Test Score Prediction Using Physiological Signals

Biomedical Signal Analysis Final Project Report, BME 772

Alex Dunn, 500880041
Biomedical Engineering
Toronto Metropolitan University
Toronto, Canada
a1dunn@ryerson.ca

Cassidy Rose, 500899027
Biomedical Engineering
Toronto Metropolitan University
Toronto, Canada
cassidy.rose@ryerson.ca

Daniel Wong, 500886434
Biomedical Engineering
Toronto Metropolitan University
Toronto, Canada
daniel.a.wong@ryerson.ca

***Abstract*—This investigation aimed to predict student examination performance using skin surface temperature, electrodermal activity (EDA), and blood volume pulse (BVP) signals recorded during test writing. Signal processing and feature extraction methodologies were implemented based on well-established techniques in the literature for studies that similarly quantified stress levels and predicted test scores. The features extracted for analysis include maximum temperature, temperature ratio, EDA mean, EDA standard deviation, EDA maximum position in time, average heart rate, and beat-to-beat interval standard deviation. The analyzed signals were recorded from 10 students during three separate examinations (n=30). A linear discriminant analysis model in MATLAB used three features to determine whether students' individual scores per examination were above or below average. Using the R-R interval standard deviation, mean heart rate, and temperature ratio features, the resulting machine learning model successfully yielded a 70% prediction accuracy, which is consistent with results obtained in the literature using other machine learning models.**

I. Introduction

Exam writing is an integral part of student evaluation in academic settings, but large portions of the student population experience adverse mental health effects and struggle to cope with stress from writing such tests [1]. According to the 2019 National College Health Assessment, 24% of students reported receiving an anxiety disorder diagnosis in the preceding year, 45.6% of students reported experiencing above-average stress levels, and 68.9% reported feeling anxious [2]. Stress can have a significant impact on psychological well-being and performance in academic and professional settings [3], [4], [5], [6]. In addition to the harmful effects of stress on mental and physical health, excessive stress has long-term financial consequences due to increases in healthcare costs, absenteeism, job loss, and reduced profits [6], [4]. While research suggests that an optimal level of stress can enhance productivity, chronic, prolonged stress can lead to the development of moderate to severe psychological disorders and physical health problems, such as cardiovascular disease [7], [8], [9]. The increase in the prevalence of anxiety disorders during the COVID-19 pandemic highlights the importance of understanding the relationship between stress and performance and presents a unique opportunity to leverage wearable technology as a tool for stress quantification and management. The following investigation, therefore, aims to quantify stress levels from various physiological signals and use extracted features to predict examination performance. A dataset titled "A Wearable Exam Stress Dataset for Predicting Cognitive Performance in Real-World Settings" was selected and obtained from Physionet for analysis [10], [11], [4]. This dataset contains heart rate, blood volume pulse (BVP) obtained from photoplethysmography (PPG), skin surface temperature, accelerometer, electrodermal activity (EDA), and inter-beat interval data recorded using Empatica E4 wristbands during two midterms and one final exam for 10 students [10]. It was collected by Amin *et al.* with the goal of enhancing understanding of how stress levels influence academic performance during examinations [10]. In this investigation, electrodermal activity, blood volume pulse, and skin surface temperature signals will be analyzed and used to predict whether a student's academic performance was high ($\geq 76\%$) or low ($< 76\%$) during a given examination.

EDA signals provide information on how the electrical conductance of skin changes over time [12], [13]. Sweating enhances the conductivity of the skin and can be evoked by stress and emotions [12], [13], [14]. Due to the innervation of sweat glands by sympathetic nerves, skin conductance can be used as a measure of sympathetic nervous system activation and is believed to represent sudomotor activity [14]. The EDA signal comprises a slow-varying baseline tonic component, and high-frequency phasic component [12]. In this investigation, the tonic component of EDA signals will be analyzed, as it best characterizes long-term fluctuations in the skin conductance level (SLC), such as those that occur during a multi-hour examination period [15]. The EDA signal is prone to artifacts caused by body motion, loss of electrode contact, and pressure or stretching of the electrodes, which must be taken into account during signal processing [12], [16].

BVP is an additional physiological signal that can quantify stress using features such as heart rate and blood pressure [17]. BVP signals are obtained using PPG, which measures the transmission or reflection of light illuminating peripheral body tissue [17]. Changes in blood volume impact the absorption and transmission of light and, thus, the amplitude of the signal detected by a photosensitive diode sensor [17]. BVP can therefore be used to quantify changes in the state and activity of the cardiovascular system and corresponding stress levels. Pre-processing of the BVP signal must address the presence of low-frequency baseline wander and motion artifacts, with frequencies of approximately 0.004-1.6 Hz and 0.1 Hz, respectively [17]. The PPG signal can be filtered to the range of 0.5 to 5 Hz (30 to 300 bpm) to isolate the frequency component caused by periodic heartbeats [17], [18].

Skin temperature analysis is based on the variance of blood circulation to the surface of the skin [19]. Skin temperature signals can be recorded using an optical thermometer that measures the temperature o body parts such as the face, wrist, or fingers [20]. Skin temperature is typically within the range of 33.5° and 36.9° C and can vary based on external factors, and disease [21].

Artificial Intelligence (AI) and machine learning (ML) are important areas of research within the last decade that look to replicate and augment human intelligence for various automated and decision-aiding tasks [22]. These powerful tools can provide tremendous benefits within a healthcare setting, especially with the rise of pandemics, lifestyle-related diseases, and an exploding world population [22]. AI and ML can provide a means to reduce healthcare costs, create more effective and personalized treatments and provide a better overall quality of patient care. This project uses ML models to help predict test scores using various biomedical signals.

II. Methods

A. Pre-Processing

Signal pre-processing was completed by loading all data files into a cell array in MATLAB, trimming the signals to the periods corresponding to exam writing, and applying frequency-domain and temporal filters. Each signal was stored in a cell array with columns corresponding to the signal type and rows corresponding to the student number and exam type. Rows 1-10, 11-20, and 21-30 contained the signals recorded for students 1 to 10 during the first midterm, second midterm, and final exam, respectively.

During data acquisition, students were provided with Empatica E4 wristbands, which they put on shortly before the exam [10]. The exams began at the same time for every student [10]. The signals were therefore trimmed such that they spanned the same duration of time per student and per exam, using manually-selected indices from graphs of the original signals. The indices were selected based on the temperature signal, using the end of the initial ramp-up and the beginning of the ramp-down sections as the starting and ending points for segmentation. For BVP and EDA signals, it was verified that the selected indices were within the sections of the signals with amplitudes above the baseline noise. Start and end indices were adjusted to ensure that sample points corresponding to baseline noise before or after exam writing were removed from the signal.

1) Blood Volume Pulse Signal Pre-Processing: Pre-processing the BVP signals involved bandpass filtering to eliminate high and low-frequency noise outside the normal frequency range for PPG signals (0.5 to 5 Hz) [18]. A comprehensive comparative study by Liang *et al.* evaluated filter performance for BVP signal processing of nine filter types and five orders per filter type [23]. The study found that a 4th-order, low-pass Chebychev II filter is optimal for BVP signal pre-processing [23]. The Chebychev II filter provides advantages over other classical analog filters, such as the Butterworth, Chebychev I, and Elliptic, because it has a flat passband and sharper frequency cutoff in the transition zone [23]. This prevents adverse effects on signal morphology for frequency components in the pass band of interest and suppresses frequencies bordering the passband more effectively [23]. In comparison, Chebychev I and Elliptic filters have ripples in the pass band, and the Butterworth filter has flat pass band behaviour but a less sharp frequency transition zone [23]. Therefore, the BVP signals were pre-processed using a 4th-order low-pass Chebychev II filter with a cutoff frequency of 5 Hz. A high-pass filter with a cutoff frequency of 0.5 Hz was designed using the bilinear z-transformation method with unit gain using the equation below:

$$H(z) = \frac{0.894168(1 - z^{-1})}{1 - 0.78834z^{-1}} \tag{1}$$

2) Electrodermal Activity Signal Pre-Processing: The two primary time-domain metrics of the electrodermal activity signal that are correlated with sympathetic nervous system activation are skin conductance responses (SCRs) and skin conductance level (SCL) [13], [12]. Metrics quantifying SCL were extracted from the slow-varying tonic component of the EDA signal. The tonic component of the EDA signal was isolated using a well-established procedure from Posada-Quintero *et al.*, in which a finite impulse response (FIR) low-pass filter is applied to the raw signal [13], [12]. This low-pass filter had a corner frequency of 0.0004 Hz and an order of 10 [13], [12]. A median filter was applied to remove motion artifacts. A median filter was appropriate for this application because it removes outlying transient spike artifacts in signals while preserving the original signal shape [24], [25], [26].

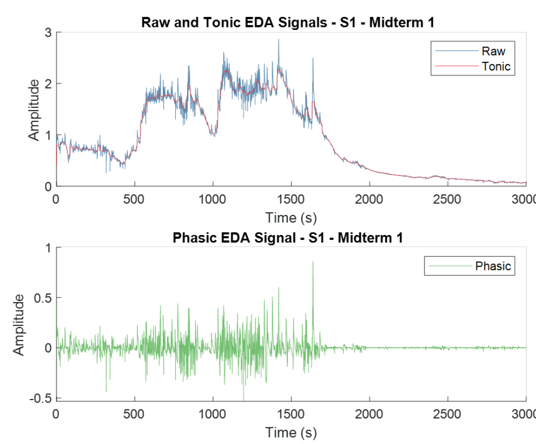

Fig. 1. Raw, tonic, and phasic EDA signals

3) Skin Surface Temperature Pre-processing: The frequency content of skin surface temperature signals is not well defined in the literature. Therefore, a time-domain moving averaging filter was used for noise removal, assuming that temperature change processes are ergodic. The size of the window was determined empirically by increasing the size of the window until rapid fluctuations were visibly removed without causing notable distortion of the signal's morphology (see the example provided in Fig. 1).

B. Feature Extraction

Test writing is a stressful experience for most students, but the stress experienced throughout test writing is not constant but rather fluctuates based on the current question they are solving, how long they have been writing the test, and most importantly, what grade the student thinks they will get on the test [4]. To determine good features for the prediction of test scores, features need to be found that quantify the emotion of the student. It is important to consider that everyone will experience different levels of stress during exam writing, therefore features needs to be compared to a baseline or normalized to some extent. A summary of the extracted features and their correlation coefficients to the students grades are outlined in Table I.

