results confirm that BST-GATNet provides a practical and deployable solution for real-time stress state detection on embedded and mobile platforms.

C. Analysis of Learned Physiological Coupling Patterns

Fig. 3 visualizes how BST-GATNet allocates attention across modalities when updating the latent representation of each signal. The heatmaps report inter-modality attention weights over EDA, TEMP, BVP, HR, and ACC, where each row corresponds to a target modality being updated and each column corresponds to a source modality providing contextual information. Across both conditions, the weights remain within a relatively narrow range, indicating that the model performs state-dependent re-weighting across modalities rather than collapsing to a single dominant channel. We therefore interpret the patterns as evidence of context-dependent information redistribution within the model, rather than as direct physiological connectivity or causal coupling.

In the non-stress condition shown in Fig. 3(a), the attention allocation is clearly structured. BVP serves as the strongest source modality for four targets, namely EDA, TEMP, BVP, and ACC, where the largest entry in each of these rows occurs in the BVP column. HR is the only target that deviates from this pattern, and it assigns its highest weight to ACC. This configuration suggests that, under baseline conditions, the model relies heavily on BVP-centered context for most updates, while motion-related information plays the primary role when refining HR-related representations.

A different allocation emerges under stress in Fig. 3(b). The overall distribution remains broadly balanced, but ACC becomes the highest-weight source modality for all targets, and HR←ACC remains among the largest entries in the matrix. This suggests that, under stress, the model places greater emphasis on ACC-derived context while still integrating information from multiple modalities. A plausible interpretation is that stress-related segments in real-world driving may coincide with movement- or posture-related variations captured by ACC.

The stress-induced changes are summarized explicitly in the difference map in Fig. 3(c). The most consistent global trend is a reduction in reliance on BVP as a source modality, as the entire BVP column exhibits negative changes across all targets. On top of this global redistribution, several localized increases are observed, including a positive shift from TEMP to BVP and a strengthened self-reliance of ACC. In addition, attention from EDA to HR increases modestly. This pattern may reflect that the model assigns greater joint relevance to these modalities in stress-related segments.

These results suggest that BST-GATNet systematically adjusts its utilization of cross-modal information across different stress states, giving rise to state-dependent attention allocation patterns within the learned multimodal representation. The observed changes in attention weights are not random fluctuations, but instead exhibit consistent global trends and coherent local redistributions, reflecting the model's ability to dynamically reconfigure multimodal contextual information under a fixed, fully connected topology. Such state-aware attention allocation provides an interpretable view of how the model integrates diverse physiological signals under varying conditions.

V. Conclusion

This paper addresses a core challenge in real-time stress monitoring from multimodal physiological data: how to robustly model cross-modal temporal dynamics and interactions under on-device computational constraints. To this end, we propose BST-GATNet, a lightweight spatio-temporal learning framework that explicitly captures dependencies among multimodal physiological signals while remaining suitable for resource-constrained edge platforms. By combining graph-based relational modeling with attention-driven temporal modeling, the proposed method achieves an effective balance between performance and computational efficiency for stress monitoring tasks, providing a technical foundation for sustainable physiological computing.

On the AffectiveROAD dataset, BST-GATNet achieves a classification accuracy of 88.43% and operates with millisecond-level latency under a CPU benchmark setting, with mean / P99 latency of 5.50 / 7.31 ms. The results further suggest that the model dynamically adjusts its cross-modal information integration across different stress states, yielding state-dependent attention allocation patterns within the learned multimodal representation.

From a biomedical and health informatics perspective, the proposed approach enables privacy-preserving, on-device stress monitoring on low-cost edge devices, reducing deployment and operational costs and facilitating scalable applications across diverse resource settings. This design aligns with the goals of sustainable and equitable healthcare technologies. Future work will evaluate the generalization of the proposed framework to larger and more diverse populations, and further investigate its long-term operational stability and energy efficiency.

VI. Acknowledgment

This work has been partially funded by Hector Stiftung.

References

[1] R. Bishop, "Intelligent vehicle r&d: A review and contrast of programs worldwide and emerging trends," in *Annales des télécommunications*, vol. 60, no. 3. Springer, 2005, pp. 228–263.

