2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)

Emotion Recognition in the Wild from Long-term Heart Rate Recording using Wearable Sensor and Deep Learning Ensemble Classification

2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) | 978-1-6654-0126-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/BIBM52615.2021.9669553

Sara A. Nasrat
Department of Biomedical Engineering
Khalifa University
Abu Dhabi, United Arab Emirates
100058006@ku.ac.ae

M. Sami Zitouni
Health Engineering Innovation Center
Khalifa University
Abu Dhabi, United Arab Emirates
mohammad.zitouni@ku.ac.ae

Soowon Kang
Graduate school of Knowledge Service Engineering
Korea Advanced Institute of Science and Technology
Daejeon, South Korea
sw.kang@kaist.ac.kr

Uichin Lee
KI for Health Science and Technology
Korea Advanced Institute of Science and Technology
Daejeon, South Korea
uclee@kaist.ac.kr

Ahsan H. Khandoker
Health Engineering Innovation Center
Khalifa University
Abu Dhabi, United Arab Emirates
ahsan.khandoker@ku.ac.ae

Herbert F. Jelinek
Health Engineering Innovation Center
Khalifa University
Abu Dhabi, United Arab Emirates
herbert.jelinek@ku.ac.ae

***Abstract*— Long-term, continuous physiological recordings are currently being intensely investigated for tracking emotions. Emotional valence has been of more interest due to its relevance to cardiac and neurophysiological disease. In this research, multiple configurable convolutional neural networks (CNNs) were developed for different image-encoding techniques used as their input. Ensemble classification was then used to achieve a combined performance of the multiple CNNs by training a simple support vector machine (SVM) classifier using the last output layers of the CNNs as its input. Valence-labelled signals from the heart rate (HR) recorded using a wearable sensor from a wristband in a daily setting for one week from 80 participants were used for the image transforms. Accuracies of more than 91% were achieved with the classification ensembling, showing an improvement of the binary classification of emotional valence by more than 19% compared to using CNNs on their own.**

Keywords—convolutional neural network, deep learning, emotion recognition, emotional valence, ensemble classification, heart rate, wearable sensor

I. Introduction

Positive emotions help increase the performance of human health and function, whereas negative emotions can cause health issues [1]. Emotion recognition is a key component of affective computing. It essentially combines computer science, artificial intelligence (AI), cognitive neuroscience and psychology [2].

The relationship of physiology and psychology in emotions is complex and the precise and timely identification of human emotions, especially over an extended time remains the goal of psychological assessment [3]. Previous research using deep learning and CNNs did not investigate peripheral signals such as heart rate (HR) in depth compared to other types of signals such as electroencephalographic (EEG) recordings, especially for long-term data [4]. The aim of this research was to develop a robust configurable CNN model for long-term physiological signal analysis to identify emotional valence and to employ ensemble classification for improved performance.

II. Background

Image-encoding techniques for use with CNN's are not well-studied in emotion recognition research, with only one significant study highlighting a comparative overview of three image encoding techniques. Anjana, Ganesan and Lavanya [5] explored emotion recognition from EEG signals using a deep learning network trained with different image encoding techniques, where the EEG time series is transformed into an image using time-frequency (TF) analysis methods including spectrogram (based on Short Time Fourier Transform -STFT), scalogram (based on Continuous Wavelet Transform -CWT) and Hilbert Huang Transform (HHT) image representations. It was found that scalogram-based images resulted in the best emotion classification performance.

In another research about encoding time series as images for CNN classification tasks done by Wang and Oates [6], a novel image encoding framework of time series data using Gramian Angular Fields (GAF) and Markov Transition Fields (MTF) was proposed. Their methods preserve the signals' temporal dependencies which could entail relevant features that might be missed by other signal conversion methods. This could be seen by the improved classification results of their approach tested on multiple datasets compared to state-of-the-art approaches.

Time-frequency features have shown correlations with emotions from different types of signals throughout literature [7]. Ensemble classification has also shown improved emotion recognition performance as opposed to stand-alone classifiers in the study conducted by Setz, Schumm, Lorenz, Arnrich and Tröster [8].

III. Methods

A. Data Collection

Emotion data was collected at the Korean Advanced Institute of Science and Technology (KAIST) using the Experience Sampling Method (ESM), which is a technique of self-reporting emotional experiences in the real world [9]. The participant rated their emotional state in terms of valence on a

7-point scale at random time instances during waking hours, whilst a Microsoft Band 2 Smartwatch was recording the corresponding heart rate (HR) data. This continued for seven days for each participant, and there were 80 healthy participants in total of both males and females with ages ranging from 19 to 40.

B. *Workflow*

The flowchart in Fig. 1 describes the methodology which consists of first converting the labelled normalized HR time series (Fig. 2) into an image and then training a configurable CNN model to classify emotional valence. Finally, a Support Vector Machine (SVM) classifier is trained using the probabilistic outputs of the CNNs by ensembling their final layers to classify them into emotion labels of valence.