1) Skin Surface Temperature Features: Skin temperature features are based on the variance of blood circulation to the surface of the skin [19]. When the body is in a state of stress or heightened emotion, blood flow is reduced to the surface of the skin through vessel contraction [19]. In a relaxed state, the constricting musculature is relaxed to allow more blood to the surface of the skin [19]. Fluctuations in the skin surface temperature signal therefore can be used to indicate changes in emotion while writing a test. For the purposes of feature extraction, the signal was broken into 3 phases, an initial warming phase (IWP), middle writing phase (MWP), and an ending test submission phase (TSP). The IWP is the period at the start of test writing where the temperature sensor is still adjusting to the heat of the skin before reaching its true temperature. The IWP consists of the first 30-60 minutes of the signal and is irrelevant for this study. The MWP of the temperature signal can be thought of as the students baseline temperature while writing the test. In the middle section, there is variation due to the changes in emotion as certain question are found challenging or are resolved, but overall this is typically the time with the lowest stress [4]. The TSP is a region of interest since it relates to how the student was feeling in the final moments of test work and completion. These phases in the temperature signal can be seen in Fig. 2. The features chosen for the temperature signal are temperature mean for, temperature max and a temperature ratio between the middle and end values. The temperature ratio quantifies how the student was feeling at the end of the test by calculating the maximum temperature value within the last 10% of the signal. To normalize this value to individuals baseline test-temperature, the minimum temperature for the center 10% of the signal was calculated and used as the numerator in the ratio. The minimum value was used to help estimate the baseline test-temperature since this won't be biased by peaks caused by individual questions. The mean and maximum values of the temperature signal were also calculated.

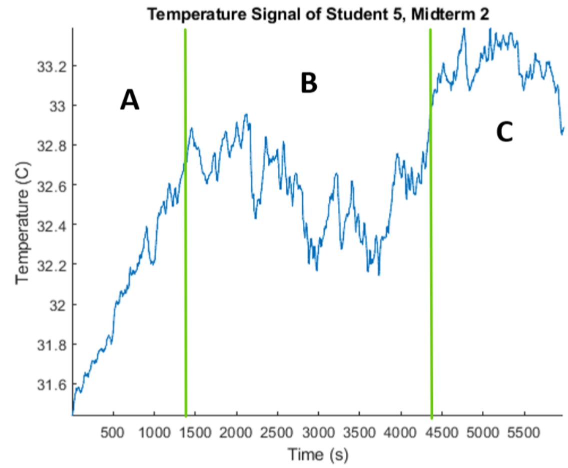

Fig. 2. Three regions within the temperature signal: A:IWP, B:MWP, C:TSP

2) Electrodermal Activity Features: Electrodermal activity is an important feature for stress indication since sweat glands are exclusively innervated by the sympathetic nervous system, allowing it to be unperturbed by other physiological processes [14], [27]. Similar to temperature, the EDA signal can be broken into phases with EDA peaks representing different stressors during test writing. EDA peaks near the start of the test represent the initial stress felt upon opening the exam package, and possibly a realization that the individual did not study enough given the magnitude of question [4]. An EDA peak near the end of test writing indicates stress that may be caused by a lack of time to complete questions or the realization that a student may receive a bad mark on the test [4]. Both starting and final EDA stress response are visualized in Fig. 3.

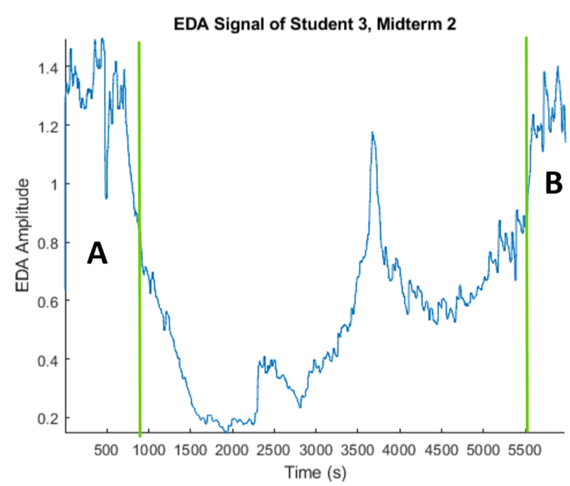

Fig. 3. Labelled sections of interest for the EDA signal: A:Initial stress response, B:End stress response

Similar to temperature, the middle of the EDA signal can be assumed as the test writing baseline for comparing to [4]. To quantify when the maximum stress occur ed, a feature named 'When Max' is created to quantify where the maximum EDA reading was observed as a fraction of total test time, i.e a value of 0.95 for a maximum reading near the time of submission or a value of 0.10 for a maximum value during the initial exam question reading. Furthermore, the mean and standard deviation of the signal were calculated as additional features.

3) Blood Volume Pulse Features: Heart activity is a complex process with many factors determining the hearts behaviour, the factor relevant to this project is the link between physiological stress and the heart [28]. Heart rate (HR) and heart rate variability (HRV) are both important features to indicate stress that can be extracted from a blood volume pulse (BVP) signal [28]. To extract features from the BVP signal, the signal was first normalized to a magnitude of one and then peak detection was used on the processed signal. To prevent artifacts in the signal from being registered during peak detection, a minimum peak-to-peak distance of 0.3s was used and a minimum value of -0.05. A minimum peak to peak distance is used to limit false detection of peaks between the beat interval and a 0.3 seconds constraint was chosen since there will be no beats at an interval higher than 200bpm. An example of the BVP signal with the peaks detected is shown in Fig. 4. To estimate HRV, the standard deviation of the detected peak locations was calculated for each signal.

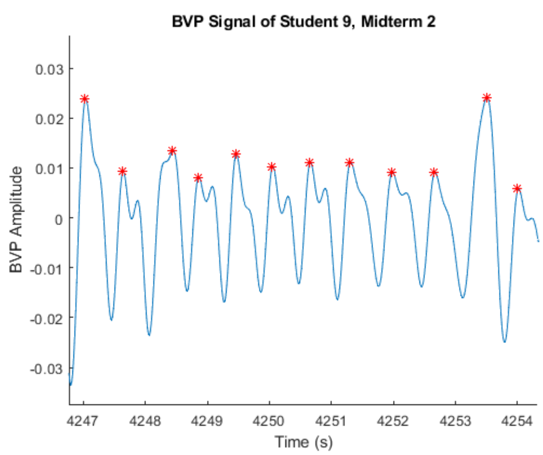

Fig. 4. Section of BVP signal with red asterisks showing peak detection

TABLE I
SUMMARY OF CORRELATION COEFFICIENTS FOR EXTRACTED FEATURES

Feature	Correlation Coefficient
Temp. Mean	0.005
Temp. Max	-0.053
Temp. Ratio	0.473
EDA Mean	0.211
EDA STD	0.234
EDA When Max	-0.316
Beat-Beat Interval STD	-0.244
BPM Average	0.634

C. Machine Learning

The decision of which machine learning model to use to analyze the data set took into account various factors which include performance, model complexity, data size and dimensionality of the samples being analyzed. All major classifiers from the MATLAB 2022b classification library with default settings were evaluated for performance accuracy. The model chosen to examine and predict the test scores of the students was linear discriminant analysis (LDA). LDA is a supervised linear machine learning model that makes use of labeled observations to construct a classification that will separate the data from each other [22]. The goal of LDA is to perform dimension reduction while also preserving class discriminatory information [22].

LDA became a clear choice for this investigation as it is advantageous in categorizing problems with multi-group classifications when the data follows a Gaussian distribution. Analyzing the histograms for the features selected indicated a set of features that exhibited a normal distribution seen in Fig. 5. These features included R-R interval standard deviation (RR_STD), mean heart rate (Mean BPM), and temperature ratio. To label the data, a high or low grade was assigned to each feature vector to indicate whether a student scored above or below 76% on their tests and shows if they are above or below the class average of all tests taken.

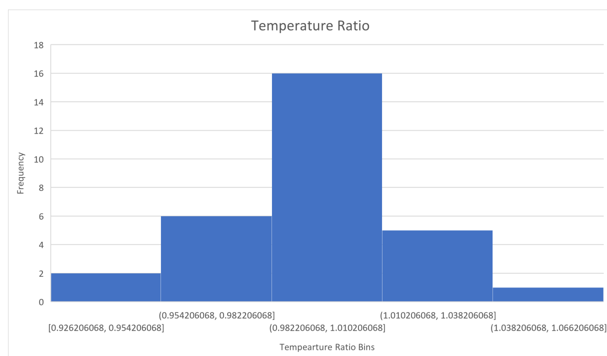

Fig. 5. Example histogram of the distribution of the Temperature ratio feature extracted

The choice of the number of features was also carefully chosen as to prevent inaccuracies in the prediction model. The leave-one-out cross validation method is a re-sampling technique used in machine learning models to prevent overfitting of the data and to evaluate the robustness of a ML model on a data pool with limited samples. Much of the literature on the number of folds suggest using 5 or 10 fold cross validation to provide a balance between protection against over fitting without compromising on computational power [29]. However, due to the small number of samples from the data to analyze, leave-one-out cross validation method was the desired choice to split the data as the computational cost was low (approx. 230obs/second) while also yielding a higher accuracy and robustness in the model prediction. While the data presented did analyze 10 students each of which took 3 tests each (midterm 1, midterm 2 and a final) a combination of the data from all three exams was not classified together as each exam has separate stimuli for each student and their performance might have been affected by different factors that is unpredictable and might provide poor and inaccurate results if correlated together. In addition to using cross validation, the one in ten rule was also employed as a method to estimate the number of predictor parameters to be used on the data. This prevents the curse of dimensionality and lowers the risk of over-fitting.

To evaluate the performance of the data set a confusion matrix was generated to understand how the selected classifier performed. After training the classification model using the data, a matrix was created to show the true class values compared to the predicted class values. The diagonal cells show where the true class and predicted class match. From the decision matrix, various performance metrics can be extracted to further evaluate the model. In this analysis classification accuracy, sensitivity, specificity, precision and f1 score were evaluated using the following equations:

$$classificationaccuracy = \frac{TN+TP}{TN+FP+FN+TP} \quad (2)$$

$$Sensitivity = \frac{TP}{FN+TP} \quad (3)$$

$$Specificity = \frac{TN}{FP+TN} \quad (4)$$

$$Precision = \frac{TP}{FP+TP} \quad (5)$$

$$F1score = \frac{2TP}{2TP+FP+FN} \quad (6)$$

In addition, a receiver operating characteristic (ROC) curve was generated and the area under the curve (AUC) was evaluated to determine overall accuracy and performance of the model.

