



MASTER'S IN INDUSTRIAL ENGINEERING

MASTER'S THESIS

**SELF-SUPERVISED LEARNING FOR
ACUTE STRESS DETECTION IN
ELECTROCARDIOGRAMS**

Author: Francisco Barragán Castro

Supervisor: Berta Ruíz González

Co-Supervisor: Álvaro López López

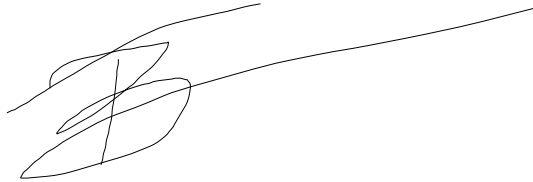
Madrid

July 2024

Declaro, bajo mi responsabilidad, que el Proyecto presentado con el título
“SELF-SUPERVISED LEARNING FOR ACUTE STRESS DETECTION IN
ELECTROCARDIOGRAMS”

en la ETS de Ingeniería - ICAI de la Universidad Pontificia Comillas en el
curso académico 2023/24 es de mi autoría, original e inédito y
no ha sido presentado con anterioridad a otros efectos.

El Proyecto no es plagio de otro, ni total ni parcialmente y la información que ha sido
tomada de otros documentos está debidamente referenciada.



Fdo.: Francisco Barragán Castro

Fecha: 11/ 07/ 2024

Autorizada la entrega del proyecto

EL DIRECTOR DEL PROYECTO

Fdo.: Berta Ruíz González Fecha: 16/07/2024

Álvaro López López Fecha: 16/07/2024



MASTER'S IN INDUSTRIAL ENGINEERING

MASTER'S THESIS

**SELF-SUPERVISED LEARNING FOR
ACUTE STRESS DETECTION IN
ELECTROCARDIOGRAMS**

Author: Francisco Barragán Castro

Supervisor: Berta Ruíz González

Co-Supervisor: Álvaro López López

Madrid

SELF-SUPERVISED LEARNING FOR ACUTE STRESS DETECTION IN ELECTROCARDIOGRAMS

Autor: Barragán Castro, Francisco

Director: Ruíz González, Berta

Entidad Colaboradora: Instituto de Investigación Tecnológica

RESUMEN DEL PROYECTO

Este estudio presenta un modelo basado en el *transformer* diseñado para la detección precisa de estrés mediante señales de electrocardiograma (ECG) de una sola derivación. El modelo combina el aprendizaje auto-supervisado con la arquitectura de transformadores para procesar eficientemente datos de ECG en bruto, garantizando al mismo tiempo la explicabilidad a través de *saliency maps*, estableciendo un nuevo estándar para la detección de estrés basada en ECG utilizando el conjunto de datos WESAD.

Palabras clave: Transformers, Electrocardiograma (ECG), Detección de Estrés, Aprendizaje Auto-Supervisado, IA en Salud, Interpretabilidad de IA Médica, Conjunto de Datos WESAD

INTRODUCCIÓN

El estrés es un problema de salud global en aumento, que afecta significativamente la salud mental y la calidad de vida. La tecnología *wearable*, particularmente los sensores que monitorean señales fisiológicas como los ECG, ofrece una solución prometedora para el análisis del estrés. Los ECG, que reflejan la actividad eléctrica del corazón, pueden revelar información sobre la respuesta al estrés de un individuo. Los avances recientes en aprendizaje profundo y tecnología de sensores han revolucionado la atención médica, haciendo posible el monitoreo continuo de la salud de manera no intrusiva.

TRABAJOS RELACIONADOS

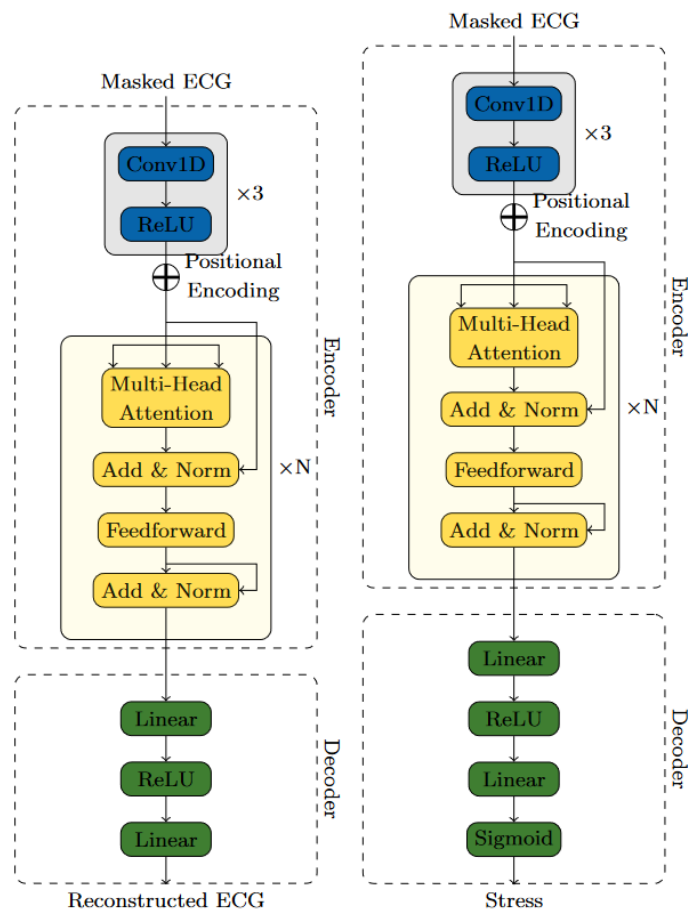
Los modelos basados en *transformers* están emergiendo como herramientas líderes para analizar señales de ECG [1], superando enfoques tradicionales como las Redes Neuronales Convolucionales (CNN). La capacidad de los transformadores para manejar datos secuenciales y capturar dependencias a largo plazo los hace ideales para el análisis de señales de ECG, donde la relación temporal entre diferentes partes de la señal es crucial.

METODOLOGÍA

El método propuesto integra el aprendizaje auto-supervisado con la arquitectura de *transformers* para procesar datos de ECG en bruto de manera efectiva, garantizando la interpretabilidad. El aprendizaje auto-supervisado permite que el modelo maneje autónomamente datos no etiquetados, mejorando su capacidad para aprender representaciones útiles. La arquitectura de *transformers* mejora la precisión del modelo al capturar dependencias a largo plazo en datos secuenciales.

Un proceso de entrenamiento en dos etapas construye una arquitectura *encoder-decoder*. La primera etapa de preentrenamiento emplea el aprendizaje auto-supervisado para desarrollar un *encoder* robusto que captura las partes esenciales de la señal de ECG en un espacio latente. La segunda etapa ajusta el *encoder* preentrenado y un nuevo *decoder* para entrenar un modelo de detección de estrés.

Figura 1: Arquitectura del Modelo



PREENTRENAMIENTO AUTO-SUPERVISADO

En esta primera etapa, el modelo predice partes ocultas de la señal de ECG utilizando una máscara binaria, obligándolo a comprender la estructura subyacente de la forma de onda del ECG. La pérdida de error cuadrático medio (MSE) se centra en las secciones enmascaradas de la señal de ECG, mejorando la capacidad del modelo para manejar datos del mundo real. La arquitectura del codificador comprende una CNN seguida de un *transformer* para capturar patrones de latidos y dependencias globales dentro de la secuencia, proporcionando una comprensión integral de las señales de ECG.

AJUSTE FINO PARA LA DETECCIÓN DE ESTRÉS

Para la segunda fase, se utiliza la pérdida de entropía cruzada binaria para la tarea de detección de estrés. El ajuste fino de todo el codificador permite que el modelo ajuste todas las características aprendidas para adaptarse mejor a los patrones de ECG relacionados con el estrés. La arquitectura del decodificador de estrés consiste en una red lineal de dos capas que transforma los datos codificados en una predicción de estrés.

EXPERIMENTOS

Para probar el modelo, se crearon múltiples configuraciones, cambiando el tamaño general del modelo. Esto se hizo cambiando la duración de la señal dada al modelo, 8 y 4 segundos, y el número de capas del *transformer encoder*. Además, se probó el efecto de desbloquear todas las capas del *transformer encoder*, o solo la última capa.

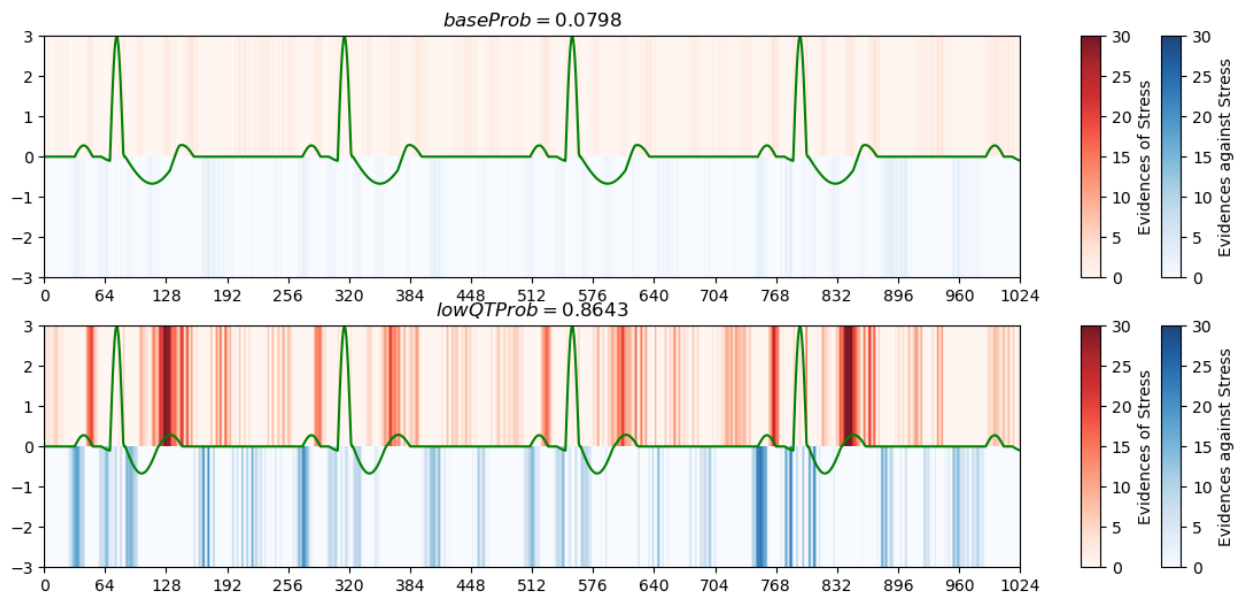
Figura 2: Comparación de resultados

Dataset	Ref.	Method		Acc. [%]	F ₁ [%]	
WESAD	[2]	QDA		85.7	-	
	[3]	LDA		85.4	81.3	
	[4]	CNN		92.0	81.8	
	[5]	Image		92.5	-	
	[1]	TF		91.1	83.3	
	Ours	SSL + TF	8s-2l	full	99.3	98.5
				last	94.6	87.4
			4s-1l	full	99.2	98.2
				last	91.4	77.9

EXPLICABILIDAD DEL MODELO

Los *saliency maps* visualizan el proceso de toma de decisiones del modelo, destacando las partes de los datos de entrada que fueron más importantes para las predicciones. Se crearon ECGs artificiales alterando características específicas de la forma de onda, confirmando que las predicciones del modelo se alinean con los indicadores médicos establecidos de estrés.

Figura 3: Análisis del efecto del intervalo QT en el Saliency map



CONCLUSIONES Y TRABAJOS FUTUROS

En general, todos nuestros modelos, excepto el más pequeño (4s-1l last), logran resultados de vanguardia en el conjunto de datos WESAD, demostrando la efectividad del aprendizaje auto-supervisado para esta tarea. Los trabajos futuros incluyen expandir el conjunto de datos, desplegar el modelo en escenarios del mundo real, reducir la complejidad del modelo y explorar otros métodos de optimización. Este estudio establece un nuevo punto de referencia para la detección de estrés basada en ECG, fusionando técnicas avanzadas de IA con aplicaciones médicas prácticas.

REFERENCIAS

- [1] B. Behinaein, A. Bhatti, D. Rodenburg, P. Hungler, and A. Etemad, "A Transformer Architecture for Stress Detection from ECG," in *2021 International Symposium on Wearable Computers*, Sep. 2021, pp. 132–134. doi: 10.1145/3460421.3480427.

- [2] P. Bota, C. Wang, A. Fred, and H. Silva, "Emotion Assessment Using Feature Fusion and Decision Fusion Classification Based on Physiological Data: Are We There Yet?," *Sensors*, vol. 20, no. 17, Art. no. 17, Jan. 2020, doi: 10.3390/s20174723.
- [3] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection," in *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, Boulder CO USA: ACM, Oct. 2018, pp. 400–408. doi: 10.1145/3242969.3242985.
- [4] E. Castets, *Edouard99/Stress_Detection_ECG*. (Aug. 03, 2023). Jupyter Notebook. Accessed: Sep. 05, 2023. [Online]. Available: https://github.com/Edouard99/Stress_Detection_ECG
- [5] S. Ishaque, N. Khan, and S. Krishnan, "Detecting stress through 2D ECG images using pretrained models, transfer learning and model compression techniques," *Mach. Learn. Appl.*, vol. 10, p. 100395, Dec. 2022, doi: 10.1016/j.mlwa.2022.100395.

SELF-SUPERVISED LEARNING FOR ACUTE STRESS DETECTION IN ELECTROCARDIOGRAMS

Author: Barragán Castro, Francisco

Supervisor: Ruíz González, Berta

Collaborating entity: Insituto de Investigación Tecnológica

ABSTRACT

This study introduces a transformer model designed for accurate stress detection via 1-lead electrocardiogram (ECG) signals. The model merges self-supervised learning with transformer architecture to efficiently process raw ECG data while ensuring explainability through saliency maps, establishing a new standard for ECG-based stress detection using the WESAD dataset.

Keywords: Transformer Models, Electrocardiogram (ECG), Stress Detection, Self-Supervised Learning, AI in Healthcare, Medical AI Interpretability, WESAD Dataset

INTRODUCTION

Stress is a growing global health issue, significantly affecting mental health and quality of life. Wearable technology, particularly sensors monitoring physiological signals like ECGs, offers a promising solution for stress analysis. ECGs, reflecting the heart's electrical activity, can reveal insights into an individual's stress response. Recent advancements in deep learning and sensor technology have revolutionized healthcare, making continuous non-intrusive health monitoring feasible.

RELATED WORK

Transformer-based models are emerging as leading tools for analysing ECG signals, outperforming traditional approaches like Convolutional Neural Networks (CNNs). Transformers' ability to handle sequential data and capture long-term dependencies makes them ideal for ECG signal analysis, where the temporal relationship between different parts of the signal is crucial.

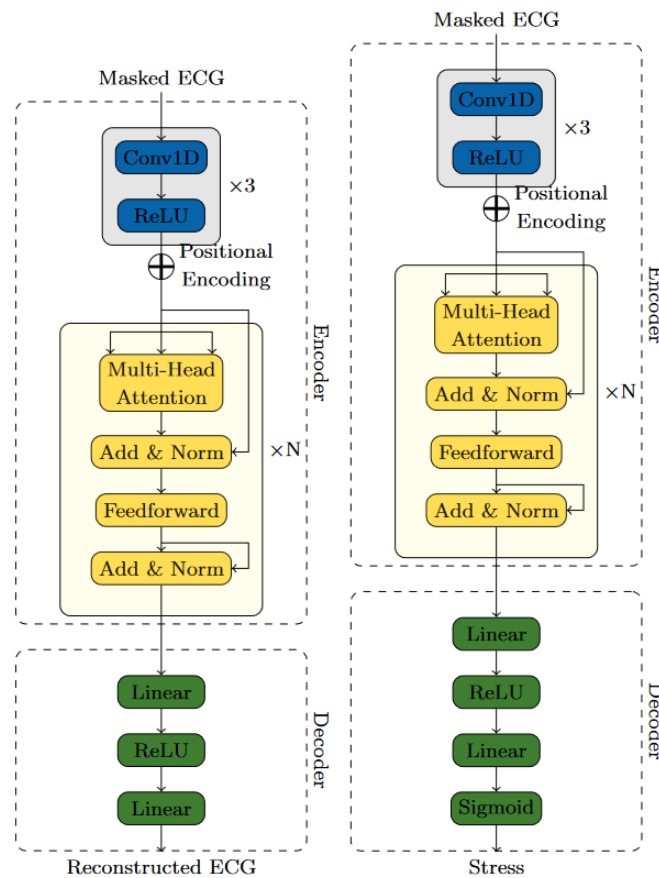
METHODOLOGY

The proposed method integrates self-supervised learning with transformer architecture to process raw ECG data effectively while ensuring interpretability. Self-supervised learning allows the model to autonomously handle unlabeled

data, improving its capability to learn useful representations. The transformer architecture enhances the model's accuracy by capturing long-term dependencies in sequential data.

A two-stage training process builds an encoder-decoder architecture. The first pretraining stage employs self-supervised learning to develop a robust encoder that captures essential parts of the ECG signal in a latent space. The second stage fine-tunes the pretrained encoder to train a stress detection model.

Figure 1: Model's architecture



SELF-SUPERVISED PRETRAINING

In this first stage, the model predicts occluded parts of the ECG signal using a binary mask, forcing it to understand the underlying structure of the ECG waveform. Mean Square Error (MSE) loss focuses on the masked sections of the ECG signal, improving the model's ability to handle real-world data. The encoder architecture comprises a CNN followed by a Transformer to capture heartbeat

patterns and global dependencies within the sequence, providing a comprehensive understanding of the ECG signals.

FINE-TUNING FOR STRESS DETECTION

For the second phase, Binary Cross-Entropy with Logits Loss is used for the stress detection task. Fine-tuning the entire encoder allows the model to adjust all learned features to better suit stress-related ECG patterns. The stress decoder architecture consists of a two-layer linear network that transforms encoded data into a stress prediction.

EXPERIMENTS

To test the model, multiple configurations were created, changing the overall size of the model. This was done by changing the duration of the signal given to the model, 8 and 4 seconds, and the number of layers of the transformer encoder.

In addition, it is also tested the effect of unlocking all the layers of the transformer encoder, or just the last layer.