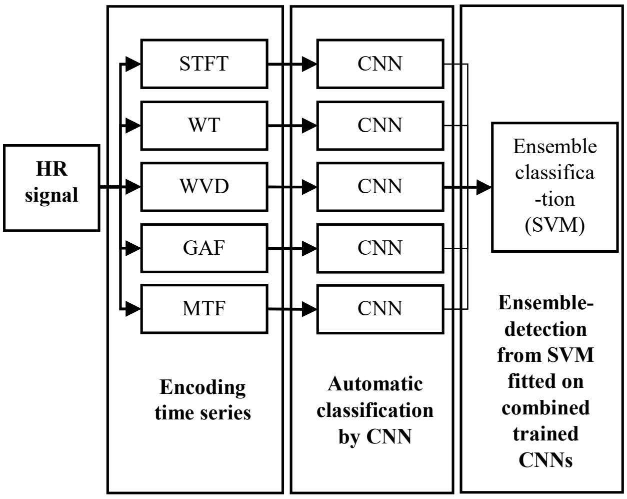
HR signal
STFT
WT
WVD
GAF
MTF
CNN
CNN
CNN
CNN
CNN
Ensemble classifica -tion (SVM)
Encoding time series
Automatic classification by CNN
Ensemble-detection from SVM fitted on combined trained CNNs

Figure 1 Flowchart of proposed system

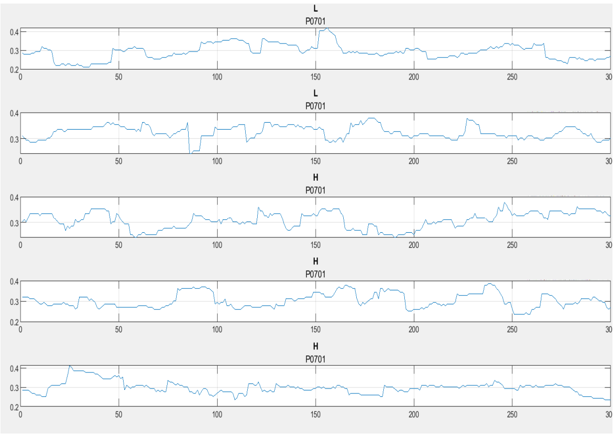

Figure 2 Five samples of normalized 5-minute-long HR time signals of High (H) and Low (L) valence labels.

C. *Image transformations*

One-hot encoding was applied to the EMS scores to obtain binary variables of the valence. Labelled HR signals were then transformed into five types of images including scalograms (Wavelet Transform -WT), spectrograms (short time Fourier transform -STFT), Wigner Ville distribution (WVD) images, Gramian angular fields (GAF) images and Markov transition fields images (MTF).

a) Spectrogram: the HR signal was divided into smaller overlapping time segments with a Hamming window. Calculating STFT can be done looking at (1).

$$X(t,f) = \int_{-\infty}^{\infty} x(t_1)\omega \times (t_1 - t)e^{-j2\pi f t_1} dt_1 \qquad (1)$$

where w(t) is a sliding window moved over the signal x(t), then the squared magnitude of the calculated STFT is found for spectral analysis [10].

b) Scalogram: an analytic Morse wavelet was used to analyze the HR time series with a time bandwidth of 60 and 10 wavelet bandpass filters per octave (10 voices per octave). Formula (2) computes CWT.

$$CWT(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\psi \times \left(\frac{t-b}{a}\right) dt \qquad (2)$$

where a is the scale, b is the time shift and ψ is the Morse wavelet [10].

c) WVD: the WVD provides a high Time-Frequency resolution using a method called instantaneous autocorrelation, calculated using (3).

$$W(t,f) = \int_{-\infty}^{\infty} R_{xx}(t,\tau)e^{-j2\pi f\tau} d\tau \qquad (3)$$

where R_{xx} is the instantaneous autocorrelation, t is the time point and τ is the lag value of the signal. A Kaiser window was used to minimize the cross terms in TF domain. for reducing the cross terms in time and frequency domains [10].

d) GAF: each GAF image is a representation of the temporal correlation between every time instance. (4-6) show the formulae used to calculate GAF summation.

$$\tilde{x}_i = a + (b - a) \times \frac{x_i - \min(x)}{\max(x) - \min(x)}, \ \forall_i \in \{1, \dots, n\} \ (4)$$

$$\phi_i = \arccos(\tilde{x}_i), \ \forall_i \in \{1, \dots, n\} \qquad (5)$$

$$GASF_{i,j} = \cos(\phi_i + \phi_j), \ \forall_{i,j} \in \{1, \dots, n\} \qquad (6)$$

where a matrix for each pair of points in time (x_i, x_j) is created from the HR signal. The time signal is rescaled in a range $[a, b]$ where $-1 \le a < b \le 1$ [6].

e) MTF: Markov transition matrix is computed by discretizing the time signal to bins, resulting in the Markov transition probabilities against temporal dependency in the two dimensions [6].

D. *CNN and SVM*

Figure 3 shows the architecture of the configurable CNN, consisting of 3 convolutional layers and two fully connected layers with a dropout layer in between. The final layer is where the image gets assigned a class based on the probability scores calculated by the softmax layer.

