International Journal of Multidisciplinary Engineering in Current Research - IJMEC
Volume 10, Issue 6, June-2025, http://ijmec.com/, ISSN: 2456-4265

Real-Time Personalized Physiologically Based Stress Detection For Hazardous Operations

Neha Fatima[1], Afifa Nazneen[2], Sadia Unissa[3], Ms. Sumayya Begum[4]
[1,2,3]B.E. Student, Department of IT, Lords Institute of Engineering and Technology, Hyderabad
[4] Assistant Professor, Department of IT, Lords Institute of Engineering and Technology, Hyderabad
sumayyabegum@lords.ac.in

Abstract:

This study explores a real-time stress detection system designed for hazardous operations, aiming to enhance performance and reduce stress. Traditional machine-learning models struggle with stress detection due to individual differences and the time-series nature of physiological signals. To address this, a personalized model was developed, selecting specific features for training. The system was tested for real-time deployment using physiological data—heart rate, blood pressure, electrodermal activity, and respiration—collected from participants performing tasks with varying stress levels.

A comparison of classifiers, including Support Vector Machine, Decision Tree, Random Forest, and an Approximate Bayes (A Bayes) classifier, showed that personalized models outperform generalized ones in classifying stress levels. Results indicate that model accuracy varies with feature selection, window size, and task type, with blood pressure emerging as a key indicator. The study highlights the advantage of personalized models in stress detection and their potential for future applications.

Keywords *— Stress monitoring, AI-driven analysis, biometric sensors, immersive simulations, astronaut training.*

I. INTRODUCTION

Stress significantly impacts an individual's response in emergencies, affecting decision-making, situational awareness, and cognitive function[1]. Effective stress management is crucial to preventing performance decline and mission failure[2]. Machine learning-based stress detection offers potential solutions but faces several challenges[3], [5].

First, stress responses vary among individuals, making it difficult to generalize models[18]. Personalized stress detection may be more effective than generalized approaches. Second, physiological signals exhibit temporal dependencies that violate common machine learning assumptions, leading to biased results[10],[12]. Third, traditional machine learning algorithms approximate stress probabilities but lack a benchmark for accuracy [6],[7]. Bayes' theorem provides an optimal classification framework by estimating stress levels using empirical density distributions[7]. Machine learning models can improve reliability by incorporating multivariate kernel density estimators to account for physiological dependencies.

For real-time stress monitoring, efficient analysis of physiological time-series data is needed[14]. Adaptive training environments can dynamically adjust scenarios based on stress responses[4],[15]. Multivariate kernel density estimators can improve detection accuracy in datasets with repeated physiological measurements by minimizing uncertainty[11]. This study evaluates personalized stress detection based on objectivity, reliability, and validity. It aims to determine if stressors create distinct physiological responses for classification, assess the reliability of a time-series interval approach across various conditions, and compare supervised classifiers with an Approximate Bayes (A Bayes) classifier [7], [10].

As part of broader VR training development, this research tests a time-series interval-based model using physiological data from stress-inducing tasks[6]. It validates classification accuracy and proposes an architecture for real-time stress monitoring. Post-hoc evaluations of machine learning pipelines will enable practical implementation in adaptive training environments for improved stress detection and management[4],[14]

II. RELATED WORK

In today's digital era, sentiment analysis has become essential for understanding customer opinions and public perception. As a subfield of Natural Language Processing (NLP), it helps analyse textual data from sources like social media, product reviews, and online discussions to classify sentiments as positive, negative, or neutral. Businesses, policymakers, and researchers use sentiment analysis to enhance customer satisfaction, monitor brand reputation, gain competitive insights, and make informed decisions.

Machine learning plays a crucial role in automating sentiment analysis, reducing dependence on manual efforts. Supervised learning methods, including Naïve Bayes, Support Vector Machines (SVM), and Random Forest, classify sentiments based on labelled datasets. Deep learning approaches, such as Recurrent Neural Networks (RNNs) and Transformer-based models like BERT, capture contextual meanings for improved accuracy. Lexicon-based techniques rely on predefined word dictionaries to analyse sentiments effectively [6-8]

This study aims to explore and evaluate machine learning approaches for sentiment analysis. It focuses on comparing the effectiveness of algorithms like Naïve Bayes, SVM, Random Forest, and deep learning models. Additionally, it examines the impact of text pre-processing techniques, including tokenization, stemming, and stop-word removal, on classification accuracy. The role of feature extraction methods, such as TF-IDF, word embeddings, and deep learning representations, is also analysed. Furthermore, the study evaluates the efficiency, accuracy, and computational cost of different sentiment analysis models[6-8]

Sentiment analysis has widespread applications across industries, including e-commerce, healthcare, finance, politics, and entertainment. It enhances product recommendations, improves patient feedback systems, tracks market sentiment, assesses public opinion, and analyses audience reactions.