III. Results

Using MATLAB's classification learner application, an LDA classification model was used to train and test the data with subsets of features extracted from BVP, EDA, and temperature signals. To avoid the curse of dimensionality [30], the 'one in ten rule' [31] was chosen and from the 30 samples of data, 3 features were used. Fig. 6 shows the resulting confusion matrix generated from MATLAB and Fig. 7 shows the subsequent ROC curve for the labels assigned.

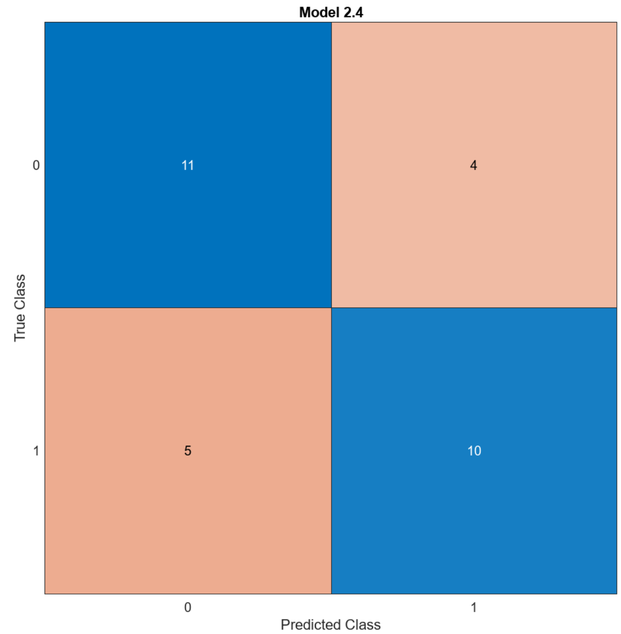

Fig. 6. Confusion matrix generated from the LDA classification model in MATLAB. '1' indicating test scores above 76%. '0' indicating test scores below 76%

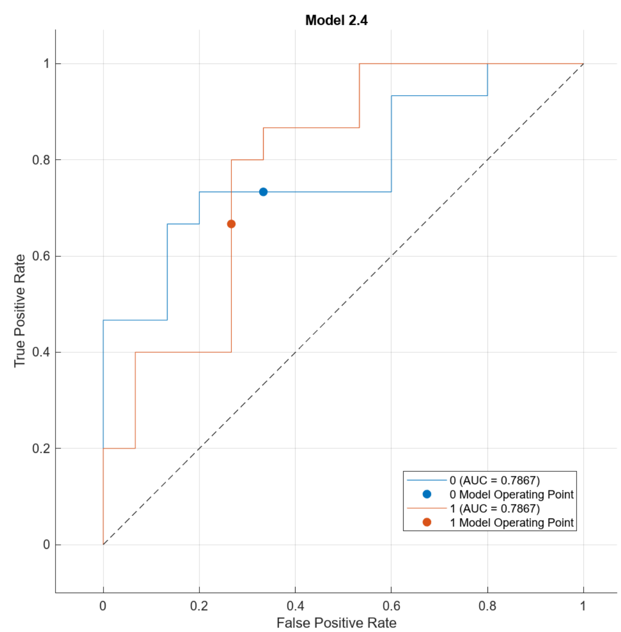

Fig. 7. ROC curve from the LDA classification model in MATLAB for both class labels. '1' label indicating scores above the 76% mean and '0' label representing the scores below the 76% mean

To better visualize the data, Figs. 8-10 show the model prediction graphs with each feature compared against each other and their associated correct and incorrect predictions.

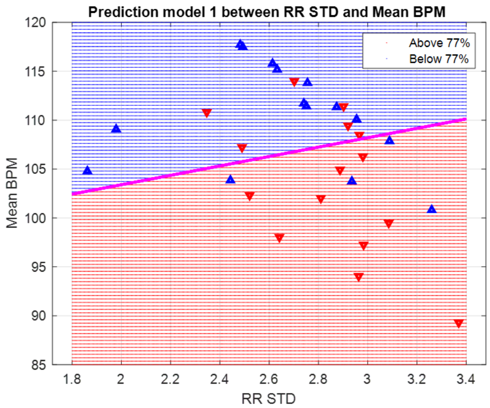

Fig. 8. Prediction Model between RR STD and Mean BPM. Pink line indicating the decision boundary

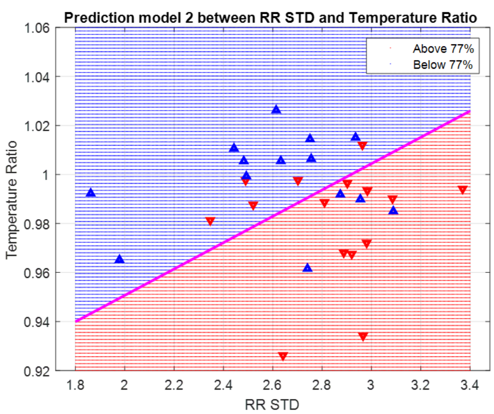

Fig. 9. Prediction Model between RR STD and Temperature Ratio. Pink line indicating the decision boundary

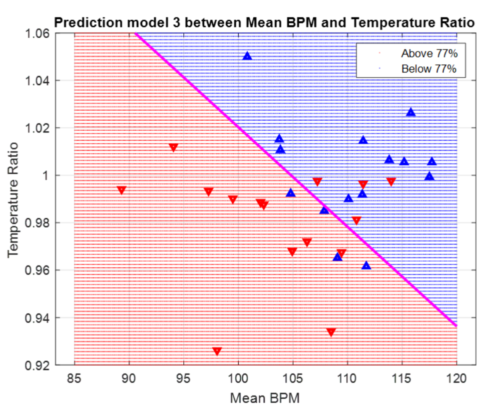

Fig. 10. Prediction Model between Mean BPM and Temperature Ratio. Pink line indicating the decision boundary

Using the data from the confusion matrix and equations 1-5, the model performance metrics were calculated using and are summarized in table II.

TABLE II
SUMMARY OF PERFORMANCE METRICS

Performance Metric	Value
Classification Accuracy	70%
Sensitivity	66.67%
Specificity	73.3%
Precision	71.4%
F1 Score	68.9%
AUC	78.67%

IV. DISCUSSION

1) Signal Processing: Signal processing was successfully used to isolate the tonic component of EDA signals, attenuate noise in BVP, EDA, and temperature signals, and reduce the impact of motion artifacts on the BVP and EDA signals. However, inherent characteristics of the signals and pre-processing techniques were observed and could be improved upon in future work. Firstly, the frequency ranges of body movement and respiratory motion artifacts overlap with the lower end of the BVP signal spectrum [17]. Therefore, it is nnot possible to fully remove these artifacts from BVP signals using conventional filtering [17]. Furthermore, EDA signals are sensitive to the pressure and stretching of the electrodes and the quality of contact made between electrodes and skin, and the frequency range of the tonic component of the EDA signal overlaps with that of motion artifacts [12], [16]. While most transient spike artifacts were removed from EDA signals using a median filter, for certain signals in which a high amount of artifacts were present, a small number of transient artifacts remained in the processed signal. A potential improvement to EDA signal pre-processing includes the use of wavelet-based artifact removal reported in recent studies [32], [33].

2) Feature Extraction: The feature extraction methods used in this study successfully yielded metrics and prediction accuracy from machine learning comparable to results in the literature. PPG systems are notoriously difficult to obtain clinically accurate results from due to their tendency to capture motion artifacts and lose signal in the sensor is moved, especially in the case with a wrist-mounted PPG signal [10], [34]. While robust pre-processing and peak detection methods are proposed in this paper to process the PPG signal, further studies can use clinical ECG monitoring to improve feature accuracy and improve machine learning results. Furthermore, for more accurate features that represent students' emotions and stress levels during exam writing, baseline data needs to be collected outside of school to better characterize changes in the student's typical electrodermal activity and resting heart rate.

3) Machine Learning: Using MATLAB's linear discriminant classification algorithm, a machine learning model was trained and predicted students' test scores were above or below average with approximately 70% accuracy based on the 3 unique features extracted from the physiological signals collected as they wrote their tests. Comparing the results gathered in this report to the literature using the same data set shows a similar accuracy range of approx. 70-80% obtained using a variety of other ML algorithms, such as support vector machines and k-nearest neighbours. In addition, an analysis of the ROC curve shows an AUC of 0.7867, indicating a high degree of accuracy when predicting true positive rates and shows that the model is performing well.

While the LDA model used in this report provides various advantages when classifying the data based on performance, model complexity, data size and dimensionality. Other more complex models were also explored, which provided a higher accuracy. For example, a medium Gaussian support vector machine yielded a higher accuracy during testing (approx. 75% accuracy) using the same model parameters used for LDA. However, the increase in accuracy is undesired for this project's purposes as the complexity of the SVM model to optimize its hyperplane using more complicated convex optimization algorithms might limit someone's trust in the model. The idea of explainable machine learning is an essential factor that must be considered, as it allows more efficient debugging and increased understanding of fairness, privacy, causality, and trust in the model [35].

One of the limitations encountered during this analysis was the small sample size (n=30) of the data set. A common issue that can occur when developing a ML model for the prediction based on biomedical signals is a lack of data available to train, validate and test the model with due to a limited pool of participants and measurement equipment [22]. The use of leave-one-out cross-validation helped limit the risk of overfitting the data by iteratively taking the average classification accuracy rate after each sample was tested against the trained model. While 30 samples may leave some to question the accuracy of this model, the samples were collected from each participant over 3 different tests, allowing for longitudinal data analysis over time and the use of 30 separate sample points [10]. However, for the purposes of this analysis, the accuracy provided, combined with consistent results obtained from others who have experimented with this data set, gives a level of confidence towards the model's accuracy in predicting test scores. Additional data from future studies would be a valuable addition to help to improve the model and potentially improve the accuracy.