[2] A. Broggi, A. Zelinsky, Ü. Özgüner, and C. Laugier, *Intelligent Vehicles*. Cham: Springer International Publishing, 2016, pp. 1627–1656. [Online]. Available: https://doi.org/10.1007/978-3-319-32552-1_62

[3] J. C. Prinet and N. B. Sarter, "The effects of high stress on attention: A first step toward triggering attentional narrowing in controlled environments," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 59, no. 1. SAGE Publications Sage CA: Los Angeles, CA, 2015, pp. 1530–1534.

[4] A. Koesdwiady, R. Soua, F. Karray, and M. S. Kamel, "Recent trends in driver safety monitoring systems: State of the art and challenges," *IEEE transactions on vehicular technology*, vol. 66, no. 6, pp. 4550–4563, 2016.

[5] S. Begum, "Intelligent driver monitoring systems based on physiological sensor signals: A review," in *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*. IEEE, 2013, pp. 282–289.

[6] M. N. Rastgoo, B. Nakisa, A. Rakotonirainy, V. Chandran, and D. Tjondronegoro, "A critical review of proactive detection of driver stress levels based on multimodal measurements," *ACM Computing Surveys (CSUR)*, vol. 51, no. 5, pp. 1–35, 2018.

[7] W.-Y. Chung, T.-W. Chong, and B.-G. Lee, "Methods to detect and reduce driver stress: a review," *International journal of automotive technology*, vol. 20, no. 5, pp. 1051–1063, 2019.

[8] G. Giannakakis, D. Grigoriadis, K. Giannakaki, O. Simantiraki, A. Roniotis, and M. Tsiknakis, "Review on psychological stress detection using biosignals," *IEEE transactions on affective computing*, vol. 13, no. 1, pp. 440–460, 2019.

[9] R. Li and Z. Liu, "Stress detection using deep neural networks," *BMC Medical Informatics and Decision Making*, vol. 20, no. Suppl 11, p. 285, 2020.

[10] V. Singh Roha, M. E. Gaber Gendy, K. Pavan, T. M. Deserno, N. Ganapathy, and M. R. Yuce, "Enhanced driver stress prediction from multiple biosignals via cnn encoder-decoder model," in *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2024, pp. 1–4.

[11] T. D. Bui, D. T. Quang, and C. T. Long, "Driving stress detection using physiological data with machine learning," *Journal of Military Science and Technology*, no. 83, pp. 22–29, 2022.

[12] D. Jaiswal, S. Mukhopadhyay, and V. Sharma, "Tinystressnet: On-device stress assessment with wearable sensors on edge devices," in *2024 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*. IEEE, 2024, pp. 166–171.

[13] S. Cvetkovic, S. Stankovic, and S. V. Nikolic, "Bioedgenet: A compact deep residual network for stress recognition on edge devices," *Biomedical Signal Processing and Control*, vol. 102, p. 107361, 2025.

[14] N. E. Haouij, J.-M. Poggi, S. Sevestre-Ghalila, R. Ghozi, and M. Jaïdane, "Affectiveroad system and database to assess driver's attention," in *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*, 2018, pp. 800–803.

[15] H. Zhang, T. Lin, F. Ji, D. Cheng, Y. Shi, and B. H. Gwee, "Long-short-gnn: A novel graph neural network for detecting fpga ip circuits for hardware assurance," in *2025 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2025, pp. 1–5.

[16] A. Demir, T. Koike-Akino, Y. Wang, and D. Erdoğmuş, "Eeg-gat: Graph attention networks for classification of electroencephalogram (eeg) signals," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2022, pp. 30–35.

[17] S. Yun, D. Han, S. J. Oh, S. Chun, J. Choe, and Y. Yoo, "Cutmix: Regularization strategy to train strong classifiers with localizable features," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 6023–6032.

[18] D. Lopez-Martinez, N. El-Haouij, and R. Picard, "Detection of real-world driving-induced affective state using physiological signals and multi-view multi-task machine learning," in *2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*. IEEE, 2019, pp. 356–361.

[19] P. Siirtola and J. Röning, "Comparison of regression and classification models for user-independent and personal stress detection," *Sensors*, vol. 20, no. 16, p. 4402, 2020.

[20] H. Rafi, Y. Benezeth, F. Y. Song, P. Reynaud, E. Arnoux, and C. Demonceaux, "Tree-based personalized clustered federated learning: A driver stress monitoring through physiological data case study," *IEEE Internet of Things Journal*, 2024.