III. RESEARCH METHODOLOGY

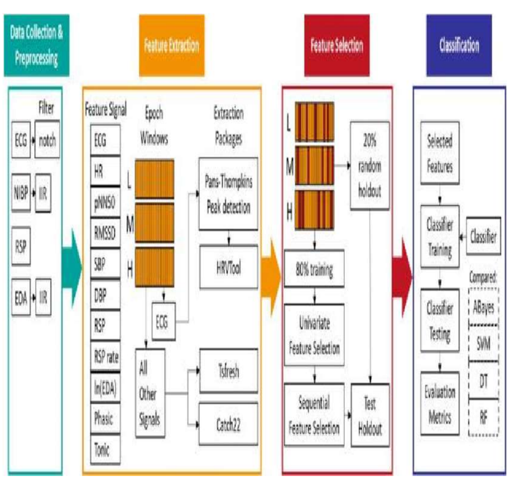

Fig.1 Stress Detection and Classification[1]

Participants:

Aggregate of 41 healthy individualities (34 males, 7 ladies) shared, with an average age of 20.9 ± 6.5 times (range 18 – 41). Actors were assigned to a VR spaceflight exigency fire task (N = 27) or a laboratory- grounded N- reverse task (N = 14). The demographic breakdown included 76 European American/ White, 12 Asian or Asian American, and 7 Hispanic or Latino.

Test Design:

The study featured two task types with three stressor situations. The VR- ISS task dissembled a spaceflight exigency fire, taking actors to detect and extinguish a fire. Stress situations were manipulated using voice adverts

(low), admonitions, bank, and flashing lights(medium), and boosted goods across all modules(high).

The N- reverse task assessed cognitive cargo by taking actors to recall the position of a multi coloured square displayed n way before. 1-reverse(low), 2- reverse(medium), and 4-reverse(high) conditions were used to control task difficulty.

Stress dimension:

Actors rated their stress using the Free Stress Scale (0 – 100) after completing trials. This measure assured stressor conditions were effective.

Procedure:

The trial lasted 120 twinkles. Actors handed informed concurrence, completed a demographic check, and passed VR navigation training. Physiological detectors were attached for birth recordings before assigning actors to VR- ISS or N-aft tasks.

VR- ISS actors entered a 20 – 30- nanosecond tutorial covering VR navigation and exigency procedures. N-reverse actors passed a brief tutorial explaining task mechanics.

Trials included low, medium, and high stress conditions, with five- nanosecond breaks between each. The Free Stress Scale was administered after the final trial.

Stress Discovery System Overview :

A machine literacy channel was developed for stress bracket using data collection, pre-processing, point birth, selection, and bracket.

Physiological data were recorded via multiple detectors, and time- series bracket was used for point birth. Dimensionality reduction meliorated features, and supervised literacy models classified stress situations.

For real- time stress discovery, the system enabled nonstop monitoring and adaptive responses. A resembling processing armature allowed real- time bracket, with a 30-alternate buffering period icing smooth operation.

A separate party group tested the real- time system in an adaptive VR stress training terrain. The system acclimated stressor intensity stoutly, perfecting stoner training. Cross validation verified trust ability, and quiescence testing assured real- time operation.

Data Collection & Pre-processing:

Four physiological signals were recorded using the Biopic , MP150 system, Electrocardiogram (ECG) (via ECG100C module), Electrodermal exertion (EDA) (indicator and middle fritters), Respiration (RSP) (Bluetooth- transmitted for mobility), Non-invasive blood pressure (NIBP) (oscillometric cutlet cuff) ECG and respiration were tried at 125 Hz. Pre-processing removed movement vestiges, power line hindrance, and electromagnetic noise. Pollutants included IIR band- pass (NIBP), electrical noise junking (ECG), and alternate- order Butterworth low-pass (EDA). G. Feature Extraction point birth captured stress- related physiological changes. Heart rate variability (HRV) features included root mean forecourt of consecutive differences (RMSSD) and chance of peak- to- peak intervals exceeding 50 ms (pNN50). Lower values indicated stress[3],[9].