V. Conclusion

This paper proposed a powerful system for quantifying physiological signals recorded using the Empatica E4 wristband. Since wrist-mounted signal collection is prone to artifacts, robust pre-processing algorithms were designed to extract only meaningful features from the data with low sensitivity to artifacts. Creating quantifiable metrics from physiological data collected during examination writing is significant for both students and academic institutions. Analysis of stress metrics using this methodology presents opportunities for educational institutions to design examination formats that reduce the levels of stress experienced by students and the impact on their well-being to improve student performance. This investigation also presents value to students, as it can help them understand how to better regulate their emotions during high-stress situations to achieve the optimal level of stress for productivity and performance. The signals used in this study yielded a classification accuracy of 70% using computationally easy methods. This result, therefore, shows the viability of a wrist-mounted system that could allow for rapid, accurate detection of stress in day-to-day life. Future work can include alarger population size, with data collection starting before the examination period to allow for better baseline signal estimation.

Acknowledgment

The authors wish to thank Dr. Krishnan and Binh Nguyen for their guidance and education throughout this project. The authors also acknowledge the work of Amin *et al.* to develop the dataset analyzed in this investigation.

References

[1] X. Wang, J. Zhang, X. Sun, and L. Zhang, "Stress mindset and mental health status among Chinese high school students: The mediating role of exam stress appraisals," *PsyCh Journal*, 2022.

[2] American College Health Association, "American College Health Association-National College Health Assessment II: Canadian Consortium Executive Summary Spring 2019," 2019. [Online]. Available: www.acha-ncha.org.

[3] M. Milenkovic, "42 Worrying Workplace Stress Statistics," 9 2019. [Online]. Available: https://www.stress.org/42-worrying-workplace-stress-statistics

[4] M. Rafiul Amin, D. S. Wickramasuriya, and R. T. Faghih, "A Wearable Exam Stress Dataset for Predicting Grades using Physiological Signals," in *Healthcare Innovations and Point of Care Technologies Conference, HI-POCT 2022*. Institute of Electrical and Electronics Engineers Inc., 2022, pp. 30–36.

[5] M. Corbett, "From law to folklore: Work stress and the Yerkes-Dodson Law," *Journal of Managerial Psychology*, vol. 30, no. 6, pp. 741–752, 8 2015. [Online]. Available: /record/2015-34595-008

[6] A. J. Felkey, "The stress-productivity shortfall," 4 2021. [Online]. Available: https://www.stress.org/the-stress-productivity-shortfall#:~:text=One%20of%20the%20most%20salient,the%20cause%20of%20their%20absence.

[7] B. Linden, H. Stuart, and A. Ecclestone, "Trends in Post-Secondary Student Stress: A Pan-Canadian Study," *https://doi.org/10.1177/07067437221111365*, 7 2022. [Online]. Available: https://journals.sagepub.com/doi/full/10.1177/07067437221111365

[8] M. Kivimäki and A. Steptoe, "Effects of stress on the development and progression of cardiovascular disease," *Nature reviews. Cardiology*, vol. 15, no. 4, pp. 215–229, 4 2018. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/29213140/

[9] C. S. Aneshensel and W. R. Avison, "The Stress Process," *http://dx.doi.org/10.1177/2156869315585388*, vol. 5, no. 2, pp. 67–85, 6 2015. [Online]. Available: https://journals.sagepub.com/doi/abs/10.1177/2156869315585388

[10] F. M. Amin, D. S. Wickramasuriya, and R. T. Faghih, "A Wearable Exam Stress Dataset for Predicting Cognitive Performance in Real-World Settings v1.0.0," 2022. [Online]. Available: https://physionet.org/content/wearable-exam-stress/1.0.0/

[11] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. C. Ivanov, R. Mark, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals." *Circulation*, vol. 101, no. 23, pp. e215–e220, 6 2000.

[12] H. F. Posada-Quintero and K. H. Chon, "Innovations in electrodermal activity data collection and signal processing: A systematic review," *Sensors (Switzerland)*, vol. 20, no. 2, 1 2020. [Online]. Available: https://click.endnote.com/viewer?doi=10.3390%2Fs20020479&token=WzI1NjIwMDgsIjEwLjMzOTAvczIwMDIwNDc5Il0.T9ROO8Wlu4o-ZQC5-l1n-J01Rgg

[13] H. F. Posada-Quintero, J. P. Florian, A. D. Orjuela-Cañón, T. Aljama-Corrales, S. Charleston-Villalobos, and K. H. Chon, "Power Spectral Density Analysis of Electrodermal Activity for Sympathetic Function Assessment," *Annals of Biomedical Engineering*, vol. 44, no. 10, pp. 3124–3135, 10 2016. [Online]. Available: https://click.endnote.com/viewer?doi=10.1007%2Fs10439-016-1606-6&token=WzI1NjIwMDgsIjEwLjEwMDcvczEwNDM5LTAxNi0xNjA2LTYiXQ.xrr6TCzUXUsr_JLbq50N0LqdwIU

[14] W. Boucsein, "Methods of Electrodermal Recording," *Electrodermal Activity*, pp. 87–258, 2012.

