Combining Deep Learning with Signal-image Encoding for Multi-Modal Mental Wellbeing Classification

KIERAN WOODWARD, Nottingham Trent University, UK
EIMAN KANJO, Nottingham Trent University, UK and Imperial College London, UK
ATHANASIOS TSANAS, University of Edinburgh, UK

The quantification of emotional states is an important step to understanding wellbeing. Time series data from multiple modalities such as physiological and motion sensor data have proven to be integral for measuring and quantifying emotions. Monitoring emotional trajectories over long periods of time inherits some critical limitations in relation to the size of the training data. This shortcoming may hinder the development of reliable and accurate machine learning models. To address this problem, this paper proposes a framework to tackle the limitation in performing emotional state recognition: 1) encoding time series data into coloured images; 2) leveraging pre-trained object recognition models to apply a Transfer Learning (TL) approach using the images from step 1; 3) utilising a 1D Convolutional Neural Network (CNN) to perform emotion classification from physiological data; 4) concatenating the pre-trained TL model with the 1D CNN. We demonstrate that model performance when inferring real-world wellbeing rated on a 5-point Likert scale can be enhanced using our framework, resulting in up to 98.5% accuracy, outperforming a conventional CNN by 4.5%. Subject-independent models using the same approach resulted in an average of 72.3% accuracy (SD 0.038). The proposed methodology helps improve performance and overcome problems with small training datasets.

CCS Concepts: • **Computing methodologies** → **Neural networks**.

Additional Key Words and Phrases: Affective Computing, Emotion Recognition, Gramian Angular Field, Transfer Learning, Artificial Intelligence

1 INTRODUCTION

Monitoring and quantifying emotional states can potentially enable people to improve their wellbeing and self management as they understand their life stressors. Computational methods to infer emotional states based on physiological and environmental measurements require further exploration as with recent advances in wearable and sensor technologies along with machine learning algorithms, the real-time monitoring, collection and analysis of multi-model signals is becoming increasingly possible. Various ubiquitous sensors can be capitalized on to monitor physiological changes that affect wellbeing. The use of these sensors to measure diverse data modalities including Heart Rate (HR), Heart Rate Variability (HRV) and Electrodermal Activity (EDA) may enable real-world emotion recognition as they directly correlate to the sympathetic nervous system [72], [70].

With the objective to have the most accurate affective classification model, multiple methods have previously been investigated starting from conventional signal processing modelling approaches to machine learning

Authors' addresses: Kieran Woodward, kieran.woodward@ntu.ac.uk, Nottingham Trent University, Nottingham, UK; Eiman Kanjo, eiman.kanjo@ntu.ac.uk, Nottingham Trent University, Nottingham, UK and Imperial College London, London, UK; Athanasios Tsanas, athanasios.tsanas@ed.ac.uk, University of Edinburgh, Edinburgh, UK.

2637-8051/2023/11-ART $15.00
https://doi.org/10.1145/3631618

algorithms, that extract features to allow the model to recognise different activities or to more recently deep learning algorithms that are trained to recognise different patterns to distinguish classes.

Most previously employed approaches used traditional supervised machine learning algorithms including popular classifiers such as support vector machines, decision-trees and K-nearest neighbor. While these models achieve very good results, a drawback of these approaches are that they entirely rely on the selection of features, meaning that a poor selection of features will result in a poor performing classification model, which will in turn result in a poor assessment of mental wellbeing. This is not desirable in when assessing wellbeing which requires an accurate evaluation to enable relevant support and interventions.

In the recent years, with the maturity of the deep learning algorithms, tremendous progress has been achieved in fields such as speech recognition and image classification. Advances in Deep Learning (DL) present new opportunities for the inference of mental wellbeing as models can be trained using raw data, alleviating the need for manual feature extraction which is often domain-driven and may be a time-consuming process. One of the models that achieved large success working with images are Convolutional Neural Networks (CNNs). CNNs have traditionally been used to classify 2D data such as images but these networks are also employed towards extracting features from 1-dimensional sensor data [27]. However, the performance of deep learning models deteriorates considerably when training data is scarce. This lack of sufficient statistical power often hinders the progress of machine learning applications in monitoring and understanding wellbeing, since collecting longitudinal and annotated training data is very challenging [45]. This is due to the following reasons:

(1) User availability, incentivisation and willingness to participate in longitudinal studies (or increasing study drop-outs beyond the first few months) [44]
(2) Privacy, ethics and data protection issues [104], [64]
(3) Data integrity and accuracy [52]
(4) Costs and availability of monitoring devices [93], [69]
(5) Requirement to set up the device and extract the data by expert personnel needing specialised equipment [98]
(6) Time consuming nature of real-time self labelling [95]