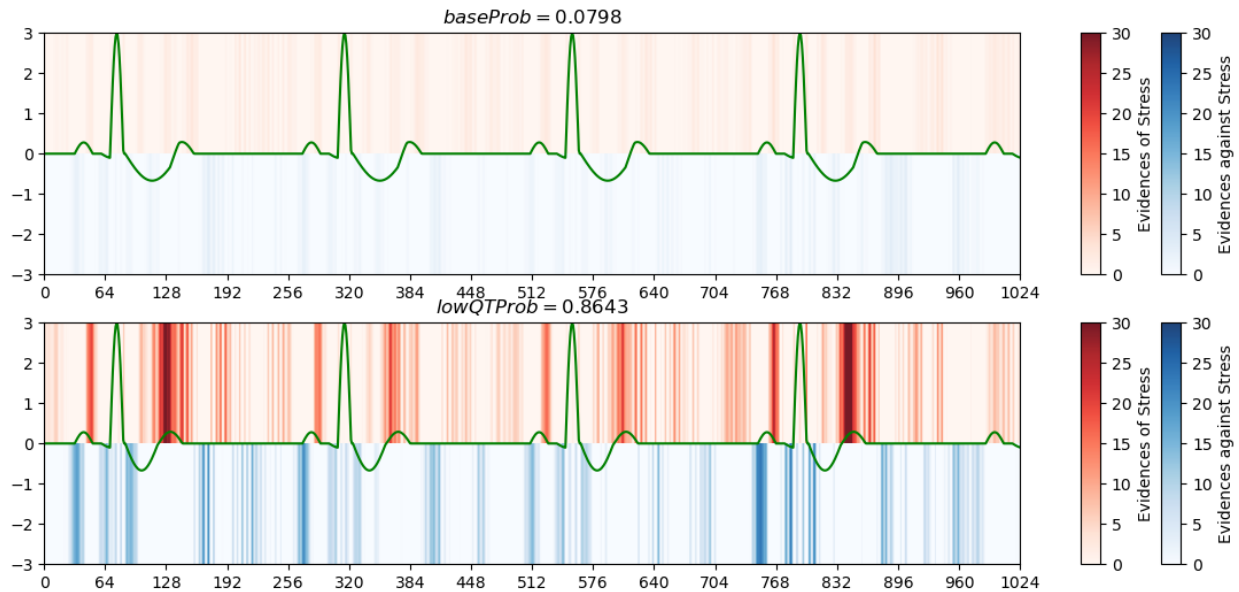
Figure 2: Performance Comparison

Dataset	Ref.	Method		Acc. [%]	F ₁ [%]	
WESAD	[2]	QDA		85.7	-	
	[3]	LDA		85.4	81.3	
	[4]	CNN		92.0	81.8	
	[5]	Image		92.5	-	
	[1]	TF		91.1	83.3	
	Ours	SSL + TF	8s-2l	full	99.3	98.5
				last	94.6	87.4
			4s-1l	full	99.2	98.2
				last	91.4	77.9

MODEL EXPLAINABILITY

Saliency maps visualize the decision-making process of the model, highlighting the parts of the input data that were most important for predictions. Artificial ECGs were created by altering specific waveform features, confirming that the model's predictions align with established medical indicators of stress.

Figure 3: QT interval saliency map analysis



CONCLUSIONS AND FUTURE WORK:

Overall, all but the smallest (4s-1l last) our models achieve state of the art results on the WESAD dataset, proving the effectiveness of self-supervised learning for this task. Future work includes expanding the dataset, deploying the model in real-world scenarios, reducing model complexity, and exploring other optimization methods. This study sets a new benchmark for ECG-based stress detection, merging advanced AI techniques with practical medical applications.

REFERENCES

- [1] B. Behinaein, A. Bhatti, D. Rodenburg, P. Hungler, and A. Etemad, "A Transformer Architecture for Stress Detection from ECG," in *2021 International Symposium on Wearable Computers*, Sep. 2021, pp. 132–134. doi: 10.1145/3460421.3480427.
- [2] P. Bota, C. Wang, A. Fred, and H. Silva, "Emotion Assessment Using Feature Fusion and Decision Fusion Classification Based on Physiological Data: Are We There Yet?," *Sensors*, vol. 20, no. 17, Art. no. 17, Jan. 2020, doi: 10.3390/s20174723.
- [3] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection," in *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, Boulder CO USA: ACM, Oct. 2018, pp. 400–408. doi: 10.1145/3242969.3242985.

- [4] E. Castets, *Edouard99/Stress_Detection_ECG*. (Aug. 03, 2023). Jupyter Notebook. Accessed: Sep. 05, 2023. [Online]. Available: https://github.com/Edouard99/Stress_Detection_ECG
- [5] S. Ishaque, N. Khan, and S. Krishnan, "Detecting stress through 2D ECG images using pretrained models, transfer learning and model compression techniques," *Mach. Learn. Appl.*, vol. 10, p. 100395, Dec. 2022, doi: 10.1016/j.mlwa.2022.100395.

Glossary

Abbreviation	Full Term
AI	Artificial Intelligence
ANS	Autonomic Nervous System
AUC	Area Under Curve
AV	Atrioventricular
AV Node	Atrioventricular Node
BiLSTM	Bidirectional Long Short-Term Memory
CNN	Convolutional Neural Network
ECG	Electrocardiogram
EKG	Electrocardiography (alternate abbreviation for ECG)
GeLU	Gaussian Error Linear Unit
HR	Heart Rate
HRV	Heart Rate Variability
IoT	Internet of Things
LDA	Linear Discriminant Analysis
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MLP	Multilayer Perceptron
MSE	Mean Square Error
PERT	Perturbation
PR	Pulse Repetition
QDA	Quadratic Discriminant Analysis
QR	Quantitative Risk
QT	Q-wave to T-wave
ReLU	Rectified Linear Unit
ROC	Receiver Operating Characteristic
SA	Sinoatrial
SDGs	Sustainable Development Goals
SGD	Stochastic Gradient Descent
SSL	Self-Supervised Learning
TF	Transformer
TSST	Trier Social Stress Test
W&B	Weights & Biases
WESAD	Wearable Stress and Affect Detection

Index

1 Introduction	4
1.1 <i>Motivation</i>	4
1.2 <i>Project Objectives</i>	4
1.3 <i>Resources to be used.</i>	5
1.3.1 Software and Libraries	5
1.3.2 Hardware Requirements	7
1.3.3 Data Resources	7
1.4 <i>Alignment with the Sustainable Development Goals (SDGs)</i>	8
2 Medical background	12
2.1 <i>Introduction to Electrocardiography</i>	12
2.2 <i>Lead systems</i>	12
2.2.1 12-lead ECG	13
2.2.2 3-lead ECG	13
2.2.3 1-lead ECG	14
2.2.4 Radar electrocardiography	14
2.3 <i>Understanding the Origins of ECG Waveforms</i>	15
2.3.1 Electrical Waves in the Heart	15
2.3.2 Polarization and Repolarization	16
2.3.3 Critical Role of the SA and AV Nodes	16
2.4 <i>Wave morphology</i>	17
2.5 <i>The Role of the Nervous System</i>	18
2.5.1 Autonomic Nervous System (ANS)	18
2.5.2 The SA and AV Nodes in the Context of the Nervous System	19
2.5.3 Integration with Cardiac Cycle	19
2.6 <i>ECG stress detection</i>	20
3 State of the art	22
3.1 <i>Problem statement</i>	22
3.2 <i>Related works</i>	22
3.2.1 Stress detection	23
3.2.2 Emotion recognition	25
3.3 <i>ECG as a Data Source for Stress Detection</i>	26
3.3.1 Heart Rate Variability (HRV)	26
3.3.2 Wave Morphology Analysis	27
3.4 <i>Selected leads</i>	27
3.5 <i>Selected technologies</i>	29
3.5.1 Transformers	29

3.5.2	Self-Supervised Learning	30
4	Methodology	32
4.1	Pre-training via Self-Supervised Learning	33
4.1.1	SSL Task	34
4.1.2	Loss function	35
4.1.3	Encoder architecture	36
4.1.4	CNN Encoder	36
4.1.5	Transformer Encoder	37
4.1.6	Decoder	38
4.1.7	Training	39
4.2	Classifier	40
4.2.1	Loss function	41
4.2.2	Encoder	41
4.2.3	Decoder	41
4.2.4	Training	42
4.3	Data acquisition	43
4.3.1	Self-Supervised Dataset	44
4.3.2	Classifier Dataset.	45
5	Results	48
5.1	Quality of the models' predictions	48
5.1.1	Confusion Matrix Analysis	48
5.1.2	Receiver Operating Characteristic (ROC) Curve and Area Under Curve (AUC)	49
5.1.3	Histogram analysis	50
6	Interpretability of the results	53
6.1	Saliency maps	53
6.2	Saliency map of the encoder	54
6.3	Saliency map of the output	56
6.3.1	Simulated ECGs	58
6.3.2	RR Interval	59
6.3.3	PR Interval	60
6.3.4	T amplitude	61
6.3.5	QT interval	62
6.4	Attention heatmaps	63
7	Conclusions and future works	67
7.1	Future works	68
8	Bibliography	71

Figure Index

<i>x</i> Figure 1 12 lead ECG	13
Figure 2 3 Lead ECG	13
Figure 3 1 lead ECG.....	14
Figure 4 Radar electrocardiogram [1]	15
Figure 5 Diagram of the heart	16
Figure 6 QRS complex [7].....	17
Figure 7 Lead position in respect to the heart.	28
Figure 8 Delineation errors regarding the automatic detections at 1000 Hz of the ptbdb (in ms) [21]	30
Figure 9 General architecture of the model.....	32
Figure 10 ECG masking strategy.....	34
Figure 11 Self-supervised task architecture	33
Figure 12 Encoder architecture.....	36
Figure 13 CNN network specifications	37
Figure 14 Transformer encoder block	38
Figure 15 SSL Training curves	39
Figure 16 Classifier architecture	40
Figure 17 Classification hyperparameter sweep.....	43
Figure 18 Placement of the RespiBAN and the ECG, EDA,[2].....	45
Figure 19 The two different versions of the study protocol	46
Figure 20 Saliency map of the encoder.....	54
Figure 21 Saliency map of the output.....	57
Figure 22 Synthetic ECG.....	58
Figure 23 RR interval.....	60
Figure 24 Low PR interval	61
Figure 25 Increase of T wave amplitude.	62
Figure 26 Low QT interval.....	63
Figure 27 Attention heatmap of a non-stressed patient	64
Figure 28 Attention heatmap of a stressed patient.....	65

1

Introduction

With the continuous progress of technology, numerous aspects of life have become more data-driven and interconnected. This project situates itself at the intersection of medical technology and artificial intelligence. We aim to harness the advancements in Electrocardiogram (ECG) technology and Deep Learning to improve stress detection using raw ECG data. This endeavour not only increases the potential of stress management techniques but also opens up novel possibilities in preventive health measures, and continuous measurement.

1.1 Motivation

The motivation behind this project is multidimensional. There is an anticipated significant increase in the use of ECG technology in the coming years, led by radar sensors, presenting a critical opportunity to enhance the current stress detection models. Furthermore, by understanding the role of stress in activities like driving, we can work towards increasing safety measures. Lastly, the potential benefits to preventive healthcare and overall human wellbeing serve as powerful motivators to undertake this project.

1.2 Project Objectives

Based on the motivation and the current state of the matter, the following are the potential objectives for this project:

1. **Development of a robust stress detection model:** We aim to construct a model capable of detecting stress from raw ECG signals. This model will utilize advances in deep learning, specifically transformer models, to understand the complex time-series data in ECG signals.
2. **Use of self-supervised learning techniques:** Considering the abundance of unlabelled ECG data and the proven success of self-supervised learning in managing sequential data, we intend to employ self-supervised strategies. This approach will enable our model to learn useful representations from the data itself, which can then be exploited for stress detection.
3. **Benchmarking and comparison with existing models:** We aim to compare our model's performance with existing stress detection models, both those that use raw ECG data and those that rely on extracted features.
4. **Assessing the interpretability of the model:** Lastly, the project will also aim to understand the decisions the model has made, and try to see if they align with the medical consensus to ensure a correct diagnosis.

1.3 Resources to be used.

1.3.1 Software and Libraries

The project is exclusively written on the Python programming language, which is a prevalent choice in the domain of deep learning and data science. Python's syntax is user-friendly and, coupled with a robust ecosystem of libraries, it forms an excellent platform for constructing complex deep learning models. The following are the primary libraries used in this project:

- **PyTorch:** PyTorch, a key component of our project, is a widely used open-source machine learning library developed primarily by Facebook's artificial intelligence research group. It's based on the Torch library and is highly

regarded for its dynamic computational graph and efficient memory usage, which makes it suitable for our project. PyTorch offers a rich set of features for deep learning.

- **PyTorch Lightning:** We've used PyTorch Lightning, a lightweight wrapper for PyTorch, to make the code more maintainable and readable. This library streamlines the PyTorch code, reduces the amount of boilerplate code, and enhances the reproducibility of the code. It enables us to focus more on the development of the model rather than routine coding tasks, without compromising performance or flexibility.
- **Scikit-learn:** Scikit-learn, renowned for its effective predictive data analysis tools, is a crucial resource for this project. This library is primarily used for data manipulation, as well as model evaluation and performance metrics.
- **Weights & Biases (W&B):** Weights & Biases, used for experiment tracking and visualization, helps in tracking machine learning experiments. It logs and visualizes metrics, system information, hyperparameters, and more, providing us with a clear and accessible method to monitor our models' performance and progress. Its 'sweep' feature has been integral to the model, enabling us to traverse the hyperparameter space using Bayesian Optimization with minimal additional code.
- **NeuroKit2:** Specializing in neurophysiological signal processing, NeuroKit2 is a Python library we use to process and analyse our ECG data. It simplifies the task of ECG signal processing and extraction of meaningful features, saving us substantial time and effort. This library ensures that our model can access the most pertinent information contained in the ECG signals for stress detection.

1.3.2 Hardware Requirements

Our project necessitates a robust hardware infrastructure, given the computationally intensive nature of deep learning tasks. We have secured powerful resources, including top-tier Graphics Processing Units (GPUs), to significantly expedite data processing and model training.

- **Server with Dual RTX 4090:** We have confirmed access to a server equipped with two RTX 4090 graphics cards. These state-of-the-art GPUs, provided by the Instituto de Investigación Tecnológica (IIT), greatly aid the project's development with their exceptional performance. They are ideally suited for our machine learning operations. The GPUs' high processing power accelerates data processing and model training, and their substantial VRAM size (24Gb) allows us to use large batch sizes, enhancing the efficiency of our models.
- **Laptop with RTX 2070:** Alongside the server equipped with RTX 4090 graphics cards, we will utilize a laptop with an integrated RTX 2070 graphics card. This setup provides a portable solution for the initial development, analysis, and review of our models. Although the RTX 2070 is not as powerful as the RTX 4090, it still delivers substantial performance and proves invaluable for preliminary model training and testing.

These dedicated hardware resources will equip us to efficiently handle the computational demands of our project, enabling us to rapidly iterate our models, perform thorough testing, and ultimately develop a reliable and robust ECG-based stress detection system.

1.3.3 Data Resources

In terms of data resources, this project will not involve any data gathering processes. Instead, we will rely on publicly available datasets for model training and validation. The use of public datasets brings a few advantages:

1. **Efficiency and Cost-effectiveness:** Data collection can be a lengthy, expensive, and often complex process involving participant recruitment, data recording, and cleaning. By using publicly available datasets, we can focus our efforts and resources on the core aspect of the project, i.e., developing and refining our stress detection model.
2. **Model Validation and Comparison:** Public datasets offer a common ground for testing various models. As these datasets are widely used within the research community, they allow for a fair and direct comparison of our model's performance against other techniques featured in related works.

The specific datasets for this project will be chosen based on their relevance to stress detection from ECG data, their size, quality, and the variety of the data they contain. This ensures that our model is trained and validated on diverse and representative data.

A later section will delve into the details of the selected datasets, including their sources, data size, and features. Understanding these details is essential, as the chosen datasets will ultimately shape and guide the development and validation of our model.

1.4 Alignment with the Sustainable Development Goals (SDGs)

In our rapidly evolving world, ensuring sustainable growth and development remains paramount. To this end, the Sustainable Development Goals (SDGs) set forth by the United Nations act as a global blueprint, guiding concerted efforts towards building a brighter and more equitable future. In particular, our project deeply resonates with two pivotal goals: Goal 3, which emphasizes Good Health and Well-being, and Goal 9, which advocates for Industry, Innovation, and Infrastructure. By aligning our objectives and strategies with these SDGs, we aim

to create a holistic impact, addressing both health concerns and technological advancements in tandem.

Goal 3: Good Health and Well-being

Stress is a significant health concern that can lead to severe physical and mental health issues if not properly managed. By developing an advanced stress detection model using ECG technology and deep learning, our project is inherently aligned with SDG 3. Our model's primary objective is to help individuals detect their stress levels accurately and in real-time, which could lead to more timely interventions and treatments.

Moreover, stress often acts as a precursor to various cardiovascular and mental health disorders. By providing an efficient and reliable stress detection tool, we may help reduce the prevalence and impact of these stress-related conditions. It's a step forward towards achieving the targets of SDG 3, which includes ending epidemics of communicable diseases and reducing mortality from non-communicable diseases and promoting mental health and well-being.

Goal 9: Industry, Innovation, and Infrastructure

The implementation of cutting-edge technology such as deep learning, ECG technology, and self-supervised learning techniques to enhance stress detection models is an embodiment of SDG 9. We're leveraging the potential of these technologies to drive innovation and provide solutions to health challenges, thus fostering technological development in healthcare.

Furthermore, this project also contributes to building resilient infrastructure and fostering innovation. The creation of robust stress detection models will augment the capabilities of current healthcare infrastructures, particularly those related to mental health care. This aspect aligns with the targets of SDG 9, which includes upgrading infrastructure and retrofitting industries to make them sustainable, with increased resource-use efficiency and greater adoption of clean and environmentally sound technologies and industrial processes.

In conclusion, this project aligns with the SDGs by capitalizing on innovative technology for promoting well-being and supporting sustainable industrial development. The positive outcomes from this project can have a ripple effect, leading to broader benefits in health, industry, and beyond, contributing to the overarching aim of the SDGs – a better and more sustainable future for all.



UNIVERSIDAD PONTIFICIA COMILLAS
ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA (ICAI)
MASTER'S IN INDUSTRIAL ENGINEERING

2

Medical background

In this section, we provide a comprehensive insight into the current state of electrocardiography. This foundational knowledge will enhance our understanding of the critical decisions made during the model development.

2.1 Introduction to Electrocardiography

Electrocardiography, abbreviated as ECG or EKG, is a method used to measure and interpret the heart's electrical activity. By placing electrodes in standardized positions on a patient's arms and chest, electrical impulses generated by cardiac tissue during polarization and depolarization are captured. These waveforms provide essential insights into the heart's rhythm and its overall state.

In the past, ECGs were primarily tools for medical professionals. However, recent technological advancements have extended their usage to non-medical domains, enabling a more accessible and less invasive mode of data collection.

2.2 Lead systems

The foundation of electrocardiography lies in the various lead systems utilized. These lead systems determine the dimensionality and depth of the heart's electrical activity representation. From the comprehensive 12-lead ECG to the simplicity of the 1-lead, each system serves unique diagnostic purposes and offers specific insights into cardiac health.

2.2.1 12-lead ECG

The 12-lead ECG is the gold standard for diagnosis. This system captures the heart's electrical activity in 3D, employing electrodes to create 6 vertical and 6 horizontal visual axes. It is widely used for both resting and stress ECGs.

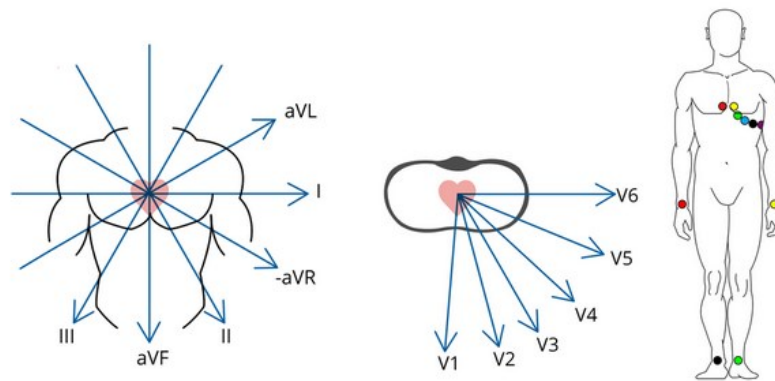


Figure 1 12 lead ECG[6]

2.2.2 3-lead ECG

Commonly used for 24-hour readings, 3-lead ECGs are instrumental in diagnosing heart problems. The Holter monitor is a typical example of this system, offering a long-term reading method.

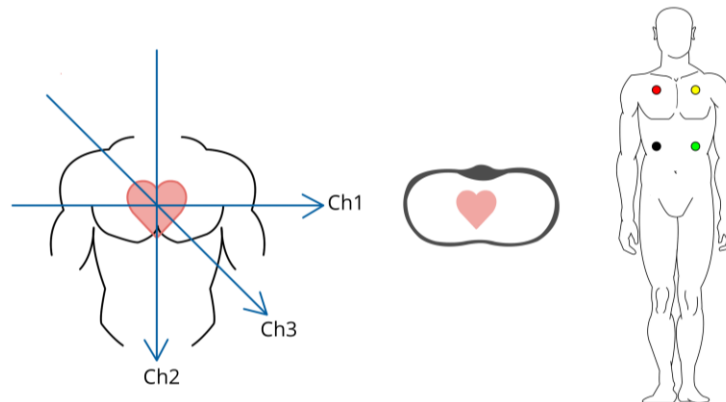


Figure 2 3 Lead ECG[6]

2.2.3 1-lead ECG

1-Lead ECG systems have gained popularity in recent years, as they provide great comfort to the user. They range from short-duration recordings, like the *AliveCor*, to extended recordings up to 14 days, as seen with *iRhythm*[6]. Despite their convenience, these systems present limitations. For instance, the intricacies of signals like Ischaemia (restricted blood flow) and extra ventricular beats are more discernible with higher lead systems. However, the 1-Lead system remains valuable, especially in the wearables domain, as it offers essential data like Heart Rate (HR) and Heart Rate Variability (HRV), which are central to deep learning applications.

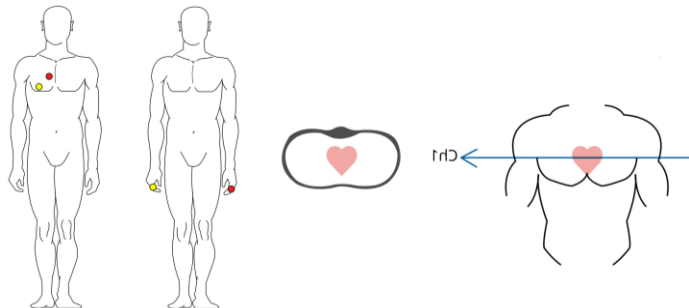


Figure 3 1 lead ECG[6]

2.2.4 Radar electrocardiography

Radar electrocardiography is an emerging research area that presents significant promise as an innovative approach for capturing ECG data with minimal intrusion. Unlike traditional methods that use electrodes, this technique employs millimetre-wave radar to detect the minute movements of the heart on the patient's chest.

It's important to note that the output from this method is susceptible to noise interference. However, through the employment of signal processing algorithms and deep learning, researchers have effectively filtered these movements and correlated them with the conventional electrical signals generated by electrodes.

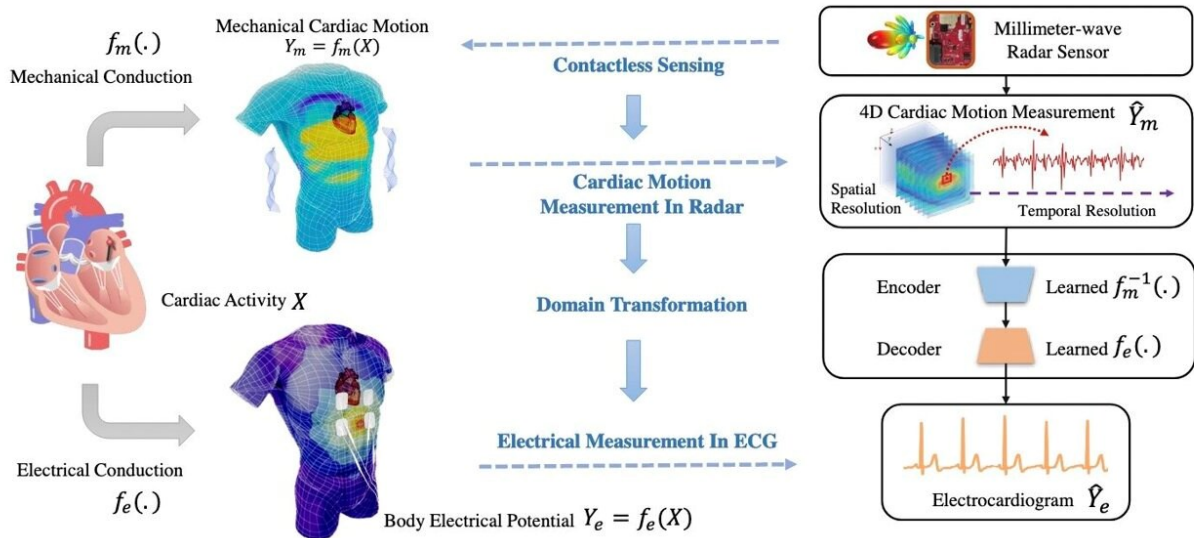


Figure 4 Radar electrocardiogram [1]

In a recent study[7], the researchers reported a median timing error of under 14 milliseconds, with a morphology accuracy surpassing 90% of standard ECG techniques. Furthermore, a 90th-percentile error of just 9ms was observed for R-R interval timing, a critical measure in diagnosing heart arrhythmias, underscoring its potential medical significance.

This advancement introduces a revolutionary approach to ECG monitoring. It not only showcases enormous potential for cardiovascular disease diagnosis but also paves the way for future medical applications rooted in non-intrusive and more dependable systems.

2.3 Understanding the Origins of ECG Waveforms

2.3.1 Electrical Waves in the Heart

The heart's electrical cycle commences at the sinoatrial node (SA node), the heart's natural pacemaker. This node generates an electrical impulse that traverses the atria, inducing their contraction. After a brief pause at the atrioventricular node

(AV node), the impulse travels through the ventricles' conduction system, prompting their synchronized contraction.

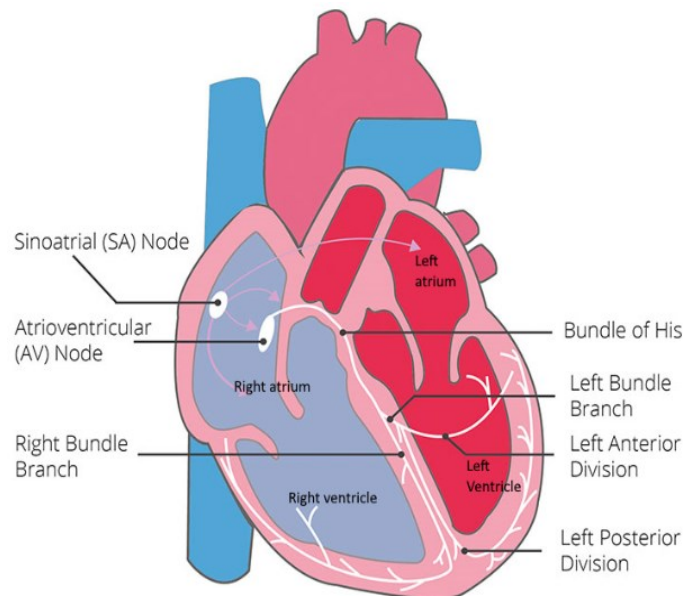


Figure 5 Diagram of the heart[8]

2.3.2 Polarization and Repolarization

Cardiac cells can undergo rapid depolarization and repolarization. Depolarization triggers cardiac cells to contract, while repolarization prepares them for the next electrical impulse. In ECG terms, depolarization corresponds to a positive deflection on the left precordial, whereas repolarization typically results in a negative deflection on them.

2.3.3 Critical Role of the SA and AV Nodes

The SA node regulates the heart's rhythm and can adjust its rate based on various signals. The AV node ensures the atria fully contract regulating the number of auricular impulses that arrive to the ventricle and their speed, by briefly delaying the impulse.

2.4 Wave morphology

Cardiac activity generates various waveforms, each offering insights into different aspects and components of a heartbeat. All these waveforms are added up together and captured by the leads of an ECG. One waveform that particularly stands out due to its pronounced amplitude is the QRS complex. While each specific lead observed might present subtle variations in the waveform's topology, a consistent general structure is seen across all leads; each lead unveiling a unique facet of the waveform.

It's imperative to delve into the morphology and timings of these waveforms, with the QRS complex being central, to unlock a more profound understanding of the heart's electrical system.

The distinct components of the signal, often called PQRST, are depicted in the following graph:

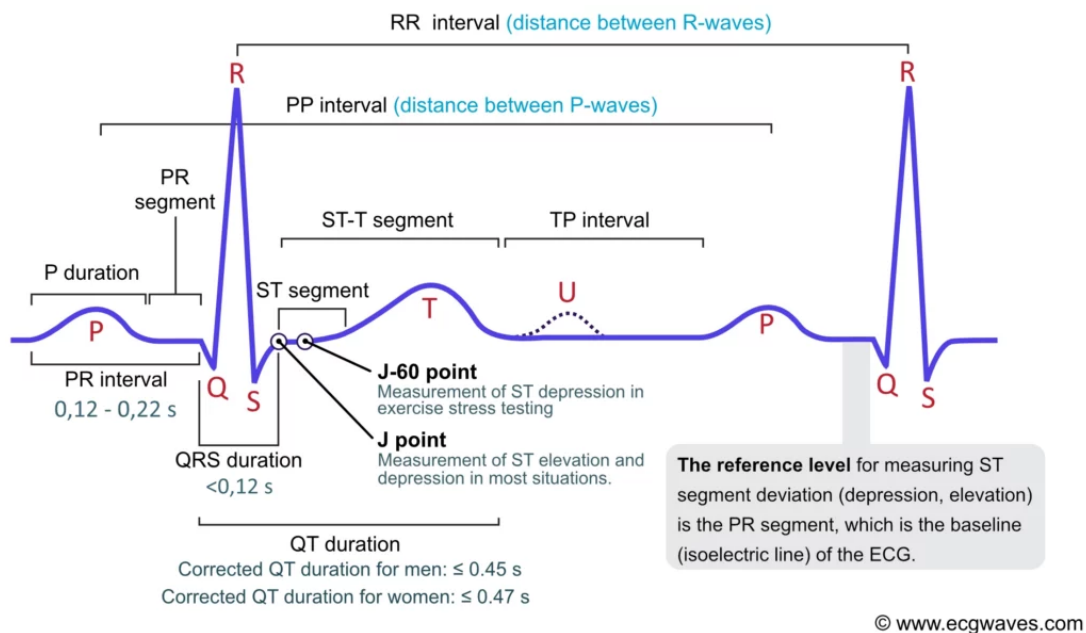


Figure 6 QRS complex [9]

The waveform comprises:

- **P segment:** Depolarization of the atria.
- **QRS complex:** Representing the ventricles' depolarization (contraction).
Within the QRS complex:
 - **Q:** Contraction of the interventricular septum (just before its curvature).
 - **R:** Contraction of the main bulk of the ventricles.
 - **S:** Final contraction phase at the ventricles' apex.

Atrial repolarization (relaxation) also occurs in parallel, but its signal is notably subdued and can be overshadowed.

- **T wave:** Indicates the repolarization (relaxation) of the ventricles.

2.5 The Role of the Nervous System

The heart's electrical activity is not just a product of its intrinsic cells and pathways but is also significantly influenced by the nervous system. This interplay between the heart and the nervous system ensures that the heart can rapidly adjust its function in response to the body's varying needs.

2.5.1 Autonomic Nervous System (ANS)

The Autonomic Nervous System (ANS) plays a vital role in modulating heart rate, the strength of cardiac contraction, and even the conduction speed of electrical impulses through the heart[10]. The ANS is divided into two main branches: the sympathetic and the parasympathetic nervous systems.

Sympathetic Nervous System: The sympathetic branch prepares the body for physical activity. When activated, it can increase heart rate, elevate the force of cardiac contraction, and enhance the conduction velocity of electrical impulses.

The release of neurotransmitters like norepinephrine from sympathetic nerve endings acts on the heart's beta receptors to produce these effects.

Parasympathetic Nervous System: Commonly referred to as the "rest and digest" system, the parasympathetic branch primarily acts to conserve energy and restore calm. Its activation usually reduces heart rate, decreases the force of contraction, and slows the conduction speed of impulses, particularly through the AV node. Acetylcholine is the main neurotransmitter here, acting on the heart's muscarinic receptors.

2.5.2 The SA and AV Nodes in the Context of the Nervous System

The SA node, as the heart's primary pacemaker, has receptors for both sympathetic and parasympathetic neurotransmitters. This means that its rate of firing (and thus the overall heart rate) can be rapidly adjusted based on the balance of sympathetic and parasympathetic stimulation. For instance, during periods of stress or exercise, sympathetic activity dominates, increasing the SA node's firing rate. In contrast, during times of rest or relaxation, parasympathetic influence prevails, reducing the heart rate.[10]

The AV node, too, is influenced by the ANS, primarily by the parasympathetic system. Strong parasympathetic stimulation can slow conduction through the AV node so much that some impulses from the atria may not be transmitted to the ventricles, leading to a phenomenon known as AV block.

2.5.3 Integration with Cardiac Cycle

In the context of the cardiac cycle, the influence of the nervous system is profound. When you consider the sequential depolarization and repolarization of cardiac cells, it's essential to realize that these processes are not just passive events. They are modulated by the nervous system to suit the physiological demands of the body. Whether it's during the excitement of a sprint, the calm of meditation, or the

stress of an emotional event, the heart's electrical activity, as evidenced on the ECG, offers clues about the prevailing autonomic influence.

2.6 ECG stress detection

As for its application in determining stress, the electrocardiogram (ECG) has proven to be a robust tool. Stress can influence several physiological parameters, and the heart is no exception. Stress-induced cardiac responses can manifest in various ways on the ECG, and understanding these patterns is paramount for timely intervention and management.

One of the primary reasons the ECG signal serves as an effective indicator is its ability to encapsulate the combined influences of both the sympathetic and parasympathetic nervous systems[11], which control stress.

Specifically, heart rate variability (HRV), which is derived from the ECG signal, represents the cardiac vagal tone. This tone mirrors the role of the parasympathetic nervous system in regulating the heart, making it valuable for psychophysiological studies.

Diving a bit deeper, HRV's low frequency (LF) variations in heart rate, ranging from 0.01 to 0.08 Hz, are shaped by both sympathetic and parasympathetic activities. On the other hand, the high frequency (HF) variations, spanning 0.15 to 0.5 Hz, predominantly result from parasympathetic activity. The ratio of these frequencies is called the autonomic balance (AB) and is often used in studies as a separate feature when studying the signal as a power spectrum[12].

These HRV components offer a lens to observe the sympathetic and parasympathetic nervous systems. However, the precise physiological mechanisms that underpin them remain somewhat elusive and are subjects of ongoing discussion [13].

A lot of information can be obtained by observing the shape form of the signal, as stress can produce both atrial and ventricular arrhythmias, that can be distinguished over a normal waveform, which researchers[14], [15] have found that includes:

- **A Shortened R-R interval** implies that the heart is beating at a faster pace (tachycardia).
- **A Shortened PR interval** can suggest that the atrial signal is being transmitted to the ventricles faster than usual.
- **An Augmented T wave amplitude** can sometimes be seen in hyperkalaemia, but can also be an indicator of increased sympathetic nervous system activity.
- **A Shortened QT interval** is regulated by cardiac sympathetic innervation.
- **An ST depression**, it occurs when the J point is displaced below baseline. Studies have shown mental stress causes transient myocardial ischaemia, which can be identified by the ST depression.[16]

However, it is important to consider that these features can't solely be used as a stress indicator, as other factors might also affect the topology. For example, external factors such as caffeine intake, medications, electrolyte imbalances, and other underlying conditions can mimic or compound the ECG changes brought about by stress.

3

State of the art

3.1 Problem statement

The goal of our project is to create a real-time stress detector that is non-intrusive. This allows users to monitor of their stress levels in various situations, helping them record and potentially mitigate the negative effects of stress. To achieve this, we need to analyse the complexities and challenges associated with the development and implementation of such a device.

3.2 Related works

The concept of monitoring stress and emotions in real-time through ECGs is not new and has been studied by many researchers over the past decades. The following subsections delve into the prior efforts and breakthroughs in the realms of stress detection and emotion recognition. These explorations shed light on the methodologies employed, the challenges encountered, and the improvements achieved, setting the context for our present work. While our main objective is stress detection, we've also investigated emotion recognition. This decision stems from the close correlation between stress and emotions, and the belief that methods used in emotion recognition most probably also apply to stress detection.

3.2.1 Stress detection

The quest to decode stress from ECG signals has been attempted with a varied set of approaches, ranging from traditional machine learning techniques to cutting-edge deep learning methodologies. Here, we present a concise overview of some most influential studies that have shaped the field:

- **SmartCar: Detecting Driver Stress**[12]: This study utilized features extracted from ECG data, specifically focusing on Heart Rate (HR) and Sympathovagal Balance (AB). The research employed the K-Nearest Neighbors Algorithm to detect stress in drivers.
- **A Transformer Architecture for Stress Detection from ECG**[1]. Raw ECG data was the primary focus of this research. The study applied a combination of Convolutional Neural Network (CNN) and Transformer architecture to detect stress, providing a baseline for transformer-based networks in this domain.
- **Contrastive Self-Supervised Learning for Stress Detection from ECG Data**[17] This paper introduced an innovative approach by focusing on Self-Supervised Learning (SSL) for ECG-based stress assessment. The study implemented a contrastive SSL model based on the SimCLR framework, showing its efficacy in stress detection.
- **Machine Learning for Stress Detection from ECG Signals in Automobile Drivers**[18] This research took a comprehensive approach by extracting features from ECG signals. The study applied various machine learning algorithms, achieving notable accuracy in detecting different stress levels in automobile drivers.
- **Real-Time Psychological Stress Detection According to ECG Using Deep Learning**[19] This study combined advanced neural network architectures. It utilized Convolutional Neural Networks (CNN) and Bi-directional Long Short-

Term Memory (BiLSTM) networks with extracted ECG features to detect stress in real-time.

- **Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection**[3] This study is the introductory paper for the WESAD dataset, and delves into various machine learning techniques to develop a baseline for the dataset.
- **Emotion Assessment Using Feature Fusion and Decision Fusion Classification Based on Physiological Data: Are We There Yet?**[2] This study centred on emotion recognition based on physiological data classification. It employed Supervised Learning (SL), Decision Fusion (DF), and Feature Fusion (FF) techniques using multimodal physiological data, providing a systematic analysis across multiple datasets.
- **Detecting stress through 2D ECG images using pretrained models, transfer learning and model compression techniques**[5] This study transformed 1D ECG data into 2D ECG images to represent stress states, eliminating the need for feature extraction. The research utilized transfer learning for enhanced results and applied model compression techniques, such as quantization, to reduce computational size without significantly compromising performance.
- **Stress Detection ECG**[4] This last reference isn't a published paper, but a GitHub repo. Regardless, it achieves great performance by applying a CNN to a set of carefully crafted features.

Table 1 Comparison of previous stress detection models

Paper	Dataset	Method	Modality	Accuracy [%]	F ₁ [%]
[17]	RML	Contrastive SSL	ECG	73.8	-

[18]	physionet	J48	Wave feat.	76.47	-
[19]	Own	BiLSTM + CNN	HR feat.	86.65	-
[12]	Own	kNN	HR feat.	88.6	-
[3]	WESAD	LDA	HR feat.	85.4	81.3
[2]		QDA	ECG	85.75	-
[5]		SSL + image	Learned feat	90.62	-
[1]		TF + CNN	ECG	91.1	83.3
[4]		CNN	HR feat.	92.0	81.8

3.2.2 Emotion recognition

Emotion recognition, while a distinct field, shares many parallels with stress detection. The underlying premise is that our physiological responses, captured through ECGs, can offer a window into our emotional states. This section reviews key studies that have made significant strides in emotion recognition using ECGs:

- **ECG Pattern Analysis for Emotion Detection**[20]

This study conducts a thorough analysis of ECG signals for emotion detection, using empirical decomposition to dynamically identify emotion patterns. The research highlights the potential of ECG morphology in determining emotional valence and differentiating arousal levels. It provides a very in-depth description of how an ECG wave works.

- **Transformer-Based Self-Supervised Learning for Emotion Recognition**[21]

His paper employs a Transformer-based model for ECG-based emotion recognition. Using attention mechanisms, the model builds contextualized signal representations. The approach leverages self-supervised learning on unlabelled ECG datasets and fine-tunes on the AMIGOS dataset for emotion recognition.

3.3 ECG as a Data Source for Stress Detection

From the presented literature, two primary methods to use the ECG data stand out as the most used. These are the wave morphology and Heart Rate Variability (HRV).

Wave morphology pertains to the shape, structure, and various intervals of the ECG waveform, which can reveal subtle changes in the heart's electrical activity due to stressors.

On the other hand, HRV measures the variation in time between consecutive heartbeats. It's a widely accepted metric for autonomic nervous system activity, and changes in HRV are directly related to the body's stress response.

Both methods provide unique insights and have been instrumental in the advancements in the field of real-time stress detection using ECG data.

3.3.1 Heart Rate Variability (HRV)

HRV is a compelling technique but requires the individual to be studied in the long term, as the patterns perceived are very low frequency (0.01 to 0.08 Hz)[22]. To reliably estimate the power density of a low-frequency signal, especially one as slow as 0.01 Hz, one would need to sample over multiple periods of the signal. Capturing for a minimum of 2 periods (200 seconds) would give a decent resolution, capturing for longer durations (e.g., 500-1000 seconds) would yield a more reliable power density estimation for the signal.

This makes it inviable for real-time or short-term stress detection, particularly in situations where quick intervention or actuation is necessary. For instance, if one were trying to monitor stress in inside a vehicle, waiting for several minutes to even half an hour to get a reliable first reading would be far from ideal.

3.3.2 Wave Morphology Analysis

Wave morphology presents both challenges and opportunities. Analysing wave morphology requires precise identification, which can be affected by noise, artefacts, or other external factors.

Another significant challenge is the considerably higher bandwidth requirements compared to HRV analysis. Processing and working with wave morphology data is much more resource-intensive, making it more expensive and demanding in terms of computational power and data transmission.

Moreover, individual variations in ECG patterns, influenced by aspects like age, gender, or health conditions, can complicate standardization.

However, if accurate algorithms or models for wave morphology analysis are developed, it may serve as a lower latency higher accuracy method to detect stress in real-time scenarios compared to HRV, therefore it has been chosen as the selected technique.

3.4 Selected leads

To reduce the device's intrusiveness, it's essential to minimize the number of required leads or sensors. Traditionally, ECGs necessitate multiple leads to obtain the necessary data. An increase in the number of leads can render a device cumbersome and potentially discourage users from wearing it. This can be mitigated by using radar technology, but using multiple of this sensors might prove too costly.

For this reason, we have chosen to use a single-lead ECG. The III projection has been selected to optimize the algorithm's performance, as it most clearly defines

the R wave and T waves. This aids in analysing wave morphology, as detailed in the medical section of the project.

One of the primary applications of this project is stress detection within a car cabin using radar. In this context, the optimal placement for the sensor would be on the car's pillar. This placement would produce a projection similar to the III, with a horizontal component. As a result, the selected wave morphology would strongly correlate.

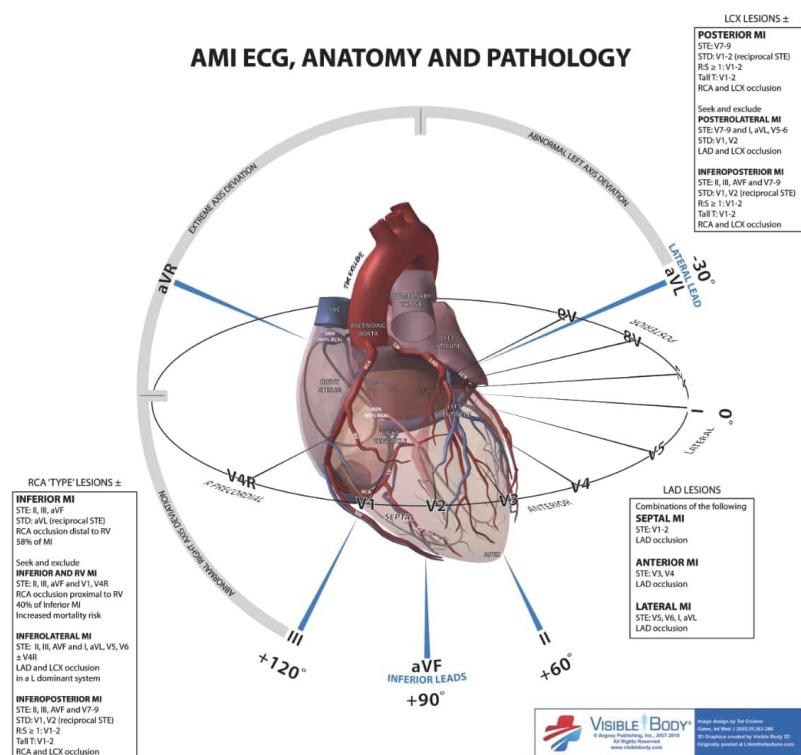


Figure 7 Lead position in respect to the heart.

3.5 Selected technologies

Looking at the preceding works and understanding the complexities involved, we have selected a set of technologies that will be instrumental in achieving our goals for real-time stress detection. First, the use of a CNN seems mandatory, as it is used by the vast majority of well performing methods as a mean provide pattern detection and to filter between noise and signal. In the case of using raw time series, it has the added effect of providing short-term pattern recognition capabilities.

In addition, the use of a long-term pattern recognition algorithm, seems to have a significant effect. Long Short-Term Memory (LSTM), and Transformers, have both shown promise in this regard. The use of transformers is particularly enticing given its proven effectiveness in capturing long-term dependencies within sequences, as seen in natural language processing applications.

These more advanced methods are harder to train, and require significant amounts of data, which might be hard to obtain. Therefore, the use of Self-supervised Learning warrants itself as an ideal solution to overcome this obstacle.

As both of these technologies are quite recent advancements in Deep Learning, they will now be discussed in detail.

3.5.1 Transformers

As previously stated, transformers have the advantage of being able to look at the entire sequence of data, rather than just short windows. By doing so, we aim to filter some noise and artefacts that can interfere with accurate wave morphology identification. Furthermore, the ability of transformers to attend to different parts of the sequence, based on the importance of those parts to the task at hand, may help in highlighting significant features of the ECG that indicate stress.

It is important to note that while transformers can analyse longer sequences, one shouldn't input an excessively large sequence. The number of parameters grows quadratically with the model's dimension[23]. Therefore, the model shouldn't have an excessively high sampling rate, and its timeframe shouldn't be too long if we don't want to excessively increase the learning time and computational hardware required.

ECGs are usually sampled at very high frequencies (~1kHz) but are usually down sampled depending on their use. In the case of HRV analysis, it has been shown that the resolution can be lowered up to 250 Hz with negligible loss in accuracy[24]. Meanwhile, for the study of the ECG morphology, it's been shown that the delineation remains essentially unaltered for sampling frequencies higher than 125 Hz [25] using common methods.

f_s (H z)	P onset	P end	QRS onset	QRS end	T onset	T end
500	0.0 ± 0.1	0.0 ± 0.1	0.1 ± 2.4	0.0 ± 1.0	0.0 ± 0.1	0.0 ± 0.1
250	-0.1 ± 2.9	0.0 ± 0.4	0.1 ± 2.5	0.1 ± 1.0	0.0 ± 0.3	0.0 ± 0.3
125	0.0 ± 5.4	0.2 ± 4.0	-0.1 ± 6.7	0.1 ± 7.7	0.0 ± 4.7	0.0 ± 3.2
62.5	-0.7 ± 14.4	0.1 ± 11.5	-3.6 ± 16.3	3.4 ± 17.5	-0.4 ± 8.9	0.4 ± 7.4

Figure 8 Delineation errors regarding the automatic detections at 1000 Hz of the ptbdb (in ms) [25]

3.5.2 Self-Supervised Learning

Self-supervised learning, on the other hand, can address the challenge of individual variations in ECG patterns derived from external factors such as age or gender, and help to create a more robust model.

By not relying on labelled data, but instead generating its own pseudo-labels from the data itself, self-supervised learning enables us to use far bigger datasets than what might be available if one depended on manually labelled datasets.

This vast amount of data, combined with the ability of self-supervised models to discern patterns, can help the system learn more generalized features and

representations from the ECG signals. This increases its accuracy and reliability without having to fine-tune the model for the individual user.

For the election of the tasks, it is critical to use a generalist task instead of a specialized one for the field, such as detecting features in the signal. By doing so, we are sure that no bias is introduced by us while developing the model.

4

Methodology

To overcome the challenges of real-time, non-intrusive stress detection, we propose a solution that brings together advanced computational methods and optimized hardware components. Our solution is a two-stage process:

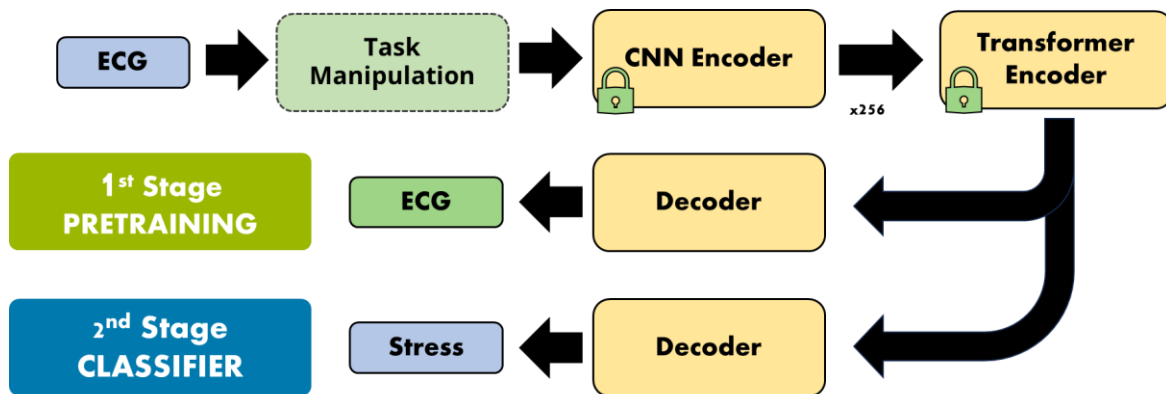


Figure 9 General architecture of the model

The first stage is designed for pre-training a self-supervised model using single-lead ECG data, with the architecture encompassing a CNN Encoder, a Transformer Encoder, and a Decoder. This pre-trained model is specifically tuned for stress detection tasks in the second stage, where the encoder layers are repurposed for a stress detection classifier. Each of these stages is comprehensively discussed in the following chapters to provide in-depth insights into their operational mechanics, data handling, and functional components.

4.1 Pre-training via Self-Supervised Learning

In the initial phase, a self-supervised learning model is employed to identify patterns and features autonomously within the data. This enables the model to establish a robust foundational representation of the ECG (Electrocardiogram) data, providing a significant advantage for the subsequent task of stress classification. This section explores the model's architecture, implementation specifics, and validation procedures.

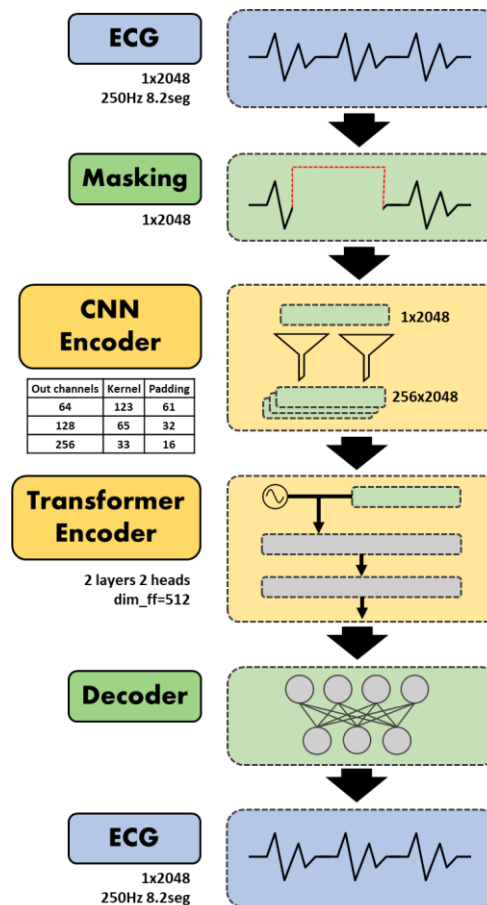


Figure 10 Self-supervised task architecture

4.1.1 SSL Task

For the purpose of representation learning, we opt for a pretext task-based approach over a contrastive learning-based approach. While there have been efforts to apply contrastive learning to univariate time-series representation [26], task-based methods remain the most prevalent. Within this area, the technique of masked reconstruction [21], [27], [28] stands out as the most commonly utilized approach, and was chosen for this project.

The task to be carried out is to reconstruct an ECG signal from an occluded view. The occlusion is guided by a binary mask M which decides what parts of the sequence should be replaced.

There are several algorithms to create the mask, but in general terms it was observed that the masked sequence shouldn't be too large. Instead, for the model to successfully learn, it was much better to create a small mask, of similar width to the individual PQRST features of the waveform. The final masking strategy consisted of masks of fixed width of 39 samples (0.16 seconds).

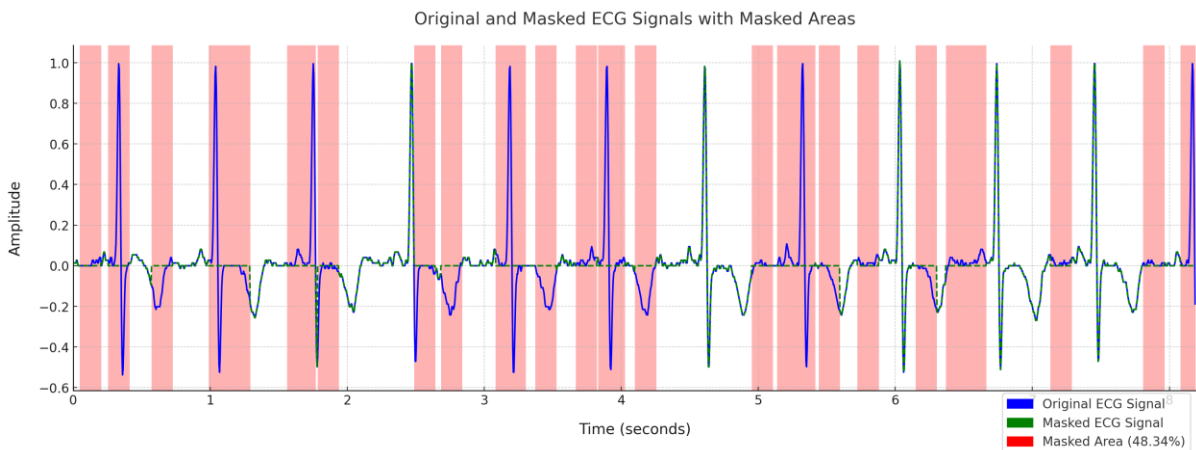


Figure 11 ECG masking strategy

The masks are generated randomly with a 1.66% likelihood and can overlap. Overall, nearly half (48%) of the original ECG waveform is obscured by these masks. This process is quite reminiscent to the one used for masking in a similar paper[21].

The choice of masking value also plays a critical role. Initially, we considered using a high value for masking, given that ECG signals are centred around zero and the values of the signal are small. The rationale was to make the masked value easily distinguishable, thereby maximizing the loss function. However, we discovered that this approach led to training instability, causing abrupt changes in the initial stages of the neural network's learning process.

We also tried to do multitask training, by alternating between this task and a reordering task. As they are very different tasks, we left each task running for many generations until the test function started to deteriorate. But ultimately, we found the learning from one task was not transferable to the next, and all the progress done by one was reversed by the other.

4.1.2 Loss function

Drawing upon the insights gained from the reviewed literature[21], [27], several options for the loss function were implemented for this project. The options included the Mean Square Error (MSE) loss, which has been commonly used, and the Mean Absolute Error (MAE), which has been found to be robust against outliers[29].

It was finally decided to use MSE, as there were no real benefits seen when using MAE on the initial testing.

$$L_{\text{MSE}} = \frac{1}{|M|} \sum_{(t,i) \in M} (\hat{x}(t,i) - x(t,i))^2$$

Where $|M|$ is the total number of masked points, $\hat{x}(t,i)$ is the predicted value at time t and position i , and $x(t,i)$ is the true value. As seen, a significant aspect of

our chosen loss function is that it focuses solely on the subset of the masked sections of the ECG signal for loss computation.

The primary rationale behind this approach is efficiency. By focusing the loss calculation on the masked areas, the model is directed to learn the most from the parts of the data that are intentionally obscured.

4.1.3 Encoder architecture

To create the latent space where the intrinsic ECG data is stored, an Encoder-Decoder architecture was used. To create the encoder, the state-of-the-art technologies for sequential data processing were used, a CNN encoder followed by a transformer encoder.

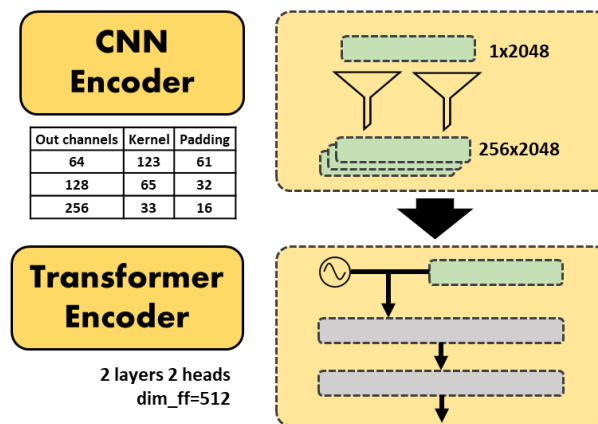


Figure 12 Encoder architecture

This architecture has been selected in recent works [1], [21], [27] over other techniques based in LSTM[19], RNN or Autoencoders[30].

4.1.4 CNN Encoder

The CNN should help to highlight specific patterns inside a heartbeat, while the transformer should capture more global dependencies between different parts of the ECG sequence.

In order for the CNN to work as intended, the receptive field must be of similar length to one heartbeat[21], therefore the CNNs kernel, stride, and padding should be calculated accordingly.

In this case, the desired network specifications for the CNN were as follows:

Layer	Input channels	Output channels	Kernel	Padding	Stride
1	1	64	123	61	1
2	64	128	65	32	1
3	128	256	33	16	1

Figure 13 CNN network specifications

In this case, using the kernel sizes creates a receptive field of 0.8 seconds (75bpm), which is a bit lower than the resting heart rate of a healthy adult (65bpm). It was intentionally chosen as to hopefully improve performance for stressed heart rates suffering with tachycardia, although further experimentation is needed to corroborate its usefulness.

The padding was picked as to not affect the dimensionality of the sequence. For each layer, the kernel always starts from the first sample and traverses the whole sequence with stride 1.

As for the output channels, they were selected considering that they would be the d_{model} for the transformer encoder, which in general is recommended to be used with tokens with numbers of features of that order of magnitude[23], [31].

4.1.5 Transformer Encoder

The output of the CNN is passed to a transformer. Each of the sample n of the sequence is an element of the input vector, with each of the elements consisting of a 1×256 feature vector.

The positional encoding used was a fixed positional encoding as in the Attention all you need paper[23], other positional encoding methods such as learnable

positional encoding were tested, but no additional advantages were seen in performance.

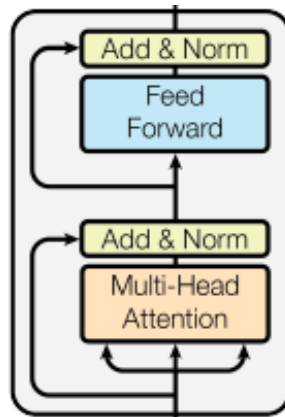


Figure 14 Transformer encoder block

The transformer block consists of 2 layers of multi-head attention with 2 heads each. A value of 0.1 for the dropout was picked as a balance between overfitting and performance. Layer and Batch normalization were studied, Batch normalization was finally selected as it led to much better performance and faster training. A Gaussian Error Linear Unit (GeLU) activation function was chosen over a Rectified Linear Unit (ReLU) activation function, as it lead to better results in BERT [28], although further research is needed to corroborate if it's a better option for this task, as BERT is a NLP (natural language processing) model.

The attention matrix of each block was saved for debugging and interpretability without separating the heads.

4.1.6 Decoder

The decoder consists of a fully connected network of two layers, where the dimensionality of the encoder's output is gradually reduced to a univariate ECG.

The first layer reduces the dimensionality from 256 to 128 and the second layer from 128 to 1. Between the two, a ReLU activation function is placed to introduce some non-linearity.

4.1.7 Training

For the initial phase of training, we used the ALSEDAS dataset[32]. The data was pre-processed to remove noise and normalized to ensure zero-mean and unit variance.

The model was trained using a batch size of 64 and was optimized using the Adam optimizer with a learning rate of 0.001. We also applied gradient clipping to prevent exploding gradients. Training was performed for 100 epochs; at each epoch, a checkpoint was saved with the parameters of the network at that point.

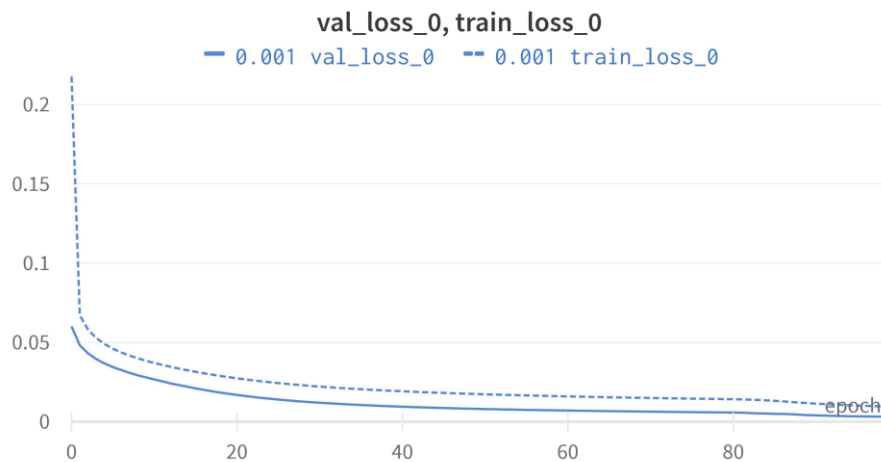


Figure 15 SSL Training curves

During training, validation loss is curiously lower than training loss, which can be seen as highly unusual, but can be partially attributed to the dropout applied. The validation loss slope was smaller than the training, so the gap between the two

decreased over time until both stagnated, becoming both rather horizontal. At the 100 epoch, the validation loss is just 0.00311, which is negligible and produces an almost perfect reconstruction.

4.2 Classifier

In the second stage, the focus shifts from pre-training to fine-tuning the model for stress detection. We utilize the weights of the encoders pre-trained in the first stage as a starting point for the stress classifier. This fine-tuning aims to adapt the pretrained model to the stress detection task, making it specialized and more accurate.

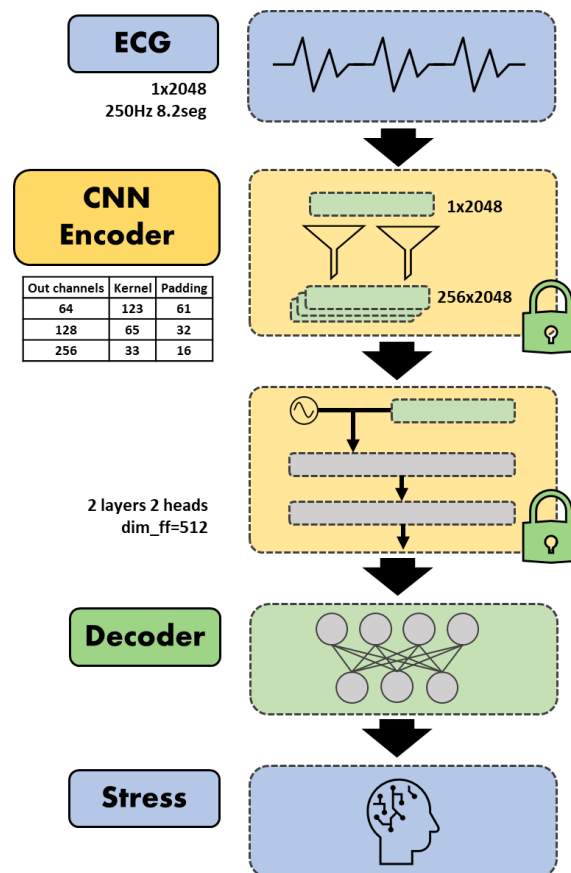


Figure 16 Classifier architecture

4.2.1 Loss function

For the stress detection task, which is essentially a binary classification problem (stress vs. no-stress), we opt for the Binary Cross-Entropy with Logits Loss, also known as Sigmoid Cross-Entropy loss. This loss function is well-suited for binary classification, especially those that may involve imbalanced datasets.

Mathematically, the loss is calculated as:

$$L_{BCE}(y, x) = y \cdot \log(\sigma(x)) + (1 - y) \cdot \log(1 - \sigma(x))$$

Where $\sigma(x)$ is the sigmoid function applied to the output x of the model and y is the ground-truth label, and N is the total number of samples.

4.2.2 Encoder

The encoder is loaded from a checkpoint of the previous stage, which loads the weights as achieved in pre-training. To give the model a bit more flexibility, it was decided to unlock the full encoder.

4.2.3 Decoder

The decoder consists of a gradual dimensionality reduction over two linear layers to a one-dimensional prediction. Between the two layers, we apply batch normalization and a ReLU activation function. After the last layer, a sigmoid function is applied to convert the output to a logit.

The first layer reduces the dimensionality of the d_{model} from 256 to 1, and the second layer reduces the dimensionality across time, and gives a single stress value for the entire sequence.

It is crucial to apply batch normalization between the two linear layers for several reasons. Firstly, it helps stabilize the training process by normalizing the

activations of the neurons in the hidden layer. This can result in a much quicker and more stable training. Secondly, batch normalization acts as regularization, mitigating the risk of overfitting the model to the training data. This is particularly important given the complexity of stress detection tasks, which often require capturing subtle patterns in the data that may easily lead to overfitting.

4.2.4 Training

For the fine-tuning of our stress detection classifier, we implemented a rigorous training regimen to ensure the best performance.

1. **Optimizer Choice:** We leveraged the Adam optimizer for training our model. Adam is a widely used optimization algorithm that can handle sparse gradients on noisy problems. Given its adaptive, it is regularly recommended by other similar papers over other methods as SGD.
2. **Parameter Sweeps:** To ensure optimal performance, we executed a sweep for certain hyperparameters. Notably, the sweeps focused on the beta1 learning rate and the weight decay parameters. Sweeping allows us to test multiple hyperparameter combinations, ultimately helping us to identify the combination that yields the best results.
3. **Bayesian Optimization:** To streamline our hyperparameter tuning, we employed Bayesian optimization, a method that uses a Gaussian Process to model the objective function and quickly identify optimal hyperparameters. We integrated this with the automatic Bayesian optimization feature of Weights & Biases (WandB), a platform for advanced machine learning experiment tracking. This combined approach not only expedited the tuning process but also provided valuable insights and visualizations.

Through this meticulous training approach, we aimed to produce a model that is adept at accurately detecting stress in diverse datasets.

Various sweeps were conducted with slightly different model architectures. In the following image an example sweep, with the WESAD dataset parted into 8 seconds (2048 samples) sections, is shown. To better illustrate the performance of each run, and its parameters, a parallel coordinates plot was created with the performance being measured with the model's validation set F1 score.

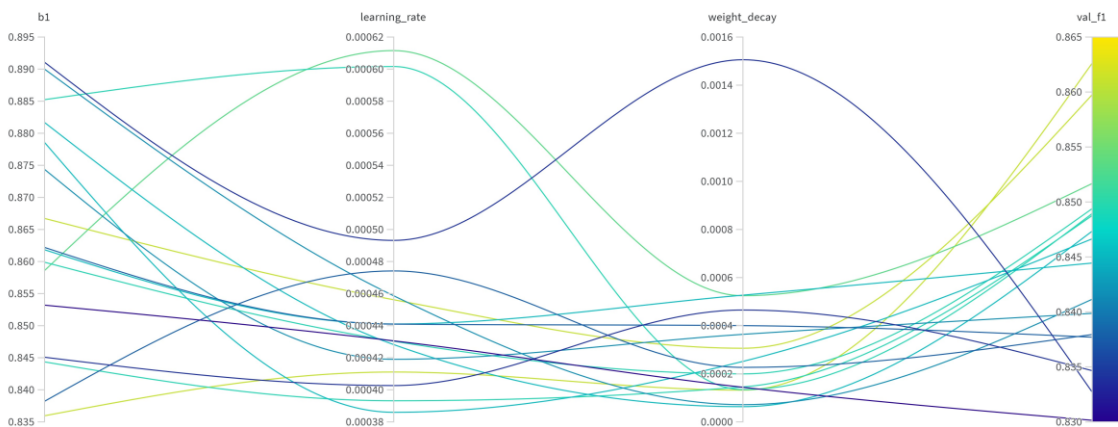


Figure 17 Classification hyperparameter sweep

The parallel coordinates plot only shows a subset of the top performing runs, as to not clog up the screen. The *learning_rate* and *weight_decay* hyperparameters are very sensitive, and the band that produces optimal results is very narrow.

4.3 Data acquisition

The acquisition of a high-quality dataset is crucial for the success of this project. The choice of dataset plays a crucial role in both training and validation of the model and can greatly impact the performance of the stress detection system.

Two different datasets will be used, one for each of the training stages of the model: the self-supervised stage and the classifier stage.

The two datasets must contain the same type of signal, as not to interfere with each other. Therefore, they must include a III lead ECG signal with a sampling frequency higher than 250Hz, as for them to be downsampled.

Publicly available datasets were chosen to ensure the reproducibility of the project and enable comparison with other studies.

4.3.1 Self-Supervised Dataset

For the self-supervised dataset, the primary focus is on the waveform, and no additional labels such as stress indicators are necessary. Therefore, this dataset can be more easily sourced. Although there are thousands of ECG datasets available online, a vast majority of them correspond to patients suffering from heart conditions, making them not representative of the general population.

After reviewing multiple datasets, it was finally decided to use the dataset: *A large scale 12-lead electrocardiogram database for arrhythmia study* [32] from now on referred as the *ALSEDAS* for simplicity's sake. It consists of 85,152 ECGs from patients from two Chinese hospitals, each of 10 seconds of duration. The dataset, therefore, contains many patients suffering from heart conditions, but because of the vast amounts of data to work with, we were able to filter these, as well as the ECGs that had significant noise. After the fact, we have 48k samples.

The III lead extracted and down sampled to the desired 250 Hz, which won't introduce any artefacts as it's a multiple of the original signal.

It must be noted that, the Icentia11k dataset [33] was originally chosen for the project, but it was finally scrapped for as the electrode location better resembled a II lead, making it not representative of the waveform types seen for the selected classifier dataset. Despite the discrepancy, performance was only marginally affected. Therefore, indicating that the selected self-supervised task didn't overfit for the lead type.

4.3.2 Classifier Dataset.

The selection of the classifier dataset, in contrast, is essential, as it includes the label used to classify stress. To help compare performance to other methods, it was decided to use a well-established and publicly available dataset. After evaluating various options, the dataset selected for this purpose is the WESAD dataset [34].

This multimodal dataset features physiological and motion data, recorded from both a wrist- and a chest-worn device, with all signals sampled at 700Hz.

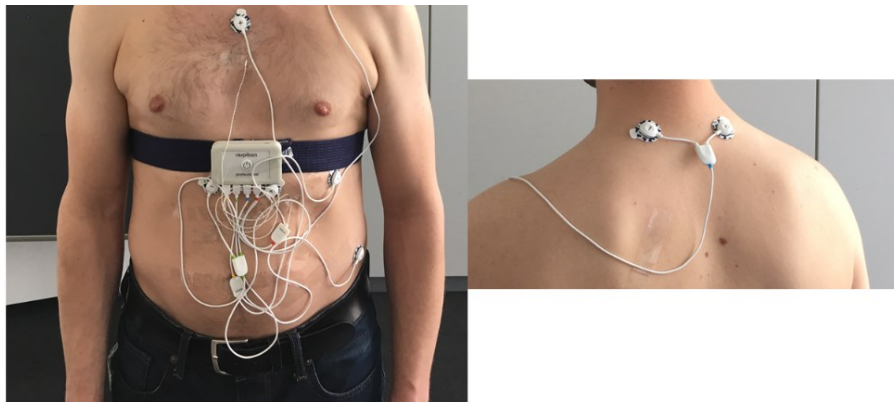


Figure 18 Placement of the RespiBAN and the ECG, EDA,[3]

The ECG used in the dataset is a standard 1-lead ECG with two signal electrodes and a ground electrode. The dataset doesn't specify, but after closely looking at the electrode placement, it is clear that it is a type III ECG.

The dataset includes data from 15 participants, averaging 27.5 ± 2.4 years of age. Of these, 12 are male, and three are female.

To measure stress levels, the study employed questionnaires administered after participants completed each of the following stations:

- Baseline: Participants sat or stood at a table for 20 minutes, with neutral reading material like magazines provided.

- Meditation: A 7-minute meditation session following other tasks.
- Amusement: Viewing humorous video clips for a total duration of 392 seconds (6 minutes and 32 seconds).
- Stress: The Trier Social Stress Test (TSST) was conducted, involving tasks like public speaking and mental arithmetic, lasting for 10 minutes.

To avoid any bias, two different versions of the study were carried out, each with a different order.

Version A



Version B

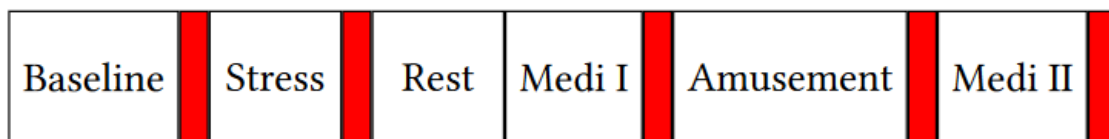


Figure 19 The two different versions of the study protocol

The original dataset featured signals sampled at 700Hz. While a higher sampling rate offers more detailed information, it may also introduce computational challenges, such as increased processing time and memory usage. Therefore, we opted to reduce the sampling rate to 250Hz.

We segmented the ECG data into 8 second chunks to emulate sensor readings for a real-time monitoring system. These 8-second intervals were labelled according to the reported stress doing that activity type.

However, it should be noted that stress responses may vary considerably within a single labelled activity. For example, during the Stress task, patients could experience a range of emotions from anticipation to relief while doing the different tasks. These details might not be fully captured by a single “Stress” label for the entire 10-minute segment but might be captured inside the 8-second intervals, leading to mislabelling.

This limitation is a by-product of our data-labelling methodology, which assigns labels at relatively long intervals to accommodate a broader spectrum of stressors.

Future work could consider employing their data acquisition, employing a more dynamic labelling approach that allows for shorter, more context-sensitive labelling intervals, which may be more relevant to the task at hand.

5

Results

5.1 Quality of the models' predictions

In this section, we delve into a more profound analysis of the models' predictive quality by investigating other evaluation metrics and analysing some individual predictions. It's crucial to understand the behaviour of the models beyond just accuracy and F1 scores to gauge their reliability and robustness in real-world scenarios. Of course, all this metrics have been made with an unseen test set.

5.1.1 Confusion Matrix Analysis

The confusion matrix provides a more detailed view of the prediction results, offering insights into the types of mistakes that the model is making. We analysed the confusion matrix for each model to observe the rate of true positives, true negatives, false positives, and false negatives.

Table 2 Confusion Matrix Analysis

8s-2l full		Predicted	
		No Stress	Stress
Actual	No stress	710	4
	Stress	10	161

8s-2l last		Predicted	
		No Stress	Stress
Actual	No stress	671	17
	Stress	31	166

4s-1l full	Predicted
------------	-----------

4s-1l last	Predicted
------------	-----------

		No Stress	Stress
Actual	No stress	1361	5
	Stress	16	389

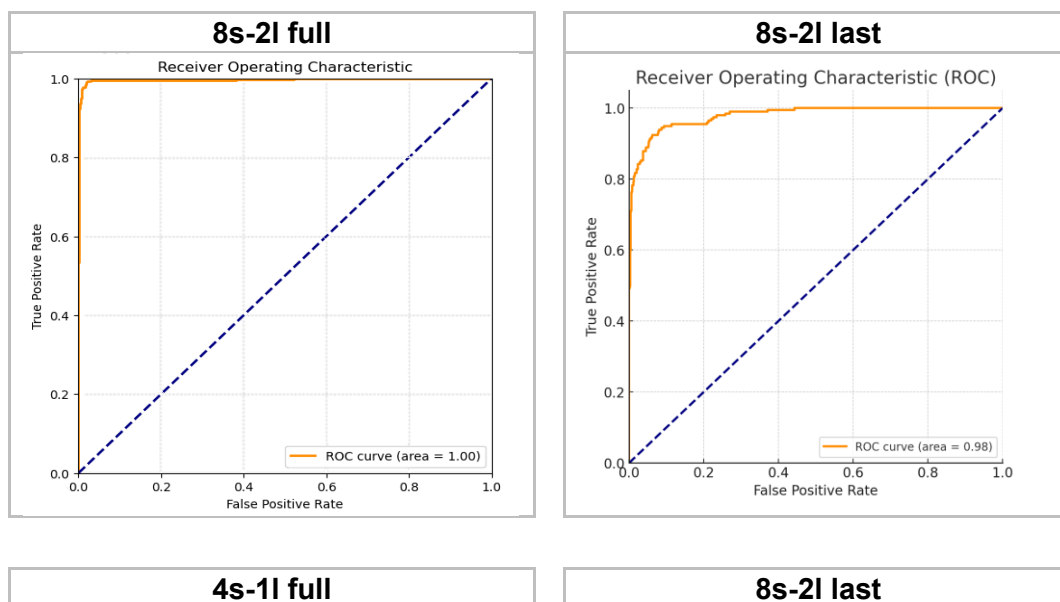
		No Stress	Stress
Actual	No stress	1349	17
	Stress	136	269

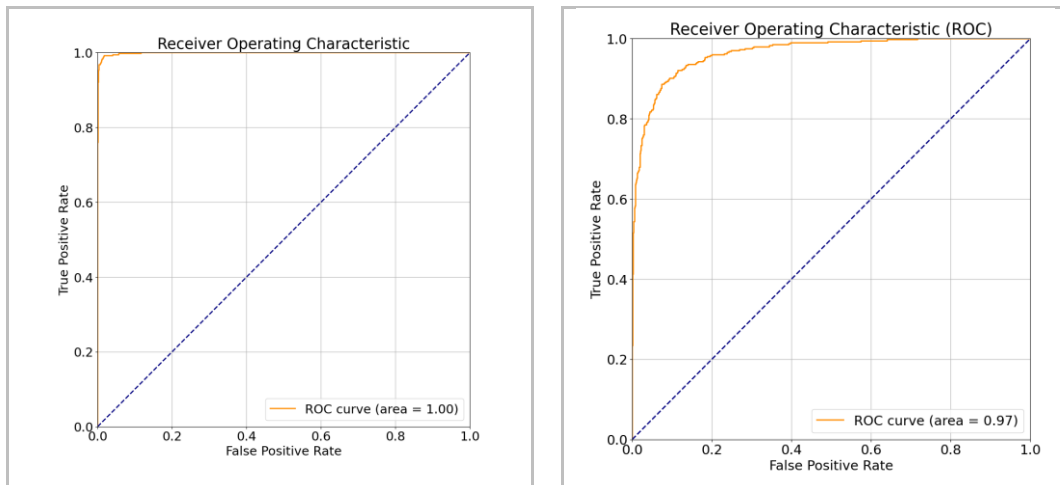
The model displays a strong capability to correctly classify "No Stress" situations, with only a minor portion being misclassified as "Stress". However, when it comes to detecting "Stress" scenarios, there is a noticeable number of instances that the model incorrectly predicts as "No Stress". This suggests that while its specificity is commendable, its sensitivity might require improvement.

5.1.2 Receiver Operating Characteristic (ROC) Curve and Area Under Curve (AUC)

The ROC curve and the associated AUC value are crucial metrics to evaluate the models' ability to distinguish between the two classes over a range of threshold settings. We computed the ROC curves and AUC values for each model.

Table 3 ROC curves





Upon examining the ROC curves, it is evident that the models have demonstrated excellent discriminatory power. Specifically, the near-square shape of the ROC curve is indicative of outstanding performance, as it closely hugs the top-left corner. This shape means that for a broad range of thresholds, the model has a high true positive rate and a low false positive rate.

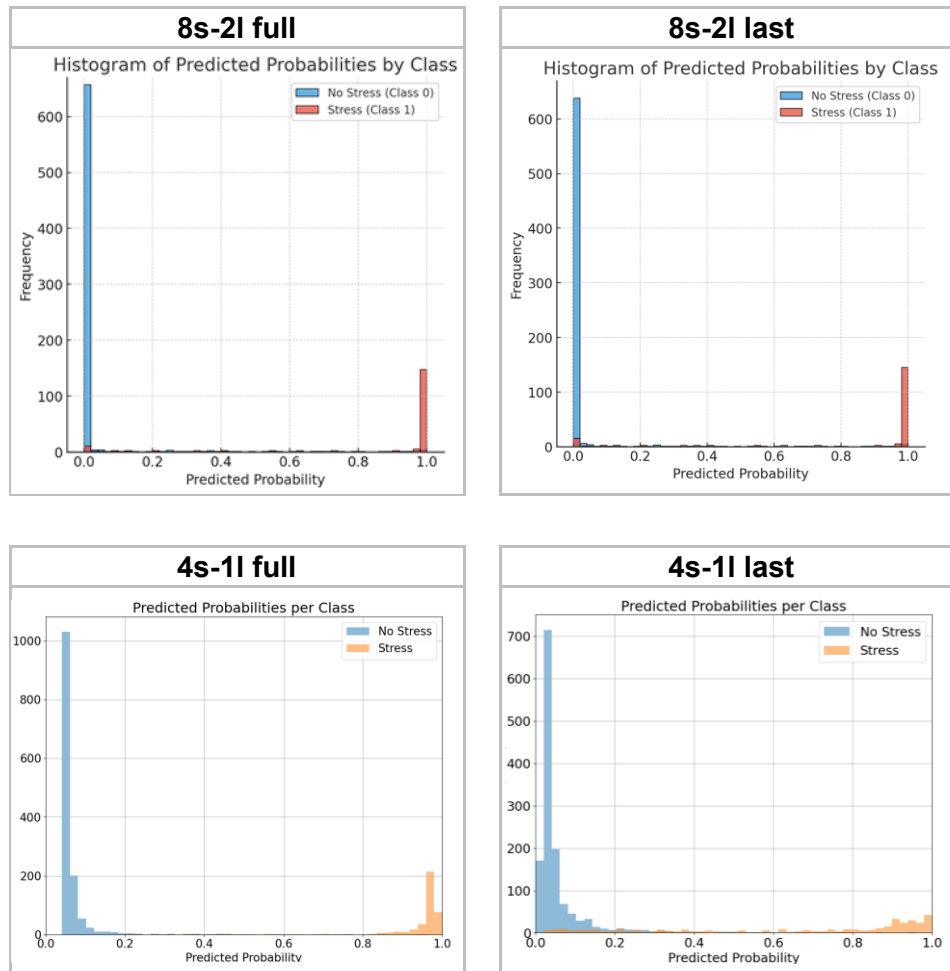
The AUC values reaffirm this observation. High AUC values, approaching 1, signal that the model has a strong ability to distinguish between the positive and negative classes across different decision thresholds. An AUC of 1 represents a perfect classifier, so our models' performance is commendable and suggests they are highly reliable in making predictions.

5.1.3 Histogram analysis

In this section, we present a histogram analysis to further dissect the performance of our models and understand the distribution of prediction scores across the two classes (stress and no-stress). Histograms provide a visual representation of the frequency of prediction scores, helping to gauge the confidence level of the model's predictions and understand any biases.

We plotted histograms of the prediction scores for both classes across different models to observe the distribution of scores.

Table 4 Histogram of the models



We find that the model is well-calibrated, with distinct peaks for the stress and no-stress classes, and minimal overlap between the two distributions. Although the “4s-1l full” predictions never get to 0, which indicate a slight miscalibration.

There are some stress samples which get completely mistaken for no stress, suggesting a possibility of outliers or mislabelled data in the dataset, or perhaps certain physiological markers of stress that were not captured well by the model.



UNIVERSIDAD PONTIFICIA COMILLAS
ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA (ICAI)
MASTER'S IN INDUSTRIAL ENGINEERING

6

Interpretability of the results

While achieving high accuracy and F1 scores is undoubtedly important for the stress detection model, the interpretability of the model is equally crucial, especially in the healthcare and medical sector. Knowing what the model “thinks” helps clinicians and healthcare providers trust the model's decision-making process, which is vital for making informed medical decisions. Additionally, interpretability helps to identify any biases or errors in the model, thus enabling fine-tuning and iterative improvements.

6.1 Saliency maps

Saliency maps are one of the techniques we used to visualize the decision-making process of the stress detection model. The maps highlight which parts of the input data were most essential for the model when making a prediction. In our context, it can be especially useful to understand which physiological or behavioural signals are most indicative of stress according to the model.

To generate saliency maps, we perform backpropagation from the output layer to the input layer, computing the gradient of the output regarding the input. The magnitude of the gradient serves as an indicator of the importance of each feature, which is then represented visually as a saliency map.

6.2 Saliency map of the encoder

The encoder plays a vital role in transforming raw input data into a feature-rich representation, which is utilized by the subsequent layers of the model to make predictions. Visualizing the saliency map for the Convolutional Neural Network (CNN) encoder helps us understand which features were deemed valuable for the pretraining task, and therefore which parts of the waveform are better represented for the Transformer and decoder.

The gradients obtained through this process are indicative of the importance of each feature extracted by the encoder. These gradients are then visualized as a saliency map, which could be colour-coded to reflect the magnitude and direction of the importance each feature has in the decision-making process. To illustrate this, the saliency of the four possible cases of our algorithm is shown, these illustrate a diverse set of the ECGs encountered by the network.

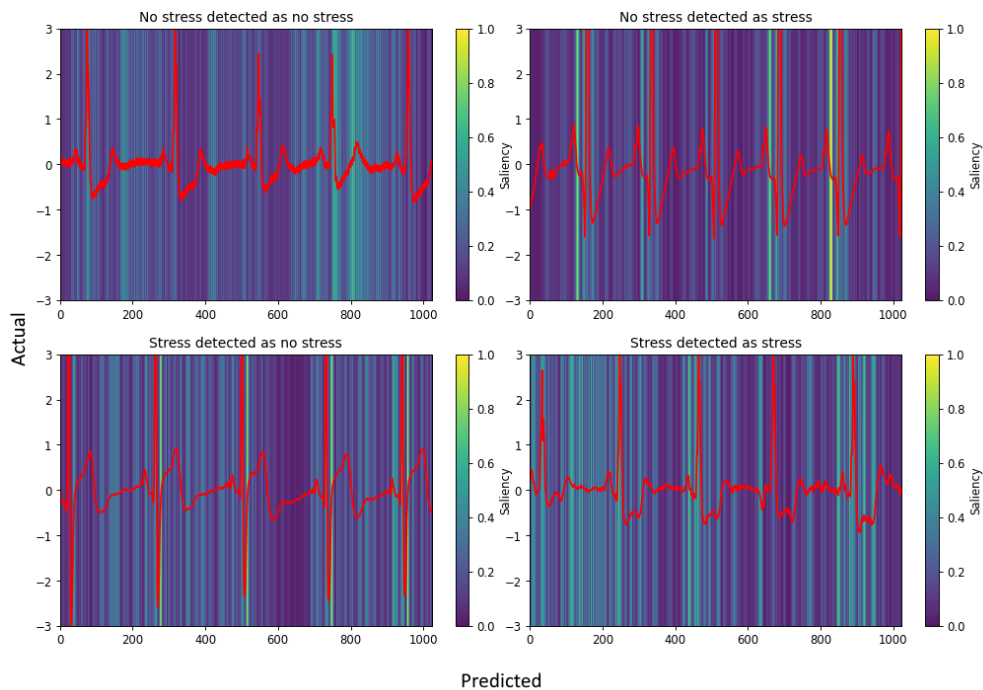


Figure 20 Saliency map of the encoder

The CNN encoder accentuates certain segments of the signal over others. Interpretation may be subjective, yet some insights can be garnered. Given the nature of our pretraining task, the encoder seems to have learned the attributes of a 'baseline ECG,' which is the most prevalent type in the ALSEDAS dataset and closely resembles the non-stressed ECG of the WESAD dataset. Utilizing this baseline, the encoder has honed its ability to detect anomalies in this pattern, highlighted within the saliency maps.

The true negative (top left) illustrates a fairly typical ECG, aside from two heartbeats with lower R amplitude. In this instance, the saliency map appears relatively flat, with the only notable area being the end of the T segment. This observation is beneficial for the downstream task, as a pronounced T wave is indicative of stress. The encoder also focuses on the ST-T segment of the second-to-last heartbeat, which deviates from the rest of the presented beat's waveform—a result of our chosen pretraining task, which emphasizes the ECG intricacies for reconstruction.

In the false positive (top right), emphasis is placed on the PR interval due to an abnormally low reference level (PQ), which typically hovers around 0.

Regarding the false negative (bottom left), attention is directed towards the ST segment, which is noticeably low, exhibiting ST depression with a negative J point—a marker of stress and indicative of a short QT. Despite being captured by the encoder, it wasn't correctly classified.

Lastly, the true positive (bottom right) appears to amalgamate all previously highlighted observations. Though the Heart Rate (HR) is notably low, the waveform exhibits abnormal characteristics typical of stress, and is therefore correctly classified.

Overall, it can be inferred from the saliency maps and corresponding cases that the encoder has developed a good representation of the signal, which correlates well with the downstream task. It must be noted, that although the encoder was

also trained on the downstream task in the case of all layers, unlocked model. The number of epochs, and number of samples it was trained on won't drastically change the overall extracted features, and its performance must be largely attributed to the pretraining task.

6.3 Saliency map of the output

Computing the saliency map from the output is also valuable. This map highlights the time steps that notably influence the classification result, acting as a visual tool to pinpoint the evidence the model utilizes for its prediction.

The saliency map gives 'importance scores' to the time steps of a given time series, demonstrating the extent to which each time step affects the classifier's prediction. Higher values on the map signify a stronger reliance by the model on those specific time steps. It's worth mentioning that we opted for a base saliency map computation over more recent methods like PERT [35].

In the following image, bright red bars signify strong evidence supporting the opposite class, stress, while blue bars indicate evidence against stress. We've chosen to display the performance in one of the four possible scenarios our algorithm could face. It should be noted, however, that there is a significant variability in the segments the network selects to focus on, with some patterns or features not displayed in this image.

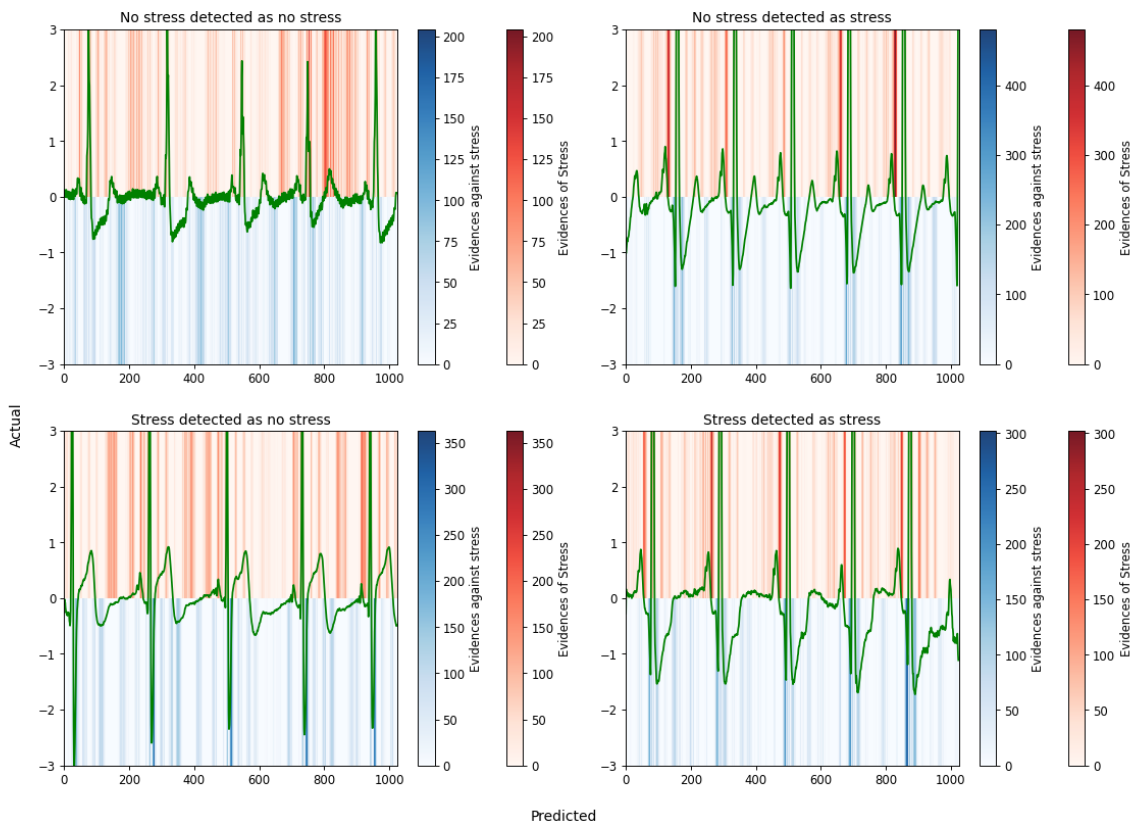


Figure 21 Saliency map of the output

Overall, there is a prevalence of the colour of which the network decided to classify.

The true negative case (top left) seems two, focuses on the T wave. Its short length is seen as evidence against stress, whilst its amplitude on the last heartbeat is a sign of stress. As this pattern is only seen on the last beat, it doesn't have enough weight as to change the overall classification.

The true positive case (bottom right) on the other hand, displays a very faint T wave, but isn't seen as evidence against stress. As probably the type of wave comes from an electrode location where there isn't a pronounced T wave.

6.3.1 Simulated ECGs

To further improve our understanding of the model, synthetic ECGs will now be made to alternate certain features of wave. There are mathematical models to create synthetic ECG waves [36], but they are hard to control as they have many non-linear parameters; therefore we decided to approximate the wave by using a collection of sine segments.