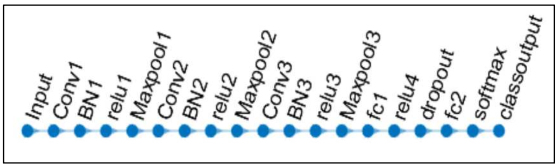

Figure 3 Layer graph of the CNN.

A support vector machine (SVM) classifier using a fast linear solver as the kernel function of the separating hyperplane was trained and fitted on the final layer output of the CNNs in an ensembling technique using the probability scores as arrays of input to the classifier, which significantly improved the classification performance. To validate the performance of the CNN models and SVM classifier, a five-fold cross validation was applied in training and testing the data (Table I).

IV. Results

TABLE I. Accuracy and F1 metrics for the Binary Classification of Valence using Heart Rate image input

Performance Evaluation	Image-encoding technique				
	WT	***STFT***	***WVD***	***GAF***	***MTF***
Testing accuracy %	70.14	64.37	70.25	68.86	72.32
F1 score %	81.33	76.95	81.54	80.83	83.34

The results show that an improved classification of the valence emotions from HR can be achieved when ensembling the CNNs by using their last output layer to train a simple SVM classifier as opposed to relying on the CNNs as standalone classifiers. The testing accuracy improved by up to more than 19 percent with ensemble classification (Fig. 3).

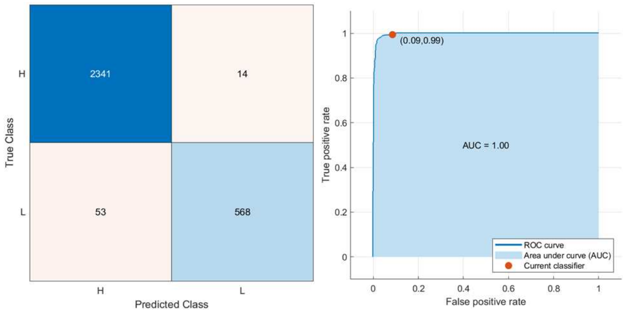

Figure 4 Confusion matrix and Area Under Curve (AUC) of ROC graph results of ensembling all five CNNs to classify binary valence using linear SVM (accuracy 97.7%).

V. Conclusion

A method to classify emotional valence from heart rate signals measured using wearable sensors for one week was proposed using configurable CNNs with different image transforms as input and improved by ensemble classification using SVM classifier. Emotional valence monitoring in daily life can support applications in physical and mental health awareness and emotional intelligence monitoring to improve treatment outcomes. The proposed method will be extended to study other signals such as galvanic skin response (GSR), with more image-encoding schemes, as such studies were not made before on long-term peripheral physiological signals in a day-to-day setting.

Acknowledgment

This work was completed in collaboration with the Korea Advanced Institute of Science and Technology (KAIST) of South Korea.

References

[1] Flynn M, Effraimidis D, Angelopoulou A, et al. Assessing the Effectiveness of Automated Emotion Recognition in Adults and Children for Clinical Investigation. Front Hum Neurosci. 2020;14:70. Published 2020 Apr 7. doi:10.3389/fnhum.2020.00070

[2] Shu L, Xie J, Yang M, et al. A Review of Emotion Recognition Using Physiological Signals. Sensors (Basel). 2018;18(7):2074. Published 2018 Jun 28. doi:10.3390/s18072074

[3] Dzedzickis A, Kaklauskas A, Bucinskas V. Human Emotion Recognition: Review of Sensors and Methods. Sensors (Basel). 2020;20(3):592.

[4] Feng K and Chaspari T (2020) A Review of Generalizable Transfer Learning in Automatic Emotion Recognition. *Front. Comput. Sci.* 2:9. doi: 10.3389/fcomp.2020.00009

[5] Anjana KA, Ganesan M, Lavanya R. Emotional Classification of EEG Signal using Image Encoding and Deep Learning. January 2021. doi:10.1109/ICBSII51839.2021.9445187

[6] Wang, Zhiguang & Oates, Tim. Encoding Time Series as Images for Visual Inspection and Classification Using Tiled Convolutional Neural Networks. Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015)

[7] P. Chettupuzhakkaran and N. Sindhu, "Emotion Recognition from Physiological Signals Using Time-Frequency Analysis Methods," *2018 International Conference on Emerging Trends and Innovations In Engineering And Technological Research (ICETIETR)*, 2018, pp. 1-5, doi: 10.1109/ICETIETR.2018.8529145.

[8] C. Setz, J. Schumm, C. Lorenz, B. Arnrich and G. Tröster, "Using ensemble classifier systems for handling missing data in emotion recognition from physiology: One step towards a practical system," *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, 2009, pp. 1-8, doi: 10.1109/ACII.2009.5349590.

[9] Myin-Germeys I, Kasanova Z, Vaessen T, et al. Experience sampling methodology in mental health research: new insights and technical developments. World Psychiatry. 2018;17(2):123-132. doi:10.1002/wps.20513.

[10] Scoll, S. Fourier, Gabor, Morlet or Wigner: Comparison of Time-Frequency Transforms. arXiv:2101.06707 [eess.SP]. Jan 2021.