Respiration rate, systolic/ diastolic blood pressure, and electrodermal exertion (EDA) factors (alcohol and phasic) were uprooted. Time- series features (e.g., absolute energy, autocorrelation, entropy) were reckoned using Ts fresh and Catch22 toolsets.

Algorithms

Decision Tree :

Decision trees are popular classifiers due to their interpretability and ease of use. They recursively resolve the dataset grounded on crucial attributes to classify instances. However, a splint knot is created, If all cases in a subset belong to the same class. else, a test trait is chosen, and branches are formed for each outgrowth. This process repeats for each subset. While decision trees can overfit complex data, ways like pruning help ameliorate delicacy and conception.

Naive Bayes:

Naive Bayes is a probabilistic classifier grounded on Bayes' theorem. It assumes point independence and is effective for large, high- dimensional datasets. Generally used in spam discovery, sentiment analysis, and document bracket, it offers fast training and reasonable delicacy. Though its independence supposition is unrealistic, it still performs well. Tools like Weka, Tanagra, Orange, and RapidMiner are used to compare Naive Bayes with styles like logistic retrogression and SVM for better perceptivity[7].

Random Forest :

Random Forest is an ensemble system that combines multiple decision trees to ameliorate vaticinator delicacy and reduce overfitting. For bracket, it uses maturity voting, and for retrogression, it pars prognostications. Introduced by Tin Kam Ho and meliorated by Bierman and Cutler, Random timbers are known for their robustness across diligence. Though generally effective, grade boosting can outperform it in certain cases[7]

Support Vector Machine (SVM):

SVM is a discrimination classifier that aims to find the optimal boundary (hyperplane) separating classes in a point space. Unlike generative models, SVM focuses only on class boundaries and is effective in high- dimensional spaces. It requires lower data and calculation, making it suitable for tasks where only bracket(not full probability distributions) is demanded.

Point Selection

A cold-blooded approach named applicable features. Univariate point selection (UFS) ranked features using ANOVA F- values, while successional point selection (SFS) iteratively meliorated them.

To help data leakage, point selection was applied only to training data, validated using 20 unseen test data.

Generalized Approach & Data Analysis Summary

A generalized model was compared with a substantiated model using a leave- one- subject-eschewal confirmation approach. Test data were formalized using training statistics.

The A Bayes classifier was compared with Support Vector Machine (SVM), Decision Tree, and Random Forest models in MATLAB. Performance was assessed using10foldcross-validation and holdout confirmation[7]

Metrics included delicacy, perfection, recall, F1-score, and particularity. Statistical analyses used

RM- ANOVA with Bonferroni correction, and Cohen's d measured effect sizes.

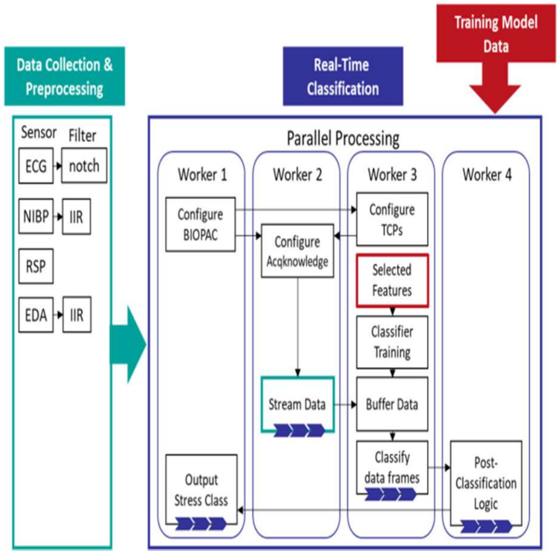

Fig.2 Stress detection implemented as a real-time[3].

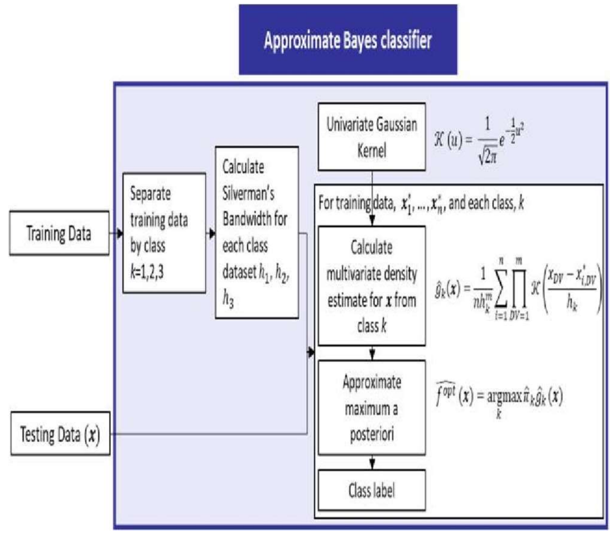

Fig.3 The approximate Bayes for stress level classification.

IV. RESULTS AND DISCUSSION

Subjective Stress Manipulation

Stress levels significantly impacted subjective stress ratings in both VR-ISS ($F(2,90) = 102$, $p < .001$, $d = 3.02$) and N-back tasks ($F(2,24) = 47.5$, $p < .001$, $d = 3.98$). Pairwise comparisons confirmed significant differences across stress levels ($p < .001$). In the N-back task, stress was highest in 4-Back compared to 1-Back ($p < .001$) and 2-Back ($p < .001$), with 2-Back significantly higher than 1Back ($p = .018$).

Machine Learning Results

Physiological data from both tasks were analysed using Sequential Feature Selection (SFS). The likelihood ratio imbalance degree (LRID) indicated higher class imbalance in VR-ISS (46%) compared to N-back (9.5%). F1-score was used as the primary performance metric, consistent across different epoch window sizes (10–40 seconds).

TABLE 1. Details of the multiclass datasets, M (± SD).

Window	N-Back		VR-ISS	
	Mean	**STE**	**Mean**	**STE**
10 sec	4.6	0.2	4.6	0.1
20 sec	5.0	0.2	4.6	0.1
30 sec	4.8	0.2	4.7	0.1
40 sec	4.4	0.2	4.2	0.2

TABLE 2. Number of features selected by SFS for each task.

Task	Number of Observations (i.e., 10-sec Windows)	Class Observations	Imbalance Ratio	LRID
VR-ISS	76.3 (±41.6)	L: 25.8 (±17.8) M: 25.5 (±19.3) H: 25.0 (±16.4)	2.37 (±2.39)	11 (±17.4)
N-back	62.9 (±17.0)	L: 20.4 (±7.26) M: 20.5 (±6.0) H: 21.9 (±5.23)	1.52 (±1.71)	1.65 (±5.08)

Classifier Performance and Validation

The highest F1-score for VR-ISS was 94% at 30 seconds (10-Fold) and 79% at 40 seconds (holdout). For N-back, it was 96% at 40 seconds (10-Fold) and 81% at 40 seconds (holdout). Random Forest performed best in N-back (98% at 30 seconds), while A Bayes achieved the highest F1-score (94% at 30 seconds) for VR-ISS [14-15]

E. Personalized vs. Generalized Approach

In leave-one-subject-out (LOSO) validation for VR-ISS, the best accuracy (62%) was achieved with Random Forest at a 30-second window [14-15]

Following are the accuracies of following machine learning models in both tabular form and bar chart[6-8].