In order to address challenges 1 and 6, Transfer Learning (TL) is often used to reduce the amount of data required thereby reducing user dropout and the time consuming nature of studies. Pre-trained models can be used to encompass methods that discover shared characteristics between prior tasks and a target task, reducing the necessity for large datasets [61]. Many outstanding models that use CNN were developed over time, such as VGG, MobileNet, and ResNet. These models can be adapted to be used in other applications without the need for fully re-training them on the new database by employing transfer learning. Transfer learning is used to improve a learner from one domain by transferring information from a related domain. This process usually involves training a base model using labelled data from a different domain and transferring the knowledge to the new target domain [16], [61].

Inspired by these developments, many approaches have been taken in order to adapt time-series data inputs to CNN-based algorithms in order to improve accuracy. Previous research shows that fast changing, continuous sensor data such as accelerometer data can be transformed into RGB images which can then be used to train DL models [87] using algorithms to encode the data into images such as Gramian Angular Field (GAF). Although the premise of presenting time series data as images is promising in extracting multi-level features and improving classification accuracy, most of the previous work only considered encoding univariate time series data as one image for a single channel of a CNN input [87], [97], [40].

One of the main reasons for the success of deep learning (DL) is that DL models can represent the raw data well. This research proposes the combination of physiological sensor data with signal-encoded images in a TL model to tackle the challenging problem of monitoring the trajectory of wellbeing. To fix problems of existing models, we propose a novel feature extraction method based on time series imaging and transfer learning that can effectively extract affective data features.

The contribution of this work is that time-series data is encoded in an image and combined with additional sensor streams to facilitate translating the highest possible number of informative characteristics through data fusion. A new CNN-TL-based approach has been developed to alleviate many challenges when classifying small datasets. We explore the use of signal-image encoding to classify accelerometer data using three techniques; Gramian Angular Summation Field (GASF), Gramian Angular Difference Field (GADF) and Markov Transition Field (MTF) [87]. We propose using these images in a novel pre-trained TL model combined with a 1D CNN trained using physiological sensor data to infer mental wellbeing. The 2D images are converted to the RGB format in order to profit from pretrained models using transfer learning and adapting prior knowledge towards a new application which can potentially lead to overall performance improvement compared to only relying on training using the available data for the problem at hand. This framework uses TL in addition to signal-encoded images to improve on the performance of conventional deep learning methods for mental wellbeing prediction.

The remainder of the paper is organized as follows: Section 2 provides a review of mental wellbeing classification and TL; Section 3 describes and explores the dataset used; Section 4 describes the methodologies of data transformation and the model implementation; Section 5 shows the results, Section 6 presents the discussion and Section 7 presents the conclusion and suggestions for future research.

2 RELATED WORK

2.1 Models of Affect

Affect, in psychology, refers to the underlying experience of feeling, emotion or mood and is an integral aspect of human life [31]. There are many aspects to monitoring affective state including measuring emotions and stress levels being felt. Where mental health conditions are clinically diagnosed [63], emotions are defined as psychological states brought on by neurophysiological changes, variously associated with thoughts, feelings, behavioural responses, and a degree of pleasure or displeasure [18]. Similarly, moods are defined as affective states typically described as having either a positive or negative valence that in contrast to emotions, are less specific, less intense and less likely to be provoked or instantiated by a particular stimulus or event [11]. In contrast, mental wellbeing is defined as a state of well-being in which an individual realises his or her own abilities to cope with the normal stresses of life and can be impacted by emotions felt [91].

Experts in assessing psychology have developed different theories to classify emotions ranging from small groups containing items such as happiness and sadness [88] and pain and pleasure [57] to groups containing a larger number of emotions. There are no universal categories for emotions but the Ekman model [24] is commonly used, which comprises of 6 basic emotions: sadness, happiness, surprise, fear, anger and disgust, all of which can be distinguished through facial expressions.

Alternatively, emotions can be measured dimensionally. The two most commonly used dimensions are arousal (from calm to excited) and valence (from attractive to aversive). Russell [68] describes how the arousal and valence dimensions are defined in a circle called the circumplex model of affect that can encompass all emotions, as shown in Figure 1.

The Self-Assessment Manikin (SAM) Scale [15] has traditionally been used to measure valence, arousal and dominance (from submissive to dominant) using a 9-point pictorial representation of humans. This method can encourage engagement with its simpler approach, allowing for the quick assessment of affective state. However, this approach can be challenging when collecting real-time, