

# An Approach using transformer architecture for emotion recognition through Electrocardiogram Signal(s)

Vincenzo Dentamaro<sup>1</sup>, Donato Impedovo<sup>1</sup>, Luigi A. Moretti<sup>2</sup>, Giuseppe Pirlo<sup>1</sup>, Prem K. Suresh<sup>1</sup>

<sup>1</sup> University of Bari Aldo Moro, Via Orabona 4, 70121, Bari Italy

<sup>2</sup> University of the West of England (UWE) - Coldharbour Ln, Stoke Gifford, Bristol BS16 1QY, UK

## Abstract

This study proposed an AI-based approach to detect seven emotional states (Happiness, Sadness, Surprise, Anger, Fear, Disgust, Neutral) based on an electrocardiogram (ECG). A well-known three-dimensional model (valence, arousal & dominance), also known as the PAD model, was used to classify the emotional spectrum. We propose a network architecture, Transformer and Temporal Convolution Network, based solely on attention mechanisms, without recurrence and convolution. A comparative analysis between different transfer learning and fine-tuning techniques was then carried out. Three databases were used, starting with the MIT BIH (Massachusetts Institute of Technology and Beth Israel Hospital) database for the characteristics of the recorded signals, and the DREAMER (Dataset for Emotion Analysis using EEG, Physiological and Video Signals) and YAAD (Young Adult Age Dataset) databases for the physiological recordings and subjective ratings of the PAD values. In this paper we address two different problems (heart disease and emotion recognition) using electrocardiogram signals. Evaluation metrics such as Mean Absolute Error and Mean Squared Error were used to assess the performance of the transfer learning models. The overall goal of this study is to analyze and compare the performance of the model and two different problems to understand the emotion in different scenarios. This includes all techniques for automatic evaluation of emotions for applications in marketing, video games, social media, website customization, healthcare, education and other fields.


## Keywords

Deep Learning, Transfer Learning, Fine Tuning, Electrocardiogram Signal (ECG), Transformer, Affective Computing

## 1. Introduction

With a rapidly growing global population, we are faced with the need to understand the feelings/emotions of others in a variety of situations, from social and entertainment settings to mental health scenarios. On the other hand, rapid improvements in technology are impacting our lifestyles and opening up new possibilities. Digital technologies can help us provide personalized care to all people in this busy society.

---

IEEESDS'23: Data Science Techniques for Datasets on Mental and Neurodegenerative Disorders, June 22, 2023, Zürich, Switzerland   
vincenzo.dentamaro@uniba.it (V. Dentamaro); premkumarsuresh@gmail.com (Prem K. Suresh);



0000-0003-1148-332X (V. Dentamaro);



© 2023 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



CEUR Workshop Proceedings (CEUR-WS.org)



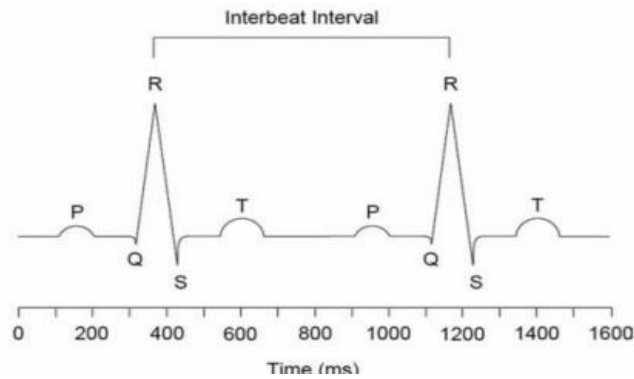
**Figure 1** VAD (Valence, Arousal, Dominance) Model

Nowadays, people are periodically monitored to ensure their physical health, but there is a lack of continuous monitoring of their psychological parameters and mental health. In recent years, many studies have reported that people of all ages are falling into depression, mental illness, stress, etc. As a result, many diseases are spreading as silent epidemics. This paper explores the possibilities of using psychological signals to detect emotions.

Emotions play an important role in our lives. People are often unable or afraid to express their emotional states, making them more vulnerable to emotional abuse, depression and other illnesses (i.e. co-morbidities). A person's emotional state enhances life through positive emotions that help prevent cognitive decline and other health problems. Conversely, negative emotions lead to a weaker health. The ability to recognise emotions is practically possible in the age of artificial intelligence. It offers enormous opportunities for human-computer interaction, robotics, healthcare, biometric security and behavioural modelling [1]. Several studies have reported that people facing stressful scenarios, such as pregnant women, are more susceptible to stress and depression. Emotion recognition can help us understand and prevent mental health problems at an early stage, improving treatment outcomes and people's quality of life.

An electrocardiogram (ECG) is a widely used medical technique that measures the electrical activity of the heart. ECGs are commonly used in cardiology wards to assess heart activity. However, it can also be useful to analyse the emotional state in real time. There are different ways to classify emotions, but in this paper we refer to the PAD model, which includes Valence, Arousal, Dominance, as shown in Figure 1. Latent dimensions are used to model emotions. Essentially these are valence (how good or bad a feeling is), arousal (the intensity of the emotion in response to a stimulus) and dominance (a sense of control over the emotion).

By combining these continuous scores, we can represent more nuanced emotional states. The underlying principle behind using the ECG for emotion recognition is the link between the autonomic nervous system (ANS) and emotional responses. The ANS is responsible for regulating various bodily functions, including heart rate, blood pressure and breathing. It consists of two antagonistic branches: the sympathetic nervous system (SNS) and the parasympathetic nervous system (PNS).



**Figure 2** QRS Interval in ECG Track

The following is a summary of the steps required for emotion recognition from the ECG trace:

1. Data acquisition: ECG signals are recorded using electrodes placed on the subject's chest and/or limbs. The electrodes record the electrical activity generated by the heart and the signals are amplified and digitised for further analysis.
2. Pre-processing: The recorded ECG signals may contain noise, artefacts or baseline drift, which can affect the accuracy of emotion recognition. Pre-processing techniques such as filtering, artefact removal and baseline correction are applied to improve the quality of the signals.
3. Feature extraction: Various features are extracted from the pre-processed ECG signals to capture relevant information related to emotional states. These features may include statistical measures, spectral analysis or non-linear measures derived from heart rate or HRV.

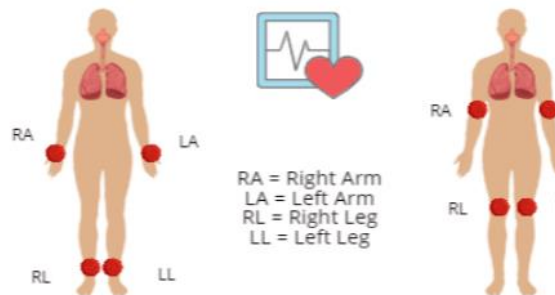
Electrocardiograms work by detecting electrical changes in the heart as it contracts and relaxes, giving us a cycle of signals [2]. As a result, an ECG detects and records the strength and timing of the electrical activity generated in the heart. Each phase of the electrical signal as it passes through your heart is plotted on a graph where this information is recorded. The PQRST complex in Figure 2 shows the combination of the P, Q, R, S and T waves that represent the deflections in an ECG signal. These waves represent the electrical potential changes in the right and left atrium and ventricle [3]. The QRS complex, part of the PQRST complex, is, as the name suggests, the combination of the Q, R and S waves of the ECG signal. Heart rate (HR) and heart rate variability (HRV) can be measured using the rate of the QRS complex or just the R-peak (i.e. left ventricular contraction). Heart rate variability (HRV) is significantly associated with average heart rate (HR), so HRV provides information on two quantities: HR and its variability [4].

## 2. Physiological Signals

It is useful to find a way of measuring the subject's emotional changes over time. This is possible because emotions can be expressed in a person's everyday life, through physical reactions, facial expressions, or through voice or intonation. The effectiveness of the machine in recognising emotions has been enhanced by the technique of deep learning with neural networks. Recent deep learning research has included a variety of human behavioural inputs, including audio-visual inputs, facial expressions, body language, bio-signals and associated brain waves. The standard sensor points on the human body to collect an ECG trace are shown in Figure 3.

Changes in the activity of the visceral motor (autonomic) system are among the most obvious indicators of emotional arousal. Thus, different emotions may be accompanied by changes in heart rate, cutaneous blood flow (flushing or pallor), piloerection, sweating and gastrointestinal motility [5]. The production of isolated or generalised emotions has been shown to correlate physiologically with the activation of brain regions and structures.

Even if a person chooses not to show their emotions, there is an inevitable change in physiological signals that can't be hidden or avoided, as the ANS sympathetic nerves are activated whenever a person is positively or negatively stimulated [6]. This sympathetic activation leads to changes in heart rate, breathing rate and blood pressure, which are considered to be some of the most common responses of the human body to a given emotion, as they are easily recorded compared to other physiological signals, such as the electroencephalogram (EEG). It is also a great source of information that can be correlated with emotional states and is a technique already developed and used in the medical field.



**Figure 3** Human body sensor point for Physiological Signal

### 3. State of the art review

A great method to summarize existing knowledge in each domain is Systematic Literature Review (SLR). It involves identifying, evaluating, and interpreting available research relevant to a certain research question [8]. In our SLR, we are posing the following research question: How do traditional methods compare to deep learning approaches, specifically transformer models, in terms of accuracy and performance in emotion recognition from ECG signals? To refine the number of studies considered in our SLR, we support our question with a set of criteria.

Inclusion criteria:

- Basic emotions of human,
- Personal device/wearable used,
- The physiological signal is monitored,
- The ECG Signal is monitored through a wearable device.

We used three databases (MIT BIH, DREAMER, YAAD) to find articles relevant to our research question: Scopus, Web of Science, and Google Scholar via Publish or Perish. Our search was narrowed down by the following terms: emotion, affective, wearable, smartwatch, smart device, smart band, transformer architecture, transfer learning. More than 2,000 papers were found. Papers based on emotion recognition phrase in the title or abstract have been given priority. Consequently, it is probable that most pertinent articles have already been found at this point.

There are some research attempts to recognize emotion through ECG as it has application in many fields such as robotics, medicine, and organization et., Since the twentieth century, Ekman et al. defined seven basic emotions, irrespective of culture in which a human grows with the seven expressions (Anger, Fear, Happy, Sad, Disgust, Surprise & Neutral). Emotions are complex processes, including feelings, body language, cognitive reactions and behaviour or thoughts [2]. Different models have been proposed for automatically recognizing emotions, considering the way all these processes may interact with each other. However, there's still no universally accepted formulation to model emotions. In recent studies the improvements in neuroscience [3] and cognitive science [4] that drive the advancement of research in the field of emotion recognition. Also, the development in computer vision [5] and machine learning [6], [21], [22], [23] and deep Learning [7] makes emotion recognition

much more accurate and accessible to the general population. As a result, emotion recognition is growing rapidly which aims to be helpful for people to understand emotions in many situations.

Data collection is a critical aspect of any research project, especially when it comes to emotion recognition using physiological signals. Below, I'll describe the data collection processes for the MIT-BIH Arrhythmia Database, the DREAMER dataset, and the YAAD dataset, focusing on their relevance to emotion recognition:

### 1. MIT-BIH Arrhythmia Database:

**Source:** The MIT-BIH Arrhythmia Database is a well-established dataset widely used for arrhythmia detection research. It was created by the Massachusetts Institute of Technology (MIT) and includes ECG recordings from a diverse population of patients.

**Data Type:** This dataset primarily contains ECG (Electrocardiogram) recordings. It is not originally designed for emotion recognition but rather for arrhythmia detection and related cardiac studies.

**Collection Process:** The ECG recordings in this dataset were collected using electrodes attached to the skin to capture the electrical activity of the heart. The patients underwent monitoring under various conditions, including normal and arrhythmic rhythms.

**Emotion Information:** The MIT-BIH Arrhythmia Database does not include explicit emotion labels. Therefore, if it is being used in emotion recognition research, additional steps would be required to associate ECG data with emotional states, either through physiological responses or annotations provided separately.

### 2. DREAMER Dataset:

**Source:** The DREAMER dataset is specifically designed for emotion recognition research. It was created by the University of Genoa and the University of Geneva.

**Data Type:** DREAMER is a multimodal dataset, meaning it includes various types of data such as ECG, audio, and video recordings, making it well-suited for studying emotions.

**Collection Process:** Data collection for DREAMER involved recording physiological signals (including ECG) alongside audiovisual stimuli designed to elicit different emotions. Participants were exposed to stimuli while their physiological responses were monitored.

**Emotion Information:** The key feature of the DREAMER dataset is that it includes explicit emotion labels corresponding to the emotional states elicited by the provided stimuli. This allows for supervised emotion recognition training and evaluation.

### 3. YAAD Dataset:

**Source:** YAAD, which stands for "You Acting Against Disinformation," is a dataset created for emotion recognition and deception detection research. It was developed by the University of Oulu, Finland.

**Data Type:** YAAD includes various types of data, including audio, visual, and physiological signals, such as ECG.

Collection Process: The YAAD dataset was collected during interviews and interactions where participants were subjected to deceptive scenarios and varying emotional states. Physiological signals, including ECG, were recorded during these interactions.

Emotion Information: Like DREAMER, the YAAD dataset includes explicit emotion labels corresponding to the emotional states of participants during the interviews. This enables emotion recognition research using supervised learning methods.

The MIT-BIH Arrhythmia Database primarily provides ECG data but lacks explicit emotion labels. On the other hand, the DREAMER and YAAD datasets are specifically designed for emotion recognition, providing multimodal data, including ECG, alongside explicit emotion annotations. Researchers interested in emotion recognition often prefer datasets like DREAMER and YAAD due to their comprehensive data collection processes and labeled emotional states.

## 4. Methods

The idea behind this research starts from Arrhythmia (Health Problem) and goes through emotion recognition. To achieve this, we employ Deep Learning and Transfer Learning techniques to distinguish between health problems (Detecting Arrhythmia) and to recognize human emotions that are classified into three emotional classes (valence, arousal, dominance).

### 4.1. DATASET:

The study revealed that emotion recognition is predominantly carried out through signals recognition semantics on standard databases such as MIT-BIH, DREAMER, YAAD.

Initially, MIT BIH Arrhythmia experiment was carried out, and analyzed ECG data from five separate classes containing 109,446 beats collected from the MIT-BIH arrhythmia database. The results were evaluated using various applications, ranging from the most basic to the most advanced. Initializing features as reading the annotation from ECG Signal Data and then assigning the symbol attribute of the resulting annotation object to variable 'symbol'. Classifying beats and labelling into Normal beats as (0), abnormal (1) and other beats as (-1). Preprocessing ECG Signal and annotation with number of seconds ECG track, sampling rate, and list of annotation symbols that are abnormal beats [V, A, F] as Ventricular, Atrial, Fibrillation rate which is a type of cardiac arrhythmia characterized by an irregular heartbeat.

DREAMER is a multimodal database that consists of EEG (electroencephalogram) and ECG (Electrocardiogram) signals recorded during emotion elicitation experiments [25]. This database has records from 23 participants while they were presented with audiovisual stimuli, consisting of 18 videos. In this way, in terms of subject and records number, it is a small database that can show some limitations. Valence and arousal scores are predicted through ECG and EEG recordings, rather than the actual emotions assessed to videos. Since assessment has been done in different study beforehand, the response of Dreamer study participants may differ from assessed.

The dataset YAAD consists of two configurations, one with single modal ECG signals and the other with multi-modal ECG and GSR signals. The provided multimodal dataset comprises seven emotional states (happy, sad, anger, fear, disgust, surprise, and neutral). Each of these seven states consists of five levels of very low, low, moderate, high, and very high annotations representing the intensity of the felt state with a total of 35 states. The multimodal sub-folder has distinct sub-folders for the raw data of the ECG and GSR signals. Twelve subjects' simultaneous ECG and GSR readings were gathered. 252

files (3 sessions x 12 people x 7 emotions) are contained in each ECG and GSR folder. 25 volunteers, including 10 women and 15 men, were included in the data that was supplied.

#### 4.2 Deep learning model:

A transformer is a deep learning [12] model introduced by Vaswani [13] in the paper “Attention is all you need” that adopts the mechanism of self-attention, differentially weighting the significance of each part of the input data. It attempts to handle long-range dependencies with ease while resolving tasks that are sequence-to-sequence in figure 5. It is used primarily in the fields of natural language processing (NLP) [14] and computer vision (CV) [15]. Attention model is different from the classic sequence-to-sequence model in two ways.