Table 3: Accuracies of Models

MODEL	ACCURACY
Decision Tree	72
Navie Bayes	91
Random Forest	85
SVM	85

The personalized stress detection system selected time-series features per individual, trained models, and classified stress levels in real time. Stress levels induced by VR-ISS and N back tasks were validated using questionnaires and physiological data. Classifier reliability was tested across tasks, window sizes, and validation methods, with better performance seen in the simpler N-back task and shorter windows (10–20 sec). F1-scores addressed data imbalance, showing 82–94% (VRISS) and 79–96% (N-back) in cross-validation. The A Bayes classifier, using kernel density estimates, outperformed or matched traditional models, achieving 94% F1 at a 30-sec window. Personalized models outperformed generalized ones, making them effective for real-time stress detection. A Bayes, Random Forest, and SVM yielded higher F1-scores for VR-ISS, with A Bayes achieving 94% (cross-validation) and SVM 84% (holdout). For N-back, Random Forest had the best F1-score (98% cross-validation), while Decision Tree topped holdout (84%). Classifier performance varied by task due to differences in task complexity and data characteristics. N-back, being a controlled cognitive task, consistently showed higher F1-scores than the more complex VR-ISS. For example, Random Forest achieved 98% (N-back) vs. 90% (VR-ISS) with the same settings. ABayes showed similar trends with slightly better results on N-back across both validation methods. This suggests the system is robust across task types. Personalized models outperformed generalized ones (82% vs. 62% on VR-ISS), highlighting their ability to adapt to individual physiological variations. Generalized models struggled with feature variability, but larger and more uniform datasets could improve their performance. Compared to other studies, this system achieved some of the highest reported multi-class stress detection scores, especially with N-back. However, many studies lacked F1score reporting and used generalized models, making direct comparisons difficult. The study's use of the SFS wrapper for feature selection further boosted accuracy, emphasizing the importance of model personalization and evaluation choices. The stress detection is more effective when using personalized models rather than generalized ones. Key features like SBP, DBP, and EDA were commonly selected, while HRV features were often excluded due to individual variability. Longer time windows shifted the focus to slower signals, and shorter windows favored fast-changing ones. This shows that feature relevance depends on both the time window and task type. Wrapper-based feature selection proved essential for tailoring models to individuals. Future systems should consider additional inputs like behavior, speech, and demographics to improve accuracy and adaptability. The stress detection approach depends heavily on sensor quality, the type of stress, and task context. Since stress responses vary by stressor and timescale, models trained on one task may not generalize well to another. Using real-time processing and wrapper-based feature selection helped reduce signal noise and adapt to individual differences. However, real-time systems face issues like model degradation over time, lack of retraining, and physiological adaptation. Personalized systems may need occasional recalibration using stress biomarkers like cortisol. Other challenges include age-related physiological changes, overfitting due to feature correlation, and the assumption of Gaussian data in the A Bayes classifier. Future systems should handle non-Gaussian features better and evaluate how classifiers and time window sizes affect feature selection.

V.CONCLUSION

This research addresses the challenges posed by individual variations in stress responses and the time-series nature of physiological signals. By implementing a personalized timeseries interval approach, we evaluated the objectivity, reliability, and validity of a real-time stress detection system.

The distinction between simple and complex tasks successfully established varying stress levels, making them suitable for machine learning ground truth.

Analysis of different window sizes provided insights into the most effective sensors and features for varying time intervals. The results demonstrated that a personalized model outperformed a generalized one. Additionally, the study assessed the impact of indirect approximations by supervised machine learning classifiers, comparing them against the benchmark optimal classifier, A Bayes. It was observed that indirect approximations could influence classifier performance, ranging from a decrease of 11% to an increase of 14% relative to A Bayes.

These findings indicate that a personalized system offers promising performance compared to previous multi-class stress detection studies. Researchers must carefully select HMIs, sensors, and features, as they may not fully account for inter- and intra-individual differences in stress physiology. Future work will focus on enhancing personalized stress detection systems by incorporating methods that adapt to temporal variations in individual stress responses and physiological signals.

VI. REFERENCES

[1] J. E. Driskell, E. Salas, J. H. Johnston, and T. N. Wollert, Stress Exposure Training: An Event-Based Approach (Performance Under Stress). London, U.K.: Ashgate, 2008, pp. 271–286.

[2] I. Barshi and D. L. Dempsey, ‘‘Risk of performance errors due to training deficiencies: Evidence report,’’ Nat. Aeronaut. Space Admin. (NASA), NASA Johnson Space Center, Houston, TX, USA, Tech. Rep., JSCCN- 35755, 2016.

[3] M. Gjoreski, M. Luštrek, M. Gams, and H. Gjoreski, ‘‘Monitoring stress with a wrist device using context,’’ J. Biomed. Inform., vol. 73, pp. 159–170, Sep. 2017, doi: 10.1016/j.jbi.2017.08.006.

[4] M. Zahabi and A. M. A. Razak, ‘‘ Adaptive virtual reality- grounded training A methodical literature review and frame,’’ Virtual Reality, vol. 24, no. 4, pp. 725 – 752, Dec. 2020, doi 10.1007/ s10055- 020- 00434-w.

[5] Y. S. Can, B. Arnrich, and C. Ersoy, ‘‘Stress detection in daily life scenarios using smart phones and wearable sensors: A survey,’’ J. Biomed. Informat., vol. 92, Apr. 2019, Art. no.103139, doi:10.1016/j.jbi.2019.103139.

[6] Akmandor, A. O., & Jha, N. K.(2017)." Keep the stress down with Soda pop A stress discovery and relief system." IEEE Deals onMulti-Scale Computing Systems, 3(4), 269 – 282.

[7] S. Tong and D. Koller, ‘‘Bayes optimal hyperplanes? Maximal margin hyperplanes,’’ in Proc. IJCAI, 1999, pp. 1–5.

[8] I. Rish, ‘‘An empirical study of the naive Bayes classifier,’’ in Proc. IJCAI Workshop Empirical Methods Artif. Intell., 2001, vol. 3, no. 22, pp. 41–46.

[9] F. Shaffer, R. McCraty, and C. L. Zerr, ‘‘ A healthy heart is n't a metronome An integrative review of the heart’s deconstruction and heart rate variability,’’ borders Psychol., vol. 5, p. 1040, Sep. 2014, doi 10.3389/ fpsyg.2014.01040.

[10] B. Kim, Y.- S. Jeong, and M. K. Jeong, ‘‘ New multivariate kernel viscosity estimator for uncertain data bracket,’’ Ann. Oper. Res., vol. 303, nos. 1 – 2, pp. 413 – 431, Aug. 2021.

[11] E. Smets, W. De Raedt, and C. Van Hoof, ‘‘ Into the wild The challenges of physiological stress discovery in laboratory and itinerant settings,’’ IEEE J. Biomed. Health Informat., vol. 23, no. 2, pp. 463 – 473, Mar. 2019, doi 10.1109/ JBHI.2018.2883751.

[12] D. Jones and S. Dechmerowski, ‘‘Measuring stress in an augmented training environment: Approaches and applications,’’ in Foundations of Augmented Cognition: Neuroergonomics and Operational Neuroscience. Berlin, Germany: Springer, 9744, pp. 23–33, 2016, doi: 10.1007/978-3- 319-39952_3.

[13] T. T. Finseth, ‘‘Adaptive virtual reality stress training for spaceflight emergency procedures,’’ Ph.D. dissertation, Aerosp. Eng., Iowa State Univ., 2021.

[14] T. Finseth, M. C. Dorneich, N. Keren,W. D. Franke, S. Vardeman, J. Segal, A. Deick, E. Cavanah, and K. Thompson, ‘‘The effectiveness of adaptive training for stress inoculation in a simulated astronaut task,’’ in Proc. Human Factors Ergonom. Soc. Annu. Meeting, Baltimore, MD, USA, 2021, pp. 1541– 1545.

[15] H. F. Posada-Quintero and K. H. Chon, ‘‘Innovations in electrodermal activity data collection and signal processing: A systematic review,’’ Sensors, vol. 20, no. 2, p. 479, Jan. 2020, doi: 10.3390/s20020479.

[16] B. M. Kudielka, D. H. Hellhammer, and S.Wust, ‘‘ Why do we respond so else?

Reviewing determinants of mortal salivary cortisol responses to challenge,'' Psychoneuroendocrinology, vol. 34, no. 1, pp. 2 – 18, Jan. 2009, doi 10.1016/ j.psyneuen.2008.10.004.

[17] B. L. Smith, A. Kay, and E. A. Shirtcliff, ''Evoking stress reactivity in virtual reality: A systematic review and meta-analysis,'' Neuro sci. Biobehavioral Rev., vol. 138, Jul. 2022, Art. no. 104709, doi: 10.1016/j.neubiorev.2022.104709.

[18] S. L. Bowers, S. D. Bilbo, F. S. Dhabhar, and R. J. Nelson, ''Stress or specific alterations in corticosterone and immune responses in mice,'' Brain, Behave., Immunity, vol. 22, no. 1, pp. 105–113, Jan. 2008, doi: 10.1016/j.bbi.2007.07.012.