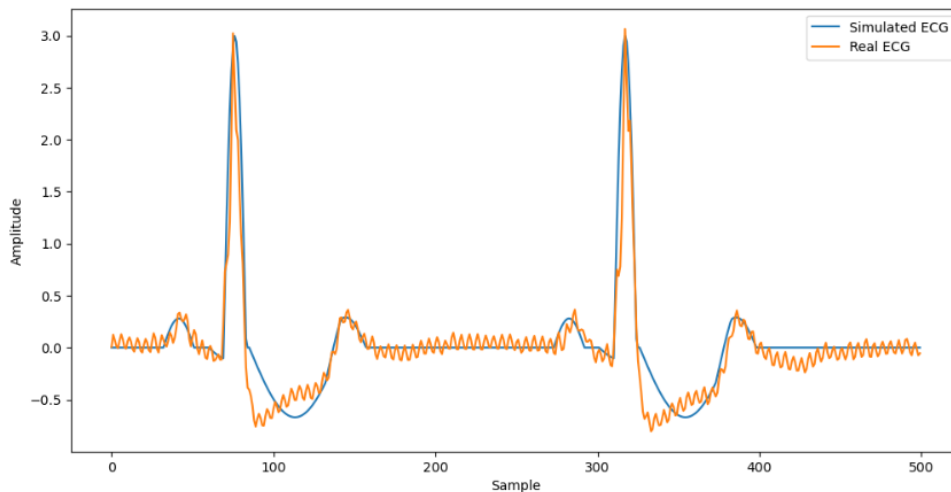
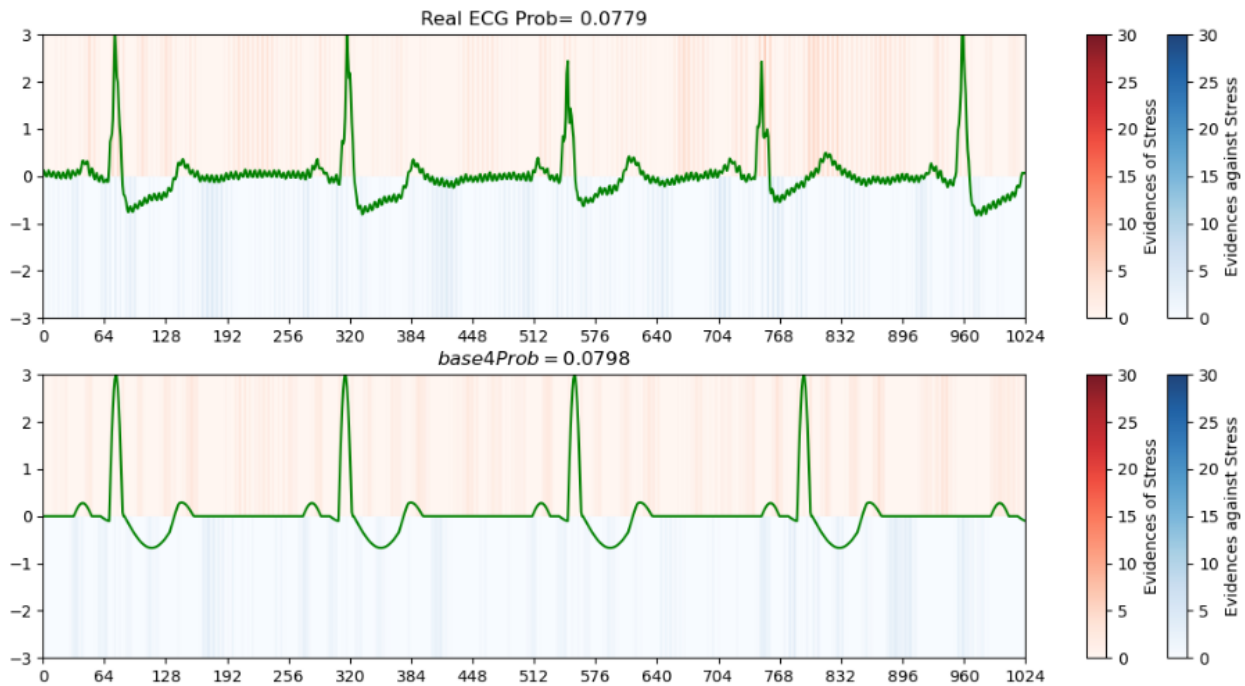


Figure 22 Synthetic ECG

Using this base ECG, the wave parameters can be changed to see the effect on the output, and corroborate they go in accordance with the medical background. The best way to test the similarity of our simulated ECGs to their real counterparts, and therefore see their viability for measuring the model's performance, is two compares this real ECG with the simulated one in the output logit and saliency.



The simulated ECG isn't a perfect model of the waveform, as there are significant differences in the saliency, and the predicted probability of the simulated waveform is higher, and there is some HRV. Nevertheless, generally, the areas of focus in the saliency maps are the same, and the predicted probability is very similar. From now on, this synthetic wave will be used as a baseline for the transformations to better resemble what scientific papers believe is a stressed waveform.

6.3.2 RR Interval

The first feature to be studied is the RR interval, in other words, the heart rate (HR). In this case, the RR interval was changed from 239(62.76 bpm) samples to 180 (83.33bpm). Which is still not considered tachycardia(100bpm), but can be considered an accelerated HR.

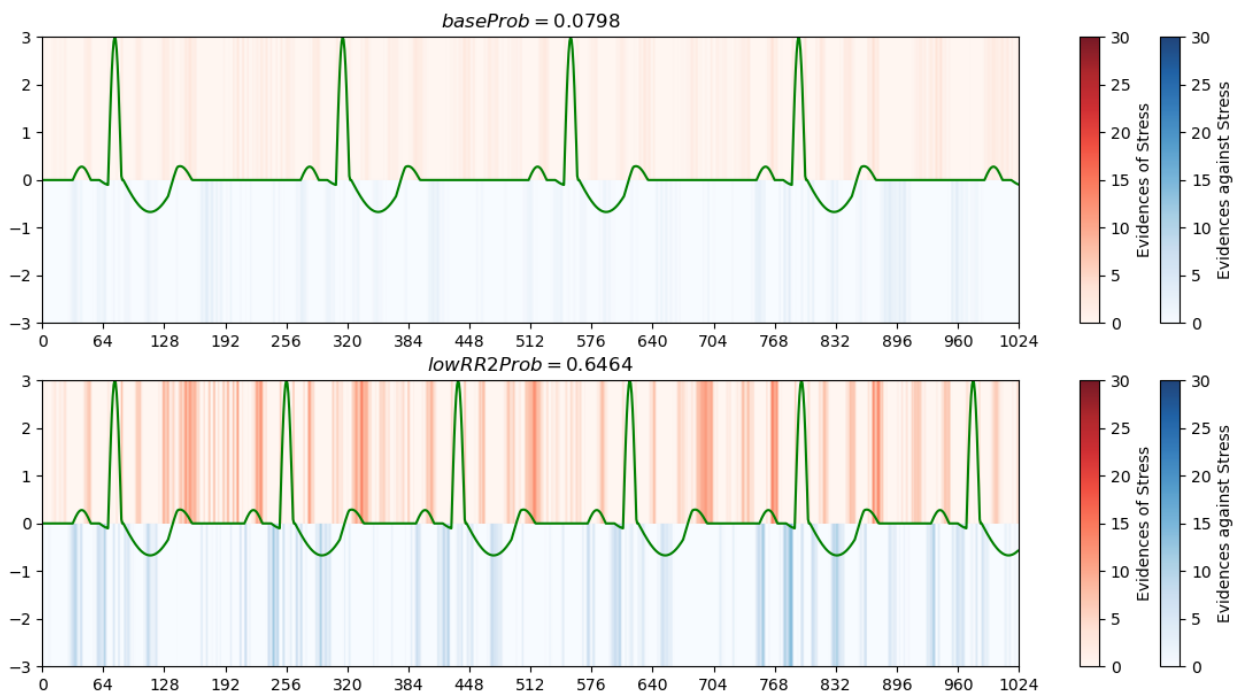


Figure 23 RR interval

This caused the predicted probability to drastically increase to 64.64% and the evidence of stress to appear much more pronounced on the waveform. It must be pointed out, however, that the QRS complex isn't being targeted as evidence of stress, this suggests that this part of the waveform wasn't deemed relevant for the model. Instead, it seems to have used other parts of the end of waveform, as the T or P as an anchor point of the waveform.

6.3.3 PR Interval

Another indicator of stress is the PR interval, which is the duration between the onset of the P wave and the QRS complex in an ECG. It reflects the time the electrical impulse takes to travel from the sinus node through the atrioventricular node, where it enters the ventricles. The PR interval is an important indicator of

atrioventricular conduction. In the context of stress, an altered PR interval may signify altered autonomic balance with increased sympathetic tone.

To show this, the PR interval of the base wave was shortened by 12 samples (48ms).

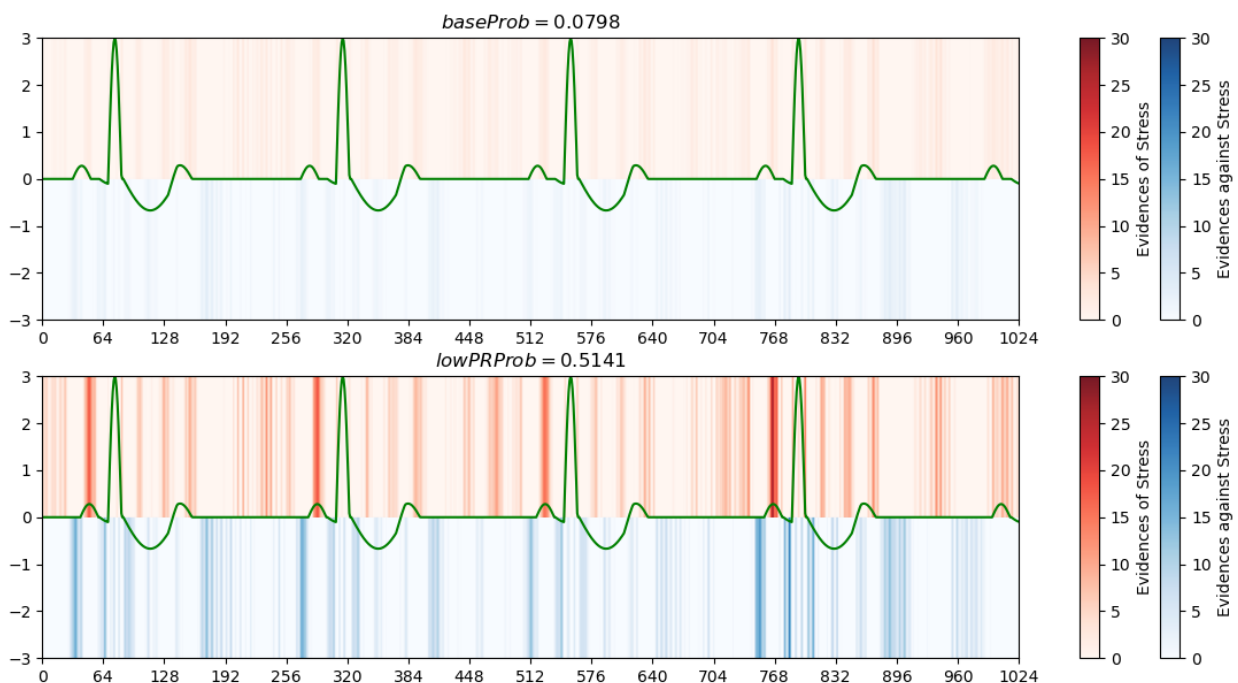


Figure 24 Low PR interval

Whilst this might look like an un significant alteration, the effect on the predicted probability is noteworthy. The probability of the waveform being associated with stress increased, highlighting the sensitivity of the model to PR interval changes.

6.3.4 T amplitude

The amplitude of the T wave is another prominent feature, it represents the repolarization of the ventricles, and its amplitude and shape can provide information of the autonomic nervous system and therefore stress. In the context of stress, an elevated T wave can be indicative of hyperkalaemia.

To analyse the model's response to changes in T amplitude, we modified the amplitude of the T wave in our base synthetic ECG doubling it and increasing its duration (4 samples) as not to distort the T waveform excessively.

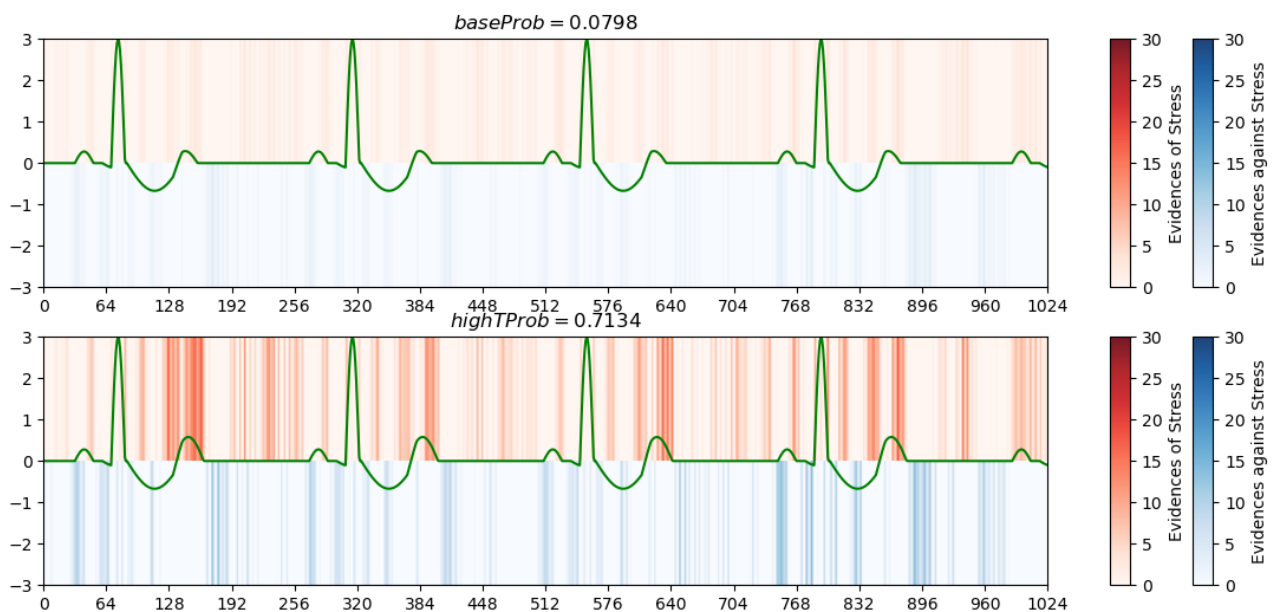


Figure 25 Increase of T wave amplitude.

This had a noticeable impact on the model's predictions. As shown in Figure 21, the heightened and prolonged T wave contributed significantly to the model's stress classification. The saliency map showed prominent evidence of stress over the T wave region, corroborating the clinical understanding that a heightened T wave can be indicative of stress.

6.3.5 QT interval

The QT interval represents the total time taken for the ventricles to depolarize and then repolarize, and it is measured from the beginning of the QRS complex to the

end of the T wave. It is, among other things, regulated by cardiac sympathetic innervation, and low intervals can be linked to stress.

To exemplify the effect, the QT interval of the base wave was shortened by 12 samples (48ms).

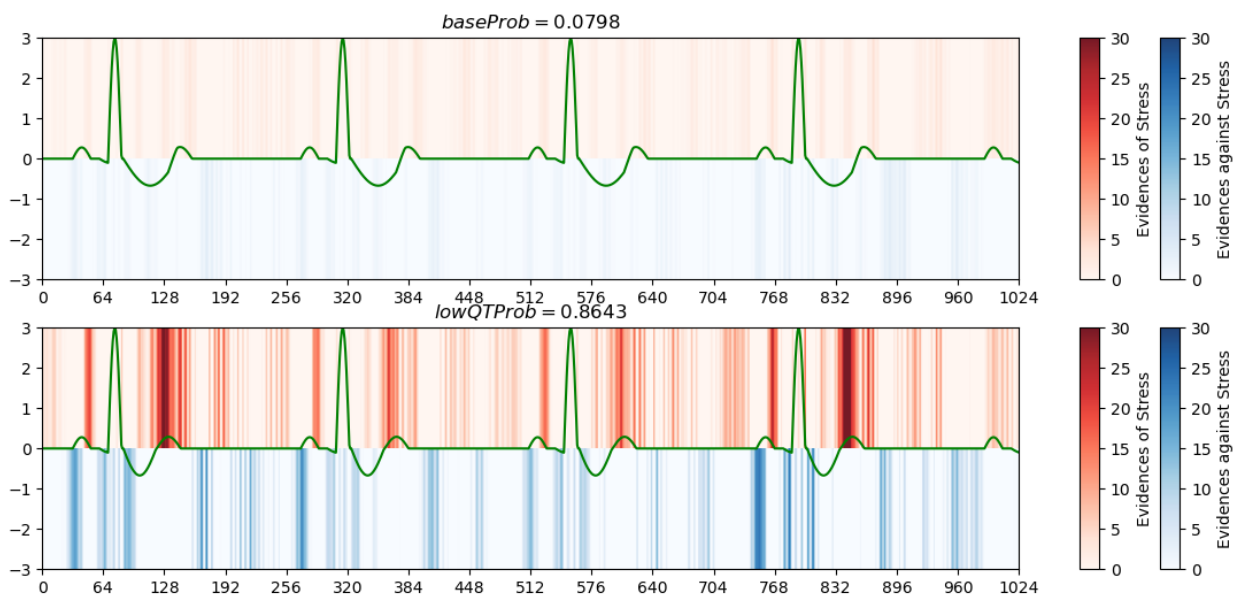


Figure 26 Low QT interval

6.4 Attention heatmaps

Attention heatmaps provide another way to visualize the decision-making process, in models that rely on self-attention mechanisms. This technique provides insight into which parts of the input data the model is paying attention to when making a prediction.

The attention mechanism allows a model to focus on different parts of the input for different tasks. By visualizing the attention weights, one can see which parts of the input the model finds relevant for a particular task.

In a trained model, the weights of the attention mechanism are learned during training. Once the model is trained, the weights themselves are fixed. However, the attention matrix produced by these weights for any given input is dynamic, meaning it depends on the input data.

The cyclic nature of the ECG signals creates the periodic lines and squares on the ECG, as the model pays attention to the same points of the signal at other cycles.

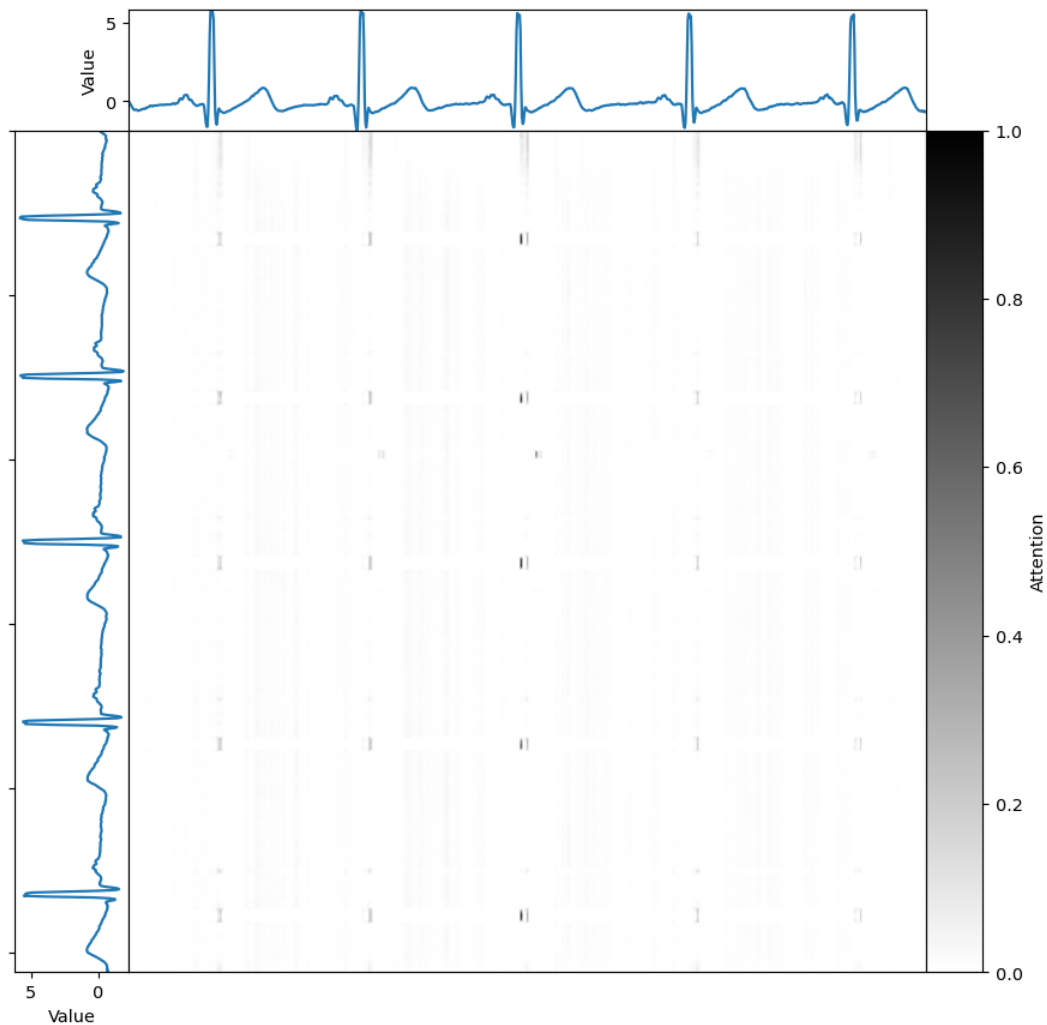


Figure 27 Attention heatmap of a non-stressed patient

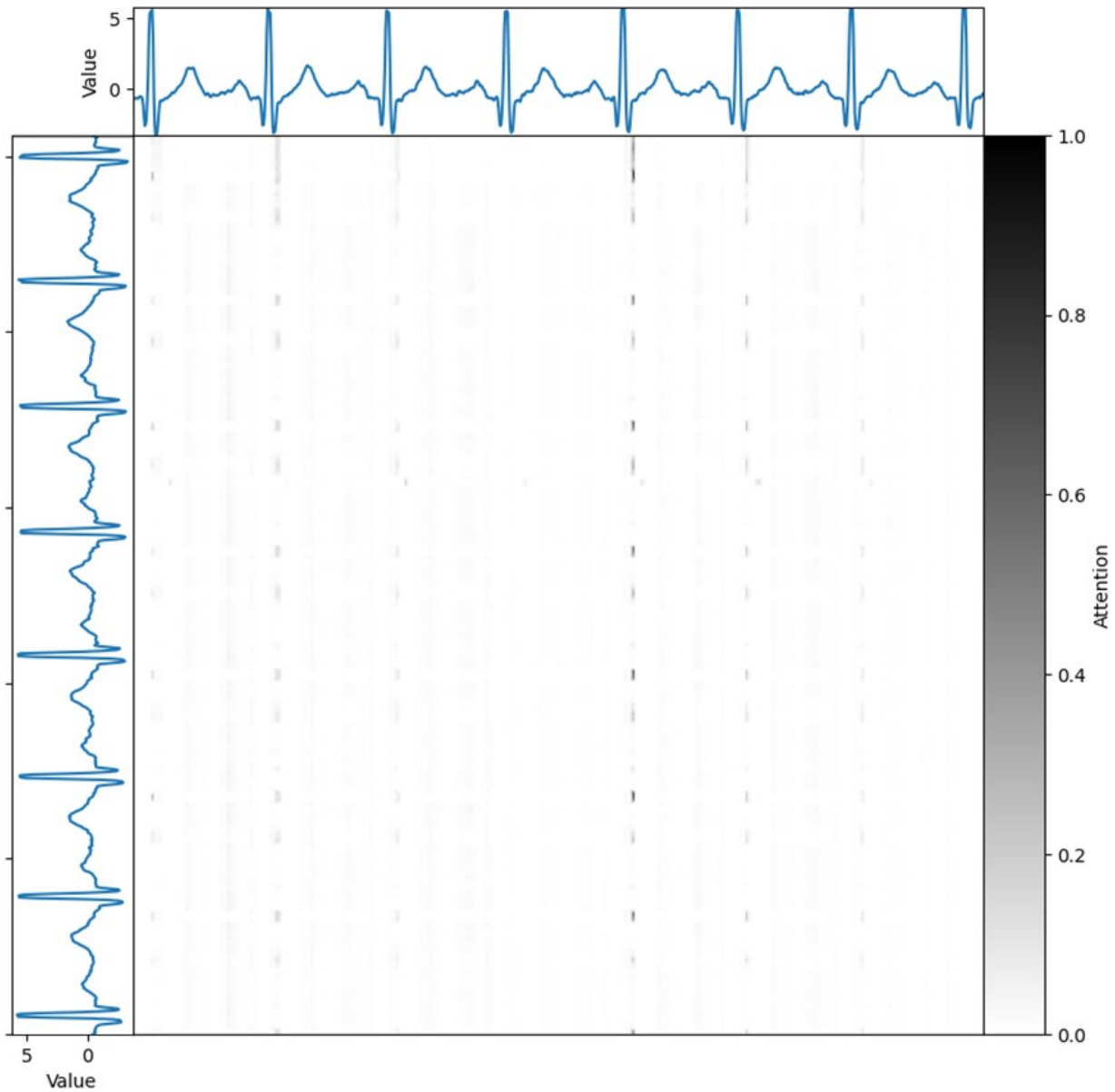


Figure 28 Attention heatmap of a stressed patient

Whether stressed or not, it's clear that the receptive field of the S-wave peak consistently focuses on signal points, like the QRS complex and the P wave. This indicates that the classifier's Multilayer Perceptron (MLP) is zeroing in on the

crucial parts of the signal. Essentially, there's a distinct correlation between the insights from the saliency maps and the model's emphasis on essential signal elements.

7

Conclusions and future works

The successful intersection of medical technology and artificial intelligence was a hallmark of this project. By drawing upon the latest advancements in Electrocardiogram (ECG) technology and integrating them with cutting-edge Deep Learning techniques, our endeavour has significantly elevated the potential of stress detection using raw ECG data over previous works. This synergy opens up promising avenues for the integration of these sensors for multiple applications, like the automotive field.

In terms of methodology, our research underscored the power of a two-stage process, which placed substantial emphasis on self-supervised learning. This pivotal decision gifted our model with an enhanced capability to recognize and decode intricate patterns from ECG signals over traditional HRV analysis, thus amplifying its accuracy and reliability manifold.

Further, by harnessing state-of-the-art computational paradigms like Transformers and Self-Supervised Learning, not only positioned the project at the forefront of innovation but also served as a testament to the transformative power of contemporary AI techniques when wielded in medical contexts. Such an adoption not only reflects our commitment to the best in technology but also casts a spotlight on the possibility of real-time and non-invasive stress detection.

Parallely, our deep-dive into interpretability studies, particularly through tools like saliency maps and attention heatmaps, has demystified the AI decision-making,

reinforcing the imperative of transparency in AI-driven outcomes which is crucial in the medical field.

7.1 Future works

Even though we are delighted with our results, we recognize that there is always room for improvement and expansion in such a dynamic field.

One immediate area of potential lies in the expansion of our dataset. The WESAD dataset used is quite limited in size and might be a limiting factor for the model's real-life adoption. By integrating a broader and more varied spectrum of ECG signals, we can potentially enable our model to achieve even greater levels of accuracy and generalization. This diversity in data can be instrumental in mirroring the true complexity and variability seen in real-world stress patterns.

A natural progression of this project would be its deployment in real-world scenarios. Given the potential applications in diverse fields like vehicular safety, continuous healthcare monitoring, and even workplace stress management, our model can be a game-changer. Experimenting with its deployment in these arenas can yield practical insights and pave the way for broader adoption.

Linking with the previous point, in a real deployment of the model, the recorded signals are likely to have some sensor distortion which needs to be calibrated for, varying for each sensor model. Our two-stage training model will ensure this retraining is minimal and more focused on the latter stages of the model, without a complete retraining, for more efficient and cost-effective adaptation to different deployment environments.

Another area that warrants further exploration is the architectural nuances of the model.

One aspect that could significantly lighten the model for IoT applications would be reducing the CNN encoder's parameters, as they largely contribute to the overall

model's size. There are several ways to do this, but it is important for the architecture to maintain the receptive field of approximately 60bpm.

One avenue would be reducing the sampling rate to 125Hz, which would almost half the number of parameters used. And, as shown while explaining the methodology, theoretically should be possible to lower the sampling rate up to 125Hz without significantly reducing the delineation, but we would rather not push the model so far in this study.

As for the transformer, it does undoubtedly lead to great results as using just one layer. But some of its hyperparameters, such as the number of heads, the K Q V dimensionality (which comes from the CNN), could also be studied.

Lastly, stress, as a physiological phenomenon, is closely tied to our emotional and mental well-being. It would be intriguing to explore the use of the model for said applications.



UNIVERSIDAD PONTIFICIA COMILLAS
ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA (ICAI)
MASTER'S IN INDUSTRIAL ENGINEERING

8

Bibliography

- [1] B. Behinaein, A. Bhatti, D. Rodenburg, P. Hungler, and A. Etemad, "A Transformer Architecture for Stress Detection from ECG," in *2021 International Symposium on Wearable Computers*, Sep. 2021, pp. 132–134. doi: 10.1145/3460421.3480427.
- [2] P. Bota, C. Wang, A. Fred, and H. Silva, "Emotion Assessment Using Feature Fusion and Decision Fusion Classification Based on Physiological Data: Are We There Yet?," *Sensors*, vol. 20, no. 17, Art. no. 17, Jan. 2020, doi: 10.3390/s20174723.
- [3] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection," in *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, Boulder CO USA: ACM, Oct. 2018, pp. 400–408. doi: 10.1145/3242969.3242985.
- [4] E. Castets, *Edouard99/Stress_Detection_ECG*. (Aug. 03, 2023). Jupyter Notebook. Accessed: Sep. 05, 2023. [Online]. Available: https://github.com/Edouard99/Stress_Detection_ECG
- [5] S. Ishaque, N. Khan, and S. Krishnan, "Detecting stress through 2D ECG images using pretrained models, transfer learning and model compression techniques," *Mach. Learn. Appl.*, vol. 10, p. 100395, Dec. 2022, doi: 10.1016/j.mlwa.2022.100395.
- [6] "Lead systems – how an ECG works | CardioSecur." Accessed: Jun. 06, 2024. [Online]. Available: <https://www.cardiosecur.com/magazine/specialist-articles-on-the-heart/lead-systems-how-an-ecg-works>

-
- [7] U. of Science and T. of China, "Researchers achieve contactless electrocardiogram monitoring." Accessed: Aug. 18, 2023. [Online]. Available: <https://medicalxpress.com/news/2022-12-contactless-electrocardiogram.html>
- [8] "Electrocardiogram (ECG) | CardioSecur." Accessed: Jun. 06, 2024. [Online]. Available: <https://www.cardiosecur.com/magazine/specialist-articles-on-the-heart/electrocardiogram-ecg>
- [9] "The ultimate ECG book & course: learn ECG interpretation, videos, test/quiz -," ECG & ECHO. Accessed: Aug. 23, 2023. [Online]. Available: <https://ecgwaves.com/>
- [10] R. Gordan, J. K. Gwathmey, and L.-H. Xie, "Autonomic and endocrine control of cardiovascular function," *World J. Cardiol.*, vol. 7, no. 4, pp. 204–214, Apr. 2015, doi: 10.4330/wjc.v7.i4.204.
- [11] K. Tzevelekakis, Z. Stefanidi, and G. Margetis, "Real-Time Stress Level Feedback from Raw Ecg Signals for Personalised, Context-Aware Applications Using Lightweight Convolutional Neural Network Architectures," *Sensors*, vol. 21, no. 23, p. 7802, Nov. 2021, doi: 10.3390/s21237802.
- [12] J. Healey and R. Picard, "SmartCar: detecting driver stress," in *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, Sep. 2000, pp. 218–221 vol.4. doi: 10.1109/ICPR.2000.902898.
- [13] A. Hernández-Vicente *et al.*, "Heart Rate Variability and Exceptional Longevity," *Front. Physiol.*, vol. 11, p. 566399, Sep. 2020, doi: 10.3389/fphys.2020.566399.
- [14] A. Bhide, R. Durgaprasad, L. Kasala, V. Velam, and N. Hulikal, "Electrocardiographic changes during acute mental stress," *Int. J. Med. Sci. Public Health*, vol. 5, no. 5, p. 835, 2016, doi: 10.5455/ijmsph.2016.19082015137.
- [15] A. Chaudhuri, M. Ray, K. Arpan, and Ghosh, "Effect of Perceived Stress on Electrocardiographic Parameters of Female Health Professionals: A Cross Sectional Observational Study," Oct. 2019.
- [16] J. Barry *et al.*, "Frequency of ST-segment depression produced by mental stress in stable angina pectoris from coronary artery disease," *Am. J. Cardiol.*, vol. 61, no. 13, pp. 989–993, May 1988, doi: 10.1016/0002-9149(88)90112-9.
-

-
- [17] S. Rabbani and N. Khan, "Contrastive Self-Supervised Learning for Stress Detection from ECG Data," *Bioengineering*, vol. 9, no. 8, Art. no. 8, Aug. 2022, doi: 10.3390/bioengineering9080374.
- [18] N. Keshan, P. V. Parimi, and I. Bichindaritz, "Machine learning for stress detection from ECG signals in automobile drivers," in *2015 IEEE International Conference on Big Data (Big Data)*, Oct. 2015, pp. 2661–2669. doi: 10.1109/BigData.2015.7364066.
- [19] P. Zhang *et al.*, "Real-Time Psychological Stress Detection According to ECG Using Deep Learning," *Appl. Sci.*, vol. 11, no. 9, Art. no. 9, Jan. 2021, doi: 10.3390/app11093838.
- [20] F. Agrafioti, D. Hatzinakos, and A. K. Anderson, "ECG Pattern Analysis for Emotion Detection," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 102–115, Jan. 2012, doi: 10.1109/T-AFFC.2011.28.
- [21] J. Vazquez-Rodriguez, G. Lefebvre, J. Cumin, and J. L. Crowley, "Transformer-Based Self-Supervised Learning for Emotion Recognition," presented at the 26th International Conference on Pattern Recognition (ICPR 2022), Aug. 2022. Accessed: May 31, 2023. [Online]. Available: <https://hal.science/hal-03634490>
- [22] J. S. Burma, A. P. Lapointe, A. Soroush, I. K. Oni, J. D. Smirl, and J. F. Dunn, "Insufficient sampling frequencies skew heart rate variability estimates: Implications for extracting heart rate metrics from neuroimaging and physiological data," *J. Biomed. Inform.*, vol. 123, p. 103934, Nov. 2021, doi: 10.1016/j.jbi.2021.103934.
- [23] A. Vaswani *et al.*, "Attention Is All You Need," Dec. 05, 2017, *arXiv*: arXiv:1706.03762. doi: 10.48550/arXiv.1706.03762.
- [24] O. Kwon *et al.*, "Electrocardiogram Sampling Frequency Range Acceptable for Heart Rate Variability Analysis," *Healthc. Inform. Res.*, vol. 24, no. 3, pp. 198–206, Jul. 2018, doi: 10.4258/hir.2018.24.3.198.
- [25] F. Simon, J. P. Martinez, P. Laguna, B. Van Grinsven, C. Rutten, and R. Houben, "Impact of Sampling Rate Reduction on Automatic ECG Delineation," in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology*
-

-
- Society*, Lyon, France: IEEE, Aug. 2007, pp. 2587–2590. doi: 10.1109/IEMBS.2007.4352858.
- [26] H. Choi and P. Kang, “Multi-Task Self-Supervised Time-Series Representation Learning,” Mar. 02, 2023, *arXiv*: arXiv:2303.01034. doi: 10.48550/arXiv.2303.01034.
- [27] G. Zerveas, S. Jayaraman, D. Patel, A. Bhamidipaty, and C. Eickhoff, “A Transformer-based Framework for Multivariate Time Series Representation Learning,” Dec. 08, 2020, *arXiv*: arXiv:2010.02803. doi: 10.48550/arXiv.2010.02803.
- [28] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” May 24, 2019, *arXiv*: arXiv:1810.04805. doi: 10.48550/arXiv.1810.04805.
- [29] Z. Shao, Z. Zhang, F. Wang, and Y. Xu, “Pre-training Enhanced Spatial-temporal Graph Neural Network for Multivariate Time Series Forecasting,” Jun. 2022. doi: 10.48550/arXiv.2206.09113.
- [30] H.-Y. S. Chien, H. Goh, C. M. Sandino, and J. Y. Cheng, “MAEEG: Masked Auto-encoder for EEG Representation Learning,” Oct. 27, 2022, *arXiv*: arXiv:2211.02625. doi: 10.48550/arXiv.2211.02625.
- [31] “Transformer — PyTorch 2.0 documentation.” Accessed: Jun. 05, 2023. [Online]. Available: <https://pytorch.org/docs/stable/generated/torch.nn.Transformer.html>
- [32] J. Zheng, H. Guo, and H. Chu, “A large scale 12-lead electrocardiogram database for arrhythmia study.” *PhysioNet*. doi: 10.13026/WGEX-ER52.
- [33] S. Tan *et al.*, “Icentia11K: An Unsupervised Representation Learning Dataset for Arrhythmia Subtype Discovery,” Oct. 21, 2019, *arXiv*: arXiv:1910.09570. doi: 10.48550/arXiv.1910.09570.
- [34] F. Wolling, “WESAD: Multimodal Dataset for Wearable Stress and Affect Detection.” Accessed: Aug. 29, 2023. [Online]. Available: <https://ubicomp.eti.uni-siegen.de/home/datasets/icmi18/>
- [35] P. S. Parvatharaju, R. Doddaiah, T. Hartvigsen, and E. A. Rundensteiner, “Learning Saliency Maps to Explain Deep Time Series Classifiers,” in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*,
-

Virtual Event Queensland Australia: ACM, Oct. 2021, pp. 1406–1415. doi: 10.1145/3459637.3482446.

- [36] P. Dolinský, I. Andras, L. Michaeli, and D. Grimaldi, "MODEL FOR GENERATING SIMPLE SYNTHETIC ECG SIGNALS," *Acta Electrotech. Inform.*, vol. 18, pp. 3–8, Sep. 2018, doi: 10.15546/aei-2018-0019.

ANNEX A:

Alignment with the Sustainable Development Goals (SDGs)

In our rapidly evolving world, ensuring sustainable growth and development remains paramount. To this end, the Sustainable Development Goals (SDGs) set forth by the United Nations act as a global blueprint, guiding concerted efforts towards building a brighter and more equitable future. In particular, our project deeply resonates with two pivotal goals: Goal 3, which emphasizes Good Health and Well-being, and Goal 9, which advocates for Industry, Innovation, and Infrastructure. By aligning our objectives and strategies with these SDGs, we aim to create a holistic impact, addressing both health concerns and technological advancements in tandem.

Goal 3: Good Health and Well-being

Stress is a significant health concern that can lead to severe physical and mental health issues if not properly managed. By developing an advanced stress detection model using ECG technology and deep learning, our project is inherently aligned with SDG 3. Our model's primary objective is to help individuals detect their stress levels accurately and in real-time, which could lead to more timely interventions and treatments.

Moreover, stress often acts as a precursor to various cardiovascular and mental health disorders. By providing an efficient and reliable stress detection tool, we may help reduce the prevalence and impact of these stress-related conditions. It's a step forward towards achieving the targets of SDG 3, which includes ending epidemics of communicable diseases and reducing mortality from non-communicable diseases and promoting mental health and well-being.

Goal 9: Industry, Innovation, and Infrastructure

The implementation of cutting-edge technology such as deep learning, ECG technology, and self-supervised learning techniques to enhance stress detection models is an embodiment of SDG 9. We're leveraging the potential of these

technologies to drive innovation and provide solutions to health challenges, thus fostering technological development in healthcare.

Furthermore, this project also contributes to building resilient infrastructure and fostering innovation. The creation of robust stress detection models will augment the capabilities of current healthcare infrastructures, particularly those related to mental health care. This aspect aligns with the targets of SDG 9, which includes upgrading infrastructure and retrofitting industries to make them sustainable, with increased resource-use efficiency and greater adoption of clean and environmentally sound technologies and industrial processes.

In conclusion, this project aligns with the SDGs by capitalizing on innovative technology for promoting well-being and supporting sustainable industrial development. The positive outcomes from this project can have a ripple effect, leading to broader benefits in health, industry, and beyond, contributing to the overarching aim of the SDGs – a better and more sustainable future for all.