##### 4.2.1 Transformer:

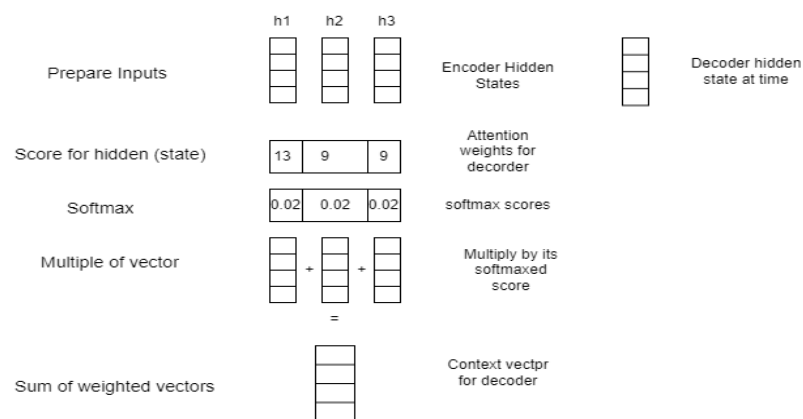


Figure 5 Transformer architecture with encoder and decoder of model

- Multispeed transformer:** To understand and use transformers (sequence-to-sequence architecture), we need to understand the attention mechanism. It has an infinite reference window. Attention mechanism based on encoder decoder type architecture. On an elevated level, the encoder maps an input sequence into an abstract continuous representation that holds all the learned information for the entire sequence. For processing ECG data, as they learn representation of data at multiple time scales and frequencies, capturing both short term and long-term patterns in data.

As compared to a simple seq-to-seq model, the encoder passes a lot more data to the decoder. The encoder sends the decoder all concealed states, including intermediate ones. It checks each hidden state that it received as every hidden state of the encoder is mostly associated with a given input. (i) Input Embedding: First step, feeding the input as feature data into the embedded layer. For each data, map to a vector with continuous value to represent the signal (ECG). It learns factor representation of each beat and classes through numbers. (ii) Positional Encoding and layer: Position and order of words are the essential parts of any language. They parse the data one by one in a sequential manner.

It injects the positional information into the embedded layer after the input is given because a transformer encoder has no recurrence like recurrent neural network.

(iii) Multi Headed attention: A specific Attention mechanism called (self-attention) which allows the model to associate each individual value in the input to the other input. It's possible that the model learns in a structured way of pattern. To achieve self-attention, feeding the

input into three distinct fully connected layers to create query, key, and value.

The Multi-Speed Transformer architecture [16], aims to learn meaningful time-dependent correlations and patterns at two different scales: fast and slow. It utilizes the concept of multiscale learning, where data is analyzed at multiple resolutions. An analogy can be drawn with a microscope slide viewed at various magnifications, where high resolution reveals small details and low resolution captures broader concepts. The Multi-Speed Transformer consists of two parallel branches. In the top branch, 1D convolution is performed with a stride of 1, followed by dilated convolution with a dilation rate of 2. In the second parallel branch, the first convolution has a stride of 3. Despite this difference, both branches share the same structure: the top branch incorporates a Positional Encoding Layer that adds the output of the dilated 1D convolution with a positional signal, using sine and cosine functions.

The Positional Encoding remolds the temporal dependency captured by the dilated convolutional layer to prevent its loss when injected into the Multi-Headed Attention Layer. The Multi-Headed Attention layer, trained through self-attention, captures correlations between elements within the same sequence. Its output is combined with the output of the positional encoding, allowing residuals to propagate forward. Subsequently, both parallel branches undergo Z-Score normalization. The outputs of the branches are concatenated and fed into a mono-dimensional global average layer, followed by a dense layer with the ELU activation function, and finally a dense layer with SoftMax. This merged representation of patterns extracted at different scales facilitates the decision-making process.

The two key components explaining the varying speed in parallel branches are the varying stride and the use of dilated convolutions. The 1D convolution with varying stride is mathematically represented by equation (4), where the input  $x$ , kernel  $h$ , and number of positions after each convolution operation are involved. A stride  $s > 1$  results in information loss, akin to sacrificing fine-grained details in favor of capturing the bigger picture. This can be understood as a moving average filter with a non-overlapping window.

Dilated convolution involves skipping some input values to cover a larger area. Dilated convolution expands the field of view without increasing computational cost and importantly, eliminates the need for a pooling layer, thus preserving resolution in the output series. In summary, the top branch focuses on capturing fine details of the movement while minimizing computational cost, while the lower branch sacrifices information to obtain a global view of the movement.

- **Vanilla transformer:** The model is based on self-attention mechanisms and does not use any convolutional or recurrent layers. The self-attention technique allows the model to pay attention to different input sequence fragments and recognize long-range correlations. In the context of natural language processing, the Vanilla Transformer [18] has been used for tasks such as language modeling, machine translation, and text classification. However, it can also be applied to other types of sequential data, such as time series data like ECG signals. It is a sequence-to-sequence model and consists of an encoder and a decoder, each of which is a stack of identical blocks. A multi-head self-attention module and a position-wise feed-forward network (FFN) make up most of each encoder block. A residual connection is used around each module, followed by a Layer Normalization module, to help develop a deeper model. Decoder blocks insert cross-attention modules between the multi-head, self-attention modules and the position-wise FFNs in addition to the encoder blocks' encoder blocks. Furthermore, the self-attention modules in the decoder are adapted to prevent each position from attending to subsequent positions.



Here, it adopts the attention mechanism with the Query-Key-Value (QKV) model. Given the packed matrix representations of queries  $Q \in R^{N \times D_k}$ , keys  $K \in R^{M \times D_k}$ , Values  $V \in R^{M \times D_v}$ .

SoftMax is applied in a row-wise manner. To address the gradient vanishing issue with the SoftMax function, the dot-products of queries and keys are split by  $d_k$ . The original  $D_m$ -dimensional queries, keys, and values are projected onto the corresponding  $D_k$ ,  $D_k$ , and  $D_v$  dimensions using  $H$  distinct learnt projection sets.

In Transformer, there are three types of attention in terms of the source of queries and key-value pairs: (i) Self-attention. In Transformer encoder, we set  $Q = K = V = X$  in Equation, where  $X$  is the outputs of the previous layer. (ii) Masked Self-attention. In the Transformer decoder, the self-attention is restricted such that queries at each position can only attend to all key-value pairs up to and including that position. To enable parallel training, this is typically done by applying a mask function to the unnormalized attention  $\hat{A} = \exp\left(\frac{QK^T}{\sqrt{d_k}}\right)$  where the illegal positions are masked out by setting. The terms autoregressive or causal attention are frequently used to describe this type of self-attention. (iii) Cross-attention projects the keys and values using the outputs of the encoder, while the queries are projected using the outputs of the preceding (decoder) layer.

- **Temporal convolutional network (TCN):** The main idea behind TCN [20] is to use dilated causal convolutions, which are 1D convolutions that preserve the temporal structure of the input sequence and can capture dependencies over large time intervals. The dilated causal convolution has a dilation rate parameter that controls how much the filter is shifted at each time step, allowing it to effectively capture long-term dependencies.

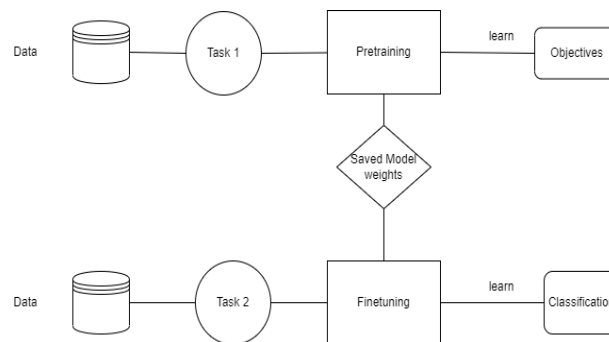
It can handle sequences of variable length. They can process sequences in parallel, which makes them computationally efficient and can capture long-term dependencies in sequences without the need for recurrent connections, which can make them easier to train. Each layer in the TCN takes in the output from the previous layer and applies a series of convolutional filters to it. The key feature of TCN architecture is the use of dilated convolutions, which increases the receptive field of the network without increasing the number of parameters. This allows the network to capture long-term dependencies in the input data, which is crucial for processing sequential data. The first layer in the TCN typically has a small dilation factor, which means that the convolutional filters are applied with a small spacing between them. As the layers progress deeper into the network, the dilation factor is increased, which allows the filters to have a larger receptive field and capture more long-term dependencies in the input data.

### 4.3 Proposed Approach

Transfer learning focuses on gathering knowledge by solving one problem and applying it to a related problem in the same domain [10]. We used transfer learning to improve ECG classifiers. Some articles make use of the similarities between various ECG circumstances [9] to transmit information between related tasks via transfer learning [20] in Figure 4. For instance, [9] presented a method for ECG heartbeat classification based on transferable representations using 1-dimensional residual networks. Like their work, we improve ECG classifiers using transfer learning and fine-tune the pretrained networks for emotion classification. First, we pretraining on the MIT-BIH Arrhythmia Classification. Next, transfer learning on the DREAMER dataset and test with YAAD to compare results. In contrast to these investigations, we exclusively concentrate on transferable ECG representations rather than

transferable picture representations. Applying information gained from solving one problem to another that is unrelated but similar is known as transfer learning.

Applying information gained from solving one problem to another that is unrelated but similar is known as transfer learning. A deep neural network (DNN) is typically pretrained on a sizable amount of data (also known as the upstream data set) [11] before being fine-tuned on the much smaller target data set in transfer learning (i.e., downstream data set). The process is divided into 3 steps: (1) Pretrained model and feed input to Feature data set used to learn and classify emotion values. (2) The weights are transferred from a pretrained model which is used as initial weights of a new neural network. (3) Then, this model is finetuned on the other feature dataset (smaller size) to classify Emotion values.



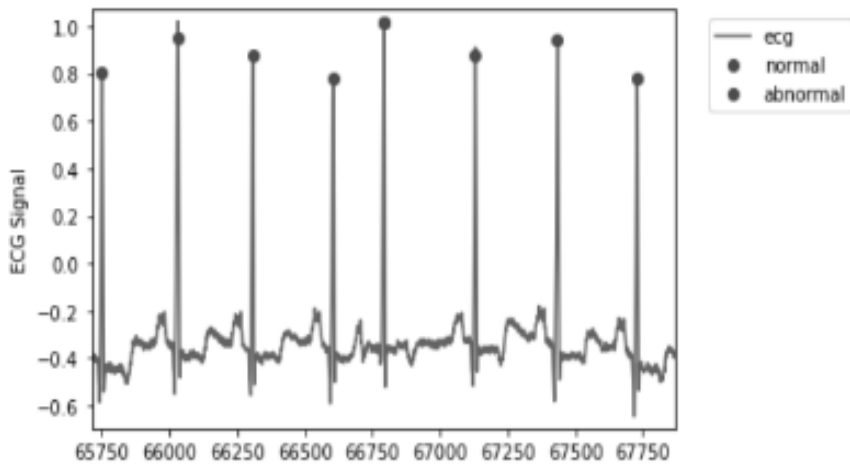
**Figure 4** Representation of Transfer Learning

## 5. Experiments

In many research, they found that ECG has been developed and remains constant over the years and upgraded to obtain more results. Using physiological signals (ECG) for emotion recognition was a recent approach compared to other types (Facial, Speech). It can show variation in emotion, through the heart rate (HR), HR variability, emotion classes [17]. The experiments carried out of this research for Emotion recognition using ECG were performed on three different architecture Multispeed transformer, Vanilla transformer, and Temporal Convolutional network with feature dataset MIT-BIH and Dreamer and testing on YAAD.

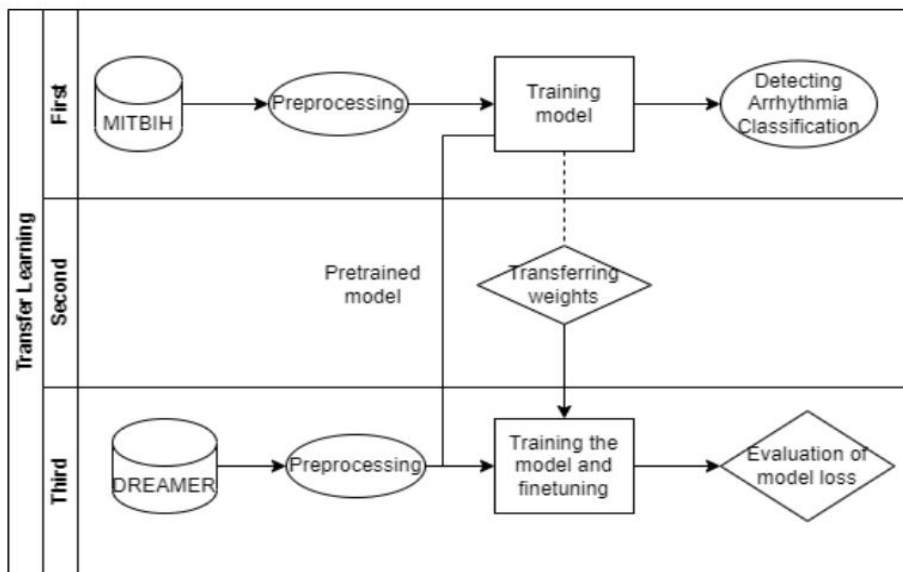
In the first experiment, we will be utilizing the Multispeed Architecture in which MIT-BIH is transfer learning. By reading and parsing ECG annotation files from the MIT-BIH ECG Database and visualizing data.

The feature labels for each beat as either normal (0) or abnormal (1), based on the annotation symbols in Figure 7. Detecting arrhythmia involves identifying abnormal beats in an ECG Signal (since arrhythmia is characterized by abnormal heart rhythms). During arrhythmia the pattern of beats becomes irregular. Preprocessing ECG Signal and annotation. The data is randomly split into training and testing sets based on a predefined ratio, such as 70% for training and 30% for testing. The scikit-learn library is used to split the input data and target labels into training and testing sets. Then a categorical function from the TensorFlow library is used to convert the target labels. This is a common technique for multi-class classification problems. The model is compiled with the categorical cross entropy loss function, the Adam optimizer, and metrics including accuracy. The model is fit to the training data with early stopping criteria and learning rate scheduling callbacks included. The training history is stored in history. The best model weights are loaded using model (load weights). The trained model is used to make predictions on the testing data, and the classification report was retrieved from model console using classification report from scikit-learn.



**Figure 6** Representation of MIT BIH ECG Data representing the different beats

At the end, training and validation accuracy over epochs is obtained from the model. After transfer learning, it continues by fine-tuning the model on DREAMER in figure 8. Because transfer learning requires large amounts of data to train the model and fine-tuning can be smaller data comparatively. The dataset is loaded from a pickle file that contains raw ECG signals collected.

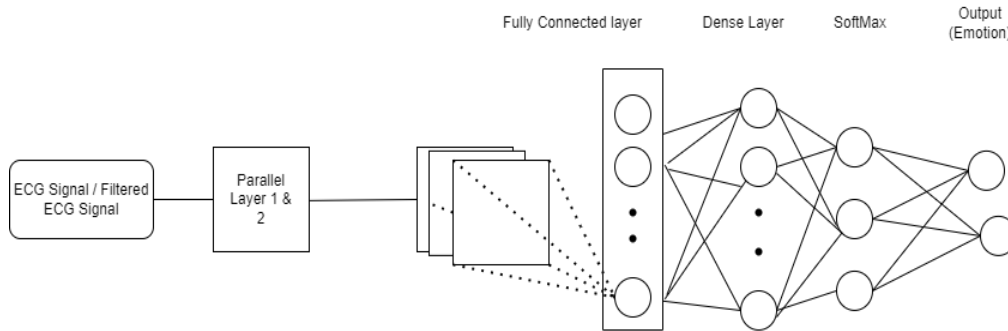


**Figure 7** Transfer learning and fine-tuning model on MITBIH & DREAMER database

The ECG signals are first filtered and then the R-peaks are detected using the biosppy (for processing biomedical signals). A label with the corresponding emotional state of the participants (valence, arousal, and dominance). The segmented beats are then saved into a list of the input features, and the emotional labels are saved as the output. The pre-trained model (MIT-BIH) is then fine-tuned on the DREAMER dataset by adding a fully connected layer on top of the output of the model and training the entire network end-to-end using the DREAMER dataset. This fine-tuning process allows the network to use the pretrained model to extract features from the ECG signals that are specifically relevant to the task of emotion recognition. The model is trained in Figure 9 with a parallel layer and fully connected layer, which is trained for a 50 number of epochs, with a checkpoint

callback that saves the weights of the best model based on the validation loss. After training, the trained model is used to make predictions on the test data, and the performance of the model is evaluated using various metrics such as mean squared error, mean absolute error, and mean absolute percentage error. Training on the Dreamer and testing with YAAD dataset to classify emotion classes (Valence, Dominance, Arousal) which are compatible.

Taking into consideration the YAAD dataset for testing. Initially dividing it into training and validation sets. Then, the Multi-speed Transformer was used and trained the model on the YAAD dataset using the training set.



**Figure 8** Model Architecture of ECG Signal with dense layer and output layer

The validation set was used to monitor the training progress and adjust the model's parameters as needed. Table I shows the training parameters that were used in training the model. Calling a model, using the architecture of the model, and its weights. We used to generate dataset, to increase the samples and comparatively check our model performance better with new data as well.

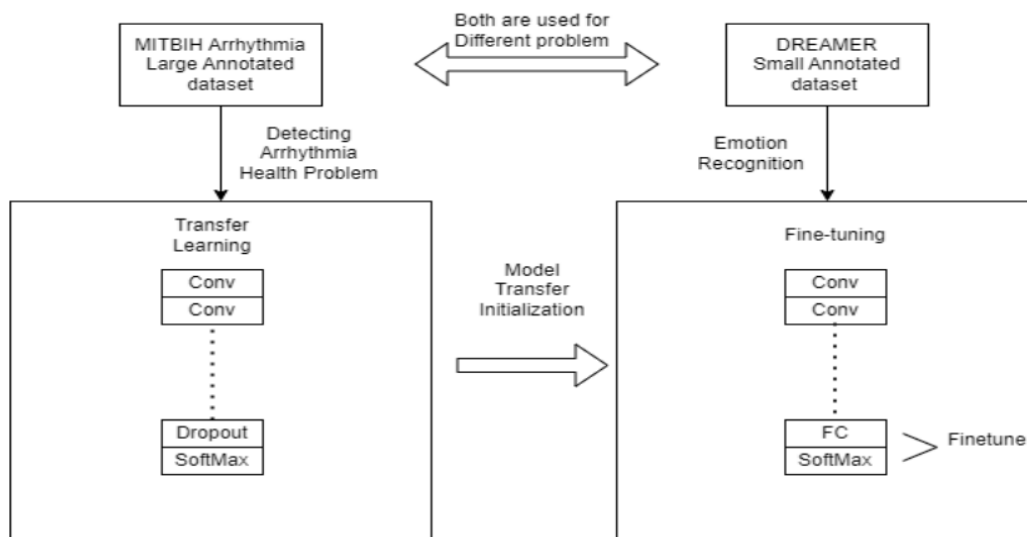
Learning rate was defined with Reduce Learning Rate on plateau which schedules the loss function for the model and evaluates till the training ends. Batch size the number of samples that will be propagated through the network. It takes the featured batch size samples from the training dataset and train network. Epochs are hyperparameters that specify how many times the learning algorithm will run over the full training dataset.

In experiment 1: Defining the model architecture (Multispeed Transformer described in section 3.5.1) and reshaping the training and testing data. By creating a model function which used to define the Transformer architecture of the deep learning model, based on the input shape of the pre-processed data. The Model Checkpoint callback is defined to save the best model weights based on the validation accuracy. Calling a model, using the architecture of the model, and its weights. We used to generate dataset, to increase the samples and comparatively check our model performance better with new data as well.

**TABLE I.** Representation of Model Optimizers for transfer learning and fine tuning

Parameter	MIT BIH	DREAMER
Total	1,122,018	292,995
Trainable	1,122,018	292,995
non-trainable	0	0

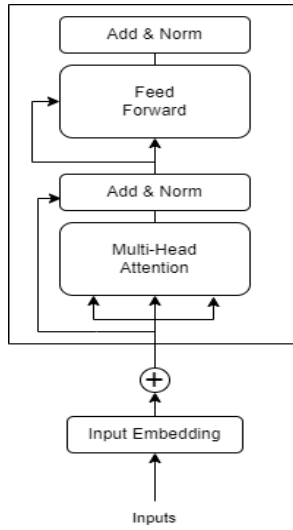
The model is trained with the following additional parameters, shuffle (true) shuffles the training data before each epoch, which can help prevent overfitting. Verbose is set to 2 shows a progress bar during the training process. Model checkpoint callback which will save the best model based on validation loss. The callbacks are used to perform some tasks during training, it can be saving the best model, early stopping, learning rate schedule etc. Early stopping is assigned to stop the training when the monitored quantity has stopped improving. And learning Scheduler to schedule the learning rate.



**Figure 9** Pre-trained model (TCN) is loaded on a new task, and the weights are frozen.

In Table I Trainable params: Trainable parameters are the values within a model that can be adjusted during training to minimize the error between the model's predictions and the true output. These values are also known as "weights" or "learnable parameters."

Non-trainable: the number of weights that are not updated during training with backpropagation as in our case(zero). It is the list of those that aren't meant to be trained. Typically, they are updated by the model during the forward pass.



**Figure 10** Pre-Model Architecture of ECG Signal with dense layer and output layer output layer

Adam optimizer was used (faster computation time, and required fewer parameters for tuning), with a learning rate of 0.00005 and the batch size of 128. The model was trained till 50 epochs, using the pretrained model, i.e., this first setting guaranteed that the data present in the training set were not in the testing set. The best model weights are loaded using model load weights. The trained model is used to make predictions on the data, and the classification report was retrieved from model, console using classification report from scikit-learn. Finally, training and validation accuracy over epochs is obtained from model.

In experiment 2: The implementation of a detecting arrhythmia based on ECG signals using a Vanilla Transformer architecture. In this experiment, Transfer learning was performed on MIT BIH Arrhythmia ECG database, and the saved model which is used for fine tuning on new task. The database takes input data (ECG Signal) and output of two classes (Normal and Abnormal beats) which builds a model using the TensorFlow and Keras libraries. The model is trained by transformer build model, which takes in several parameters such as input shape, head size, number of heads, feed-forward dimension, number of transformer blocks, MLP units, dropout, and number of classes. It takes the input features as X and target labels (Normal beat or Abnormal beat) as Y and splits them into two sets: X train, y train for training the model, and X test, y test for evaluating the model's performance.

**TABLE II.** Experiment 2 represents the Trainable params and non-trainable

Parameter	MIT BIH	DREAMER
Total	4,486,218	3,783
Trainable	4,486,218	3,783
non-trainable	0	0

The training of a model in which fit to train the model for 50 epochs with a batch size of 128. The validation data is passed to the function to evaluate the model's performance. Initially MIT-BIH model weights were transferred as the initial weight for the model and added a layer on top for feature dataset DREAMER (fine-tuning). As the Vanilla Transformer was used in this experiment, the transformer encoder which builds the transformer also is defined as a transformer encoder, which takes in inputs and several hyperparameters. The function applies multi-head attention in transformer

architecture and layer normalization to the inputs and includes a feed-forward part with convolutional layers. The build model function takes in several parameters in Table II including input shape, hyperparameters for the transformer encoder, and parameters for a multi-layer perceptron (MLP) that will be used as the final output layer. It creates an input layer and applies the transformer encoder in a loop for the specified number of transformer blocks in Figure 11. The output of the transformer encoder is then passed through the MLP to produce the final output.

In experiment 3: Temporal Convolution Network, we used a loop to calculate the receptive field of the TCN, which defined the number of time-steps the model can see in each direction. By defining the TCN model using the functional API (Keras), with an embedding layer, a TCN layer with 64 filters, and a SoftMax as output layer. The input is ECG channel and output has a label (Normal as 0, Abnormal beat as 1). Training the model, we used the Model Checkpoint callback from callbacks function to save the weights of the best model based on the validation accuracy. By saving the weights can be used to finetune the pretrained model on feature dataset DREAMER. Also used early stopping with a patience of 300 epochs and a learning rate scheduler with a factor of 0.5, a patience of 5, a cooldown of 5, and a minimum learning rate of 5e-6. The inputs and outputs of the TCN model are then used to construct a network that is trainable. Finally, the new model is compiled using the Adam optimizer and mean squared error loss and trained using the extracted beats and their associated labels. During training, several callbacks are used, such as model checkpoint, early stopping, and a learning rate scheduler, to improve the accuracy of the model and prevent overfitting. Once training is complete, the best weights are saved and used for prediction on the test set.

Overall, the process of fine-tuning TCN model on a new dataset involves freezing the weights of the pre-trained model and using its output as input to a new model. Finally, the evaluation metrics are obtained and compared for the performance of the model with the test set YAAD. The model takes feature dataset MIT BIH transfer learning and fine tuning on DREAMER and testing on YAAD in which output has emotion classes compatible. By defining some parameters for the model such as the number of filters and kernel size, it also defines callbacks such as Model Checkpoint and Early Stopping. Then define the input for the model and create an instance of the TCN layer with the defined parameters.

The model monitors loss for each epoch. Reduce Learning rate on plateau scheduler has been used to understand the behavior of models with data. Model checkpoints keep the model that has achieved the "best performance" so far, or whether to save the model at the end of every epoch regardless of performance. As it monitors whether it should be maximized or minimized. The model is compiled, compile() includes required losses and metrics. Finally, performing prediction and evaluation metrics has been used. The model is compiled with Adam optimizer, mean squared error loss, and metrics. The model is then fitted on the training data with specified number of epochs, batch size and validation data.

**TABLE III. (Experiment 3) Representation of MIT BIH and DREAMER Parameters**

Parameter	MIT BIH	DREAMER
Total	144,290	190,403
Trainable	144,290	190,403
non-trainable	0	0

## 6. Results

The experiment results are carried out for three architectures individually to evaluate and identify the performance of the model. The evaluation metrics like Mean Absolute Error (MAE) and Mean Squared Error (MSE) are used to analyze the performance of the model and compare it with other

architectures (Experiment 1: Multispeed transformer, Experiment 2: Vanilla transformer, Experiment 3: Temporal Convolution Network).

Experiment 1 – (i) Pretraining result: The result of the model for MIT-BIH Arrhythmia ECG classification. The result for ECG classification was achieved by the dataset MIT-BIH detecting abnormal beats occurring on or not on ECG signal. Accuracy (Table IV) was calculated for the train and test set in model. The results indicate that the model performs well on both classes, with high precision, recall, and F1- score values. The macro-average and weighted average metrics are also high, indicating good overall performance of the model. Accuracy: the fraction of correctly predicted labels among all instances. The accuracy score of 0.98 means that 98% of the predictions made by the model were correct. Macro Average: the average precision, recall, and F1-score across all classes. Weighted Average: the average precision, recall, and F1-score weighted by the number of instances in each class.

**TABLE IV. Accuracy of model after training with precision, recall and f1-score**

	Precision	Recall	F1-Score
Accuracy			0.98
Macro Average	0.98	0.97	0.98
Weighted Average	0.98	0.98	0.98

(ii) Fine Tuning result: This experiment conducted metric evaluation in Table V with the results obtained for DREAMER Database. The result of emotion recognition with the transfer learning (Arrhythmia ECG Classification) weights are loaded and the results achieved. Mean Absolute Error (MAE): A lower MAE value indicates a better fit. In this case, the MAE value is 0.93, which means that on average, the model is off by 0.93 units from the actual values. Mean Squared Error (MSE): In this case, the MSE value is 1.26, which means that on average, the model is off by 1.11 units from the actual values (iii)Testing results: The result of the model with the Architecture of Transformer. The result for emotion recognition was achieved by the dataset YAAD (Table V). After Training models on YAAD, test models with DREAMER and compare the results. Mean Absolute Error (MAE): value is 2.58, which means that on average, the model is off by 2.58 units from the actual values. Mean Squared Error (MSE): The loss of 9.80, which means that on average, the model is off by 3.13 units from the actual values.

**TABLE V. Loss function results**

Database	MAE	MSE
DREAMER	0.93	1.26
Macro Average	0.98	0.97

Experiment 2 – (i) Pretraining result with MIT-BIH Arrhythmia ECG classification was achieved by the model detecting whether abnormal beats occur or not on ECG signal. Accuracy (Table VI) was calculated for train and test set in model. The results indicate that the model appears to have performed very well in terms of accuracy and the various precision, recall, and F1-score measures. The accuracy of the model is reported to be 0.97, which means that the model was able to correctly predict the class of 97% of the instances in the test set. The macro average measures for precision, recall, and F1-score are also very high, with values of 0.97, 0.96, and 0.96 respectively. It’s calculated as the average of the precision, recall, or F1-score for each class, without considering class imbalance. The weighted average measures for precision, recall, and F1-score are also high, with values of 0.97 respectively and it’s calculated as the average of the precision, recall, or F1- score for each class, weighted by the number of instances in each class. Overall, the evaluation metrics suggest that the model was able to accurately classify the instances and performed well across all classes.



**TABLE VI. Accuracy of model after training with precision, recall and f1-score**

	Precision	Recall	F1-Score
Accuracy			0.97
Macro Average	0.97	0.96	0.96
Weighted Average	0.97	0.97	0.97

(ii) Fine-Tuning results evaluating the performance of models with loss function respectively. The result obtained (Table VII), Mean Absolute Error (MAE): A lower MAE value indicates a better fit. In this case, the MAE value is 0.98, meaning that, on average, the model is off by 0.98 units from the actual values. Mean Squared Error (MSE): The MSE value obtained was 1.39, which indicates that, on average, the model is off by 1.18 units from the actual values.

(iii) Testing results: In this experiment, vanilla architecture has been used and evaluated the performance of a model. Loss function and the performance of a model using YAAD Dataset. The results are obtained (Table VII) and tested with DREAMER in which both output (emotion classes) is compatible. MAE and MSE are both error metrics. MAE is the average of the absolute differences, while MSE is the average of the squared differences. The values for these metrics are 3.01 and 12.80, respectively. From the values provided, the MAE value is relatively low, indicating that the model's predictions are close to the true values on average. However, the MSE value is relatively high, indicating that the model's predictions diverge more from the true values for some instances.

**TABLE VII. Loss function of model testing**

Database	MAE	MSE
DREAMER	0.98	1.39
YAAD	3.01	12.80

Experiment 3 – (i) Pretraining results for the MIT-BIH Arrhythmia ECG classification model are shown in Table VIII. The dataset MIT-BIH determined whether irregular beats occur in the ECG signal to get the desired result for ECG classification (Classification). Accuracy was calculated for the train and test set in the model. The results indicate that the model appears to have performed very well in terms of accuracy and the various precision, recall, and F1-score measures. The accuracy of the model is reported to be 0.70, which means that the model was able to correctly predict the class of 70% of the instances in the test set. It takes training time of 153 minutes (2.5 Hours) to complete its training for 50 epochs. The macro average is calculated as the average of the precision, recall, or F1-score for each class, without considering class imbalance. The weighted average is calculated as the average of the precision, recall, or F1-score for each class, weighted by the number of instances in each class. The model showed promising results, with high precision, recall, and F1-score values across all classes. (ii) Fine-tuning results (Table IX) that is trained using the "Dreamer" dataset in which the model's performance is evaluated using the following metrics: Mean Absolute Error (MAE): The MAE is achieved the error of 2.94, meaning that, on average, the model is off by 2.94 units from the actual values. Mean Squared Error (MSE): This metric is like MAE, but it gives more weight to larger errors. The lower the value of MSE, the better the model is performing. In this case, the MSE is 10.06, which means that on average, the model is off by 3.17 units from the actual values (square root of 10.06).

**TABLE VIII. Accuracy of model after training with precision, recall and f1-score**

	Precision	Recall	F1-Score
Accuracy			0.70
Macro Average	0.72	0.54	0.49
Weighted Average	0.71	0.70	0.61

(iii) Testing results (Table IX), "YAAD" feature dataset has been utilized. The model's performance is evaluated using several evaluation metrics: Mean Absolute Error (MAE) and Mean Squared Error (MSE). The MAE value is 3.351, which indicates that on average, the model's predictions are 3.351. The MSE value is 22.05, which is a measure of the average of the squares of the errors. It is generally used to indicate how far the predictions deviate from the true values. MAE and MSE are relatively high which means that the model is not performing well.