[15] G. Fink, *Encyclopedia of stress*. Academic Press, 2000. [Online]. Available: https://go.exlibris.link/wLsYfNZD

[16] J. J. Braithwaite, D. Derrick, G. Watson, R. Jones, and M. Rowe, "A Guide for Analysing Electrodermal Activity (EDA) & Skin Conductance Responses (SCRs) for Psychological Experiments," 2015. [Online]. Available: https://www.birmingham.ac.uk/documents/college-les/psych/saal/guide-electrodermal-activity.pdf

[17] J. L. Moraes, M. X. Rocha, G. G. Vasconcelos, J. E. Vasconcelos Filho, V. C. Hugo de Albuquerque, and A. R. Alexandria, "Advances in Photopletysmography Signal Analysis for Biomedical Applications." [Online]. Available: www.mdpi.com/journal/sensors

[18] S. Bagha, S. Hills, P. Bhubaneswar, and L. Shaw, "A Real Time Analysis of PPG Signal for Measurement of SpO2 and Pulse Rate Computational Neuroscience View project A robust motifs based artifacts removal technique from EEG View project A Real Time Analysis of PPG Signal for Measurement of SpO 2 and Pulse Rate," *International Journal of Computer Applications*, vol. 36, no. 11, pp. 975–8887, 2011. [Online]. Available: https://www.researchgate.net/publication/220043686

[19] F. Isiaka, S. A. Abdulkarim, K. Mwitondi, and Z. Adamu, "Emotion detection on webpages using biosensors integrated to a window-based dynamic control system," *International Journal of Intelligent Computing and Cybernetics*, vol. 15, no. 2, pp. 277–301, 4 2022.

[20] E. Y. Ng, W. Muljo, and B. S. Wong, "Study of facial skin and aural temperature," *IEEE Engineering in Medicine and Biology Magazine*, vol. 25, no. 3, pp. 68–74, 5 2006.

[21] W. Bierman, "THE TEMPERATURE OF THE SKIN SURFACE," *Journal of the American Medical Association*, vol. 106, no. 14, pp. 1158–1162, 4 1936. [Online]. Available: https://jamanetwork-com.ezproxy.lib.ryerson.ca/journals/jama/fullarticle/1155856

[22] Sridhar Krishnan, "Biomedical Signal Analysis for Connected Healthcare - Sridhar Krishnan," 2021. [Online]. Available: https://books.google.ca/books?hl=en&lr=&id=EiMZEAAAQBAJ&oi=fnd&pg=PP1&dq=biomedical+signal+analysis+for+connected+healthcare&ots=nraU8gQoFy&sig=lbHQHcyQPD1RFJv7Hu6fszs4nLc&redir_esc=y#v=onepage&q=biomedical%20signal%20analysis%20for%20connected%20healthcare&f=false

[23] Y. Liang, M. Elgendi, Z. Chen, and R. Ward, "Analysis: An optimal filter for short photoplethysmogram signals," *Scientific Data*, vol. 5, 5 2018. [Online]. Available: https://click.endnote.com/viewer?doi=10.1038%2Fsdata.2018.76&token=WzI1NjIwMDgsIjEwLjEwMzgvc2RhdGEuMjAxOC43NiJd.ysJpshNjHPqO41AB7-XbOAtxAcM

[24] V. Gopal, "EDA Analysis and Correction."

[25] T. Pakarinen, J. Pietila, and H. Nieminen, "Prediction of Self-Perceived Stress and Arousal Based on Electrodermal Activity," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pp. 2191–2195, 7 2019.

[26] Y. Kong, H. F. Posada-Quintero, and K. H. Chon, "Sensitive Physiological Indices of Pain Based on Differential Characteristics of Electrodermal Activity," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 10, pp. 3122–3130, 10 2021.

[27] A. S. Anusha, P. Sukumaran, V. Sarveswaran, S. Surees Kumar, A. Shyam, T. J. Akl, S. P. Preejith, and M. Sivaprakasam, "Electrodermal Activity Based Pre-surgery Stress Detection Using a Wrist Wearable," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 1, pp. 92–100, 1 2020.

[28] P. Fauquet-Alekhine, L. Rouillac, J. Berton, and J.-C. Granry, "Heart Rate vs Stress Indicator for Short Term Mental Stress," *British Journal of Medicine and Medical Research*, vol. 17, no. 7, pp. 1–11, 1 2016.

[29] T. Fushiki, "Estimation of prediction error by using K-fold cross-validation," *Stat Comput*, vol. 21, pp. 137–146, 2011.

[30] B. Richard Bellman and R. Kalaba, "A Mathematical Theory OF Adaptive Control Process," 1959. [Online]. Available: https://www.pnas.org

[31] F. E. Harrell, K. L. Lee, and D. B. Mark, "Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors," *Statistics in Medicine*, vol. 15, no. 4, 1996.

[32] J. Shukla, M. Barreda-Ángeles, J. Oliver, and D. Puig, "Efficient wavelet-based artifact removal for electrodermal activity in real-world applications," *Biomedical Signal Processing and Control*, vol. 42, pp. 45–52, 4 2018.

[33] W. Chen, N. Jaques, S. Taylor, A. Sano, S. Fedor, and R. W. Picard, "Wavelet-based motion artifact removal for electrodermal activity," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, vol. 2015, pp. 6223–6226, 11 2015. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/26737714/

[34] W. J. Lin and H. P. Ma, "A physiological information extraction method based on wearable PPG sensors with motion artifact removal," in *2016 IEEE International Conference on Communications, ICC 2016*. Institute of Electrical and Electronics Engineers Inc., 7 2016.

[35] A. Heuillet, F. Couthouis, and N. Díaz-Rodríguez, "Explainability in deep reinforcement learning," *Knowledge-Based Systems*, vol. 214, p. 106685, 2 2021.