**TABLE IX. Loss function of model testing**

Database	MAE	MSE
DREAMER	2.94	10.03
YAAD	3.351	22.03

## 7. Discussion

We delve into the research, experiments, and various considerations involved in our project. Our research on emotion recognition based on physiological signals. We begin by exploring the possibilities of using pretrained models on feature data sets, including heart rate and emotion scales. We then leverage these pretrained weights as the initial weights for a new neural network. This transfer learning approach is inspired by the concept of using a pretrained model for detecting arrhythmia on ECG data and subsequently applying it to emotion recognition within the DREAMER dataset. Our goal is to apply this network to both the DREAMER and YAAD datasets to classify emotions. These topics align with our research questions and will be thoroughly discussed in this chapter.

- Transfer Learning with MIT-BIH Arrhythmia Database:

We draw inspiration from the MIT-BIH Arrhythmia Database, a widely recognized dataset in this research domain. This dataset boasts over 100,000 ECG recordings collected from a diverse group of patients. Its substantial size facilitates a more in-depth analysis and ensures the robustness and reliability of any deep learning models developed. By fine-tuning our model on DREAMER, which is relatively smaller in comparison to MIT-BIH, we can reap several advantages.

- Advantages of Using DREAMER:

**Task Specificity:** Smaller datasets like DREAMER are often tailored to a specific task or domain, making them more conducive to targeted research. DREAMER, for instance, is designed explicitly for emotion recognition and incorporates multimodal data, including physiological signals, audio, and video recordings. This focus streamlines the model's ability to learn pertinent features and patterns associated with emotion recognition, eliminating the need to sift through a larger, more diverse dataset.

Efficiency: Smaller datasets are more manageable in terms of computational resources and time. Training deep learning models on extensive datasets can be computationally intensive and time-consuming, posing practical challenges for many researchers and institutions. Leveraging a smaller dataset like DREAMER enables us to train and fine-tune models more efficiently.

Overfitting Mitigation: Overfitting, the phenomenon where a model becomes overly specialized to training data and struggles to generalize to new data, is less of a concern with smaller datasets. These datasets offer fewer examples for the model to memorize, forcing it to generalize from a more restricted data pool.

- Importance of Cross-Dataset Evaluation:

When assessing the performance of a deep learning model for a specific task, it is crucial to evaluate its performance across multiple datasets. Relying solely on a single dataset can be misleading, as a model that excels on one dataset may struggle to generalize to new and unseen data. Therefore, our research emphasizes the evaluation of the model using various metrics, optimizing its performance, and scrutinizing the effectiveness of transfer learning, fine-tuning, and testing, particularly when applied to the YAAD dataset.

As this research involves a meticulous exploration of emotion recognition using physiological signals, transfer learning from the MIT-BIH Arrhythmia Database to DREAMER, and cross-dataset evaluation to ensure the robustness and generalizability of our model. The advantages of working with a smaller, task-specific dataset like DREAMER, combined with a thorough evaluation process, contribute to the reliability and applicability of our research findings.

## 8. Comparison of the results

Table IX is the loss function of this experiment, calculated evaluation metrics MAE and MSE, with the two different physiological signal datasets (DREAMER and YAAD). There are indications that using a transfer learning model can achieve a better result and reduce cost than using the trained model which takes too long duration and cost. However, Pretrained model has more trainable parameters than that of the transfer learning model because it takes some of the trainable layer and can be retrained with the model.

**TABLE X. Transfer Learning (Loss Function) of different architecture**

Architecture	MAE	MSE
Multispeed Transformer	0.93	1.26
Vanilla Transformer	0.98	1.39
Temporal Convolution network	2.94	10.06

Having more training parameters will increase both training time and evaluating error for each iteration. In many cases, the transfer learning model takes less time to complete its training and is less expensive. However, this also means that the training process will take longer, as the model needs to fine-tune its weights to the new task. Additionally, transfer learning requires a lot of computational resources, which can also contribute to less training time. The results provided for three different models: Multispeed Transformer, Vanilla Transformer and Temporal Convolutional Network, applied to a dataset for emotion recognition task and testing on Feature Dataset YAAD to DREAMER. The

performance difference could be the use of the MIT-BIH dataset as a pre-trained model for the Multispeed Transformer architecture. As mentioned earlier, MIT-BIH dataset is a large dataset of electrocardiogram (ECG) recordings, which can be helpful for tasks related to ECG analysis. Based on the collected results, the multispeed transformer and 87 vanilla transformer perform well. By using this pre-trained model, the Multispeed Transformer architecture may have been able to learn more effective representations of the DREAMER dataset. On the other hand, Multispeed Transformer's better performance could be the use of multispeed attention, which allows the model to attend to different parts of the input sequence at different speeds. This can be especially useful for tasks where different parts of the input sequence may be more important than others. In contrast, the Vanilla Transformer has not been to effectively learn from the DREAMER dataset without the benefit of pre-training on the MIT-BIH dataset, and the Temporal Convolutional Network also has not been as effective at capturing temporal dependencies in the data. Overall, the Multispeed Transformer's performance is likely to be attributed to a combination of factors, including the use of pre-training on the MIT-BIH dataset and the use of multispeed attention help in achieving best result. The results are compared with the testing approach as YAAD database has been trained with three architectures and obtained results. As (Table XI) result is achieved by YAAD dataset with different architecture and test with DREAMER (transfer learning). Based on the comparison between Table X and XI, the transfer learning approach appears to have achieved better results than the test approach for emotion recognition. The Multispeed Transformer model achieved the best results in the transfer learning approach with a MAE of 0.93, and MSE of 1.26. In contrast, the best-performing model in the test approach is the Multispeed Transformer, with a MAE of 2.58, and MSE of 9.80.

**TABLE XI. Loss Function of different architecture YAAD (Test)**

Architecture	MAE	MSE
Multispeed Transformer	2.58	9.80
Vanilla Transformer	3.01	12.80
Temporal Convolution network	3.35	22.05

It clearly demonstrates that comparing the test set results (YAAD) with transfer learning results, the loss is higher comparatively. It's important to note that the test approach is only evaluating the models on the YAAD dataset, whereas the transfer learning approach is training the models on the dataset and then fine-tuning them on the DREAMER dataset. Therefore, the transfer learning approach has an advantage in that it can leverage the knowledge learned from the MIT-BIH dataset and apply it to the DREAMER dataset. However, the test approach may be more suitable in cases where there is no pre-existing dataset that can be used for transfer learning. In such cases, the model must be trained from scratch on the target dataset. Nonetheless, based on the research, the transfer learning approach appears to have outperformed the test approach for emotion recognition on the DREAMER compared to YAAD datasets.

## 9. Conclusions

This paper contributes to the fields of both affective computing and deep learning, specifically the application of transfer learning techniques. The intention was to investigate a possible way to improve emotion recognition using physiological signals and to analyse the performance of different models. A three-dimensional emotion model of valence, arousal and dominance was used to classify seven basic emotions using neural networks. Three experiments were attempted, generating the physiological signal from a given sample input and smaller dataset to increase the data samples, and performing the transfer learning to transfer the trained parameter to a new task and the model to predict the emotion and compare the different architectures. The task of emotion recognition from electrocardiogram (ECG) data was investigated using three different datasets: MIT-BIH, DREAMER and YAAD.

To improve the performance of our machine learning models, we used transfer learning by pre-training them on the MIT-BIH dataset, fine-tuning them on the DREAMER dataset, and testing them on the YAAD dataset. We used three different architectures: multispeed transformer, vanilla transformer, and temporal convolutional network, and evaluated their performance using mean absolute error (MAE) and mean squared error (MSE). The results of the experiments show that the multispeed transformer and vanilla transformer architectures performed better than the temporal convolutional network architecture, achieving lower values for MAPE, MAE, and MSE. Additionally, we found that the use of transfer learning improved the performance of our models on the YAAD dataset compared to training them from scratch on this dataset. Specifically, the achieved result is lower values of MAE and MSE for the DREAMER dataset than for the YAAD dataset, indicating that the fine-tuning step helped the models better adapt to the target dataset. Comparing the performance of the models on the DREAMER and YAAD datasets, we found that the results for the multispeed transformer and vanilla transformer architectures were comparable between the two datasets, with similar values for MAE and MSE. This suggests that these architectures are robust and can generalize well to new datasets with compatible output classes, such as DREAMER and YAAD.

One of the challenges of the research has been the fine-tuning of the models in the database. The DREAMER database contains a variety of emotional stimuli and different modalities, which may require more complex models and longer training times. It is also a challenge to generalise the models to other datasets or real-world applications. In terms of the implications of the results, this research contributes to the growing literature on emotion recognition and arrhythmia detection using physiological signals. The use of deep learning and transfer learning techniques has shown promise in improving the accuracy of both tasks. However, further research can help to validate the performance of these models in real-world settings and identify the factors that influence their accuracy. The relationship between emotions and physiological signals could be explored in more depth. For example, how different emotions are associated with specific changes in physiological signals, and how this information can be used to improve emotion recognition accuracy.

In addition, examining how physiological signals vary across different populations, such as individuals with different cultural backgrounds or medical conditions, can also provide valuable insights into the use of physiological sensors for emotion recognition in diverse settings. In conclusion, while the current study provides a valuable foundation for the use of physiological sensors for emotion recognition, there are still many avenues for further research and improvement in this area.

## Acknowledgements

This work was partially supported by the project FAIR - Future AI Research (PE00000013), Spoke 6 - Symbiotic AI, under the NRRP MUR program funded by the NextGenerationEU.

Special thanks to Lorenzo Diomeda for his valuable reflections on the IntelliHearts project, which inspired this research team to engage in the field of affective computing.

## References

- [1] Kaur, S., & Kulkarni, N. (2021). Emotion recognition-a review. *International Journal of Applied Engineering Research*, 16(2), 103-110. URL:[https://www.ripublication.com/ijaer21/ijaerv16n2\\_04](https://www.ripublication.com/ijaer21/ijaerv16n2_04).
- [2] Simran Kaur, Richa Sharma, Emotion AI: Integrating Emotional Intelligence with Artificial Intelligence in the Digital Workplace, *Innovations in Information and Communication Technologies (IICT-2020)*. URL:[https://link.springer.com/chapter/10.1007/978-3-030-66218-9\\_39](https://link.springer.com/chapter/10.1007/978-3-030-66218-9_39). doi: [https://doi.org/10.1007/978-3-030-66218-9\\_39](https://doi.org/10.1007/978-3-030-66218-9_39)
- [3] Jia, S., Wang, S., Hu, C., Webster, P. J., & Li, X. (2021). Detection of genuine and posed facial expressions of emotion: databases and methods. *Frontiers in Psychology*, 11, 580287. URL: <https://www.frontiersin.org/articles/10.3389/fpsyg.2020.580287>. doi: <https://doi.org/10.3389/fpsyg.2020.580287>.
- [4] Jerritta, S., Murugappan, M., Wan, K., & Yaacob, S. (2013, September). Emotion detection from QRS complex of ECG signals using hurst exponent for different age groups. In *2013 Humaine association*

- conference on affective computing and intelligent interaction (pp. 849-854). IEEE. URL: <https://ieeexplore.ieee.org/abstract/>. doi: 10.1109/ACII.2013.159
- [5] Tracy, J. L., Randles, D., & Steckler, C. M. (2015). The nonverbal communication of emotions. *Current opinion in behavioral sciences*, 3, 25-30. URL: <https://www.sciencedirect.com/science/article/>. doi: <https://doi.org/10.1016/>
- [6] Kassam, K. S., & Mendes, W. B. (2013). The effects of measuring emotion: Physiological reactions to emotional situations depend on whether someone is asking. *PloS one*, 8(6), e64959. URL: <https://journals.plos.org/plosone/>. doi: 10.1371/journal.pone.0064959.
- [7] Rani, P., Liu, C., Sarkar, N., & Vanman, E. (2006). An empirical study of machine learning techniques for affect recognition in human–robot interaction. *Pattern Analysis and Applications*, 9, 58-69. URL: <https://link.springer.com/article/10.1007/s10044-006-0025-y>. doi: <https://doi.org/10.1007/s10044-006-0025-y>.
- [8] Pantano, E., & Scarpi, D. (2022). I, robot, you, consumer: Measuring artificial intelligence types and their effect on consumers emotions in service. *Journal of Service Research*, 25(4), 583-600. URL: <https://journals.sagepub.com/doi/pdf/>. doi: <https://doi.org/10.1177/1094670522110353>
- [9] He, L., Hou, W., Zhen, X., & Peng, C. (2006, October). Recognition of ECG patterns using artificial neural network. In *Sixth international conference on intelligent systems design and applications* (Vol. 2, pp. 477-481). IEEE. URL: <https://ieeexplore.ieee.org/abstract/document/>. doi: 10.1109/ISDA.2006.253883
- [10] Revina, I. M., & Emmanuel, W. S. (2021). A survey on human face expression recognition techniques. *Journal of King Saud University-Computer and Information Sciences*, 33(6), 619-628. URL: <https://www.sciencedirect.com/science/article/pii/S1319157818303379>. doi: <https://doi.org/10.1016/j.jksuci.2018.09.002>.
- [11] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248-255). Ieee. URL: <https://ieeexplore.ieee.org/abstract/document/>. doi: 10.1109/CVPR.2009.5206848.
- [12] Salai, M., Vassányi, I., & Kósa, I. (2016). Stress detection using low-cost heart rate sensors. *Journal of healthcareengineering*, 2016. URL: <https://www.hindawi.com/journals/jhe/2016/5136705/>. doi: <https://doi.org/10.1155/2016/5136705>
- [13] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30. URL: <https://proceedings.neurips.cc/paper/2017/file>
- [14] Wang, X., Ren, Y., Luo, Z., He, W., Hong, J., & Huang, Y. (2023). Deep learning-based EEG emotion recognition: Current trends and future perspectives. *Frontiers in Psychology*, 14, 1126994. URL: <https://www.frontiersin.org/articles/10.3389/>. doi: <https://doi.org/10.3389/fpsyg.2023.1126994>.
- [15] Poria, S., Majumder, N., Mihalcea, R., & Hovy, E. (2019). Emotion recognition in conversation: Research challenges, datasets, and recent advances. *IEEE Access*, 7, 100943-100953. URL: <https://ieeexplore.ieee.org/abstract/document/8764449>. doi: 10.1109/ACCESS.2019.2929050.
- [16] Cheriet, M., Dentamaro, V., Hamdan, M., Impedovo, D., & Pirlo, G. (2023). Multi-Speed Transformer Network for Neurodegenerative disease assessment and activity recognition. *Computer Methods and Programs in Biomedicine*, 107344. URL: <https://www.sciencedirect.com/science/article/> doi: <https://doi.org/10.1016/j.cmpb.2023.107344>
- [17] Lin, Y. P., & Jung, T. P. (2017). Improving EEG-based emotion classification using conditional transfer learning. *Frontiers in human neuroscience*, 11, 334. URL: <https://www.frontiersin.org/articles/10.3389/fnhum.2017.00334/full>. doi: <https://doi.org/10.3389/fnhum.2017.00334>.
- [18] Lin, T., Wang, Y., Liu, X., & Qiu, X. (2022). A survey of transformers. *AI Open*. URL: <https://www.sciencedirect.com/science/>. doi: <https://doi.org/10.1016/j.aiopen.2022.10.001>.
- [19] Salza, P., Schwizer, C., Gu, J., & Gall, H. C. (2022). On the effectiveness of transfer learning for code search. *IEEE Transactions on Software Engineering*. URL: <https://ieeexplore.ieee.org/abstract/document/>. doi: 10.1109/TSE.2022.3192755
- [20] He, Z., Zhong, Y., & Pan, J. (2022). An adversarial discriminative temporal convolutional network for EEG-based cross-domain emotion recognition. *Computers in biology and medicine*, 141, 105048. URL: <https://www.sciencedirect.com/science/article/pii/>. doi: <https://doi.org/10.1016/j.combiomed.2021.105048>.
- [21] Dentamaro, V., Impedovo, D., & Pirlo, G. (2021, January). Fall detection by human pose estimation and kinematic theory. In *2020 25th international conference on pattern recognition (ICPR)* (pp. 2328-2335). IEEE.

- [22] Impedovo, D., Dentamaro, V., Abbattista, G., Gattulli, V., & Pirlo, G. (2021). A comparative study of shallow learning and deep transfer learning techniques for accurate fingerprints vitality detection. *Pattern Recognition Letters*, 151, 11-18.
- [23] Convertini, N., Dentamaro, V., Impedovo, D., Pirlo, G., & Sarcinella, L. (2020). A controlled benchmark of video violence detection techniques. *Information*, 11(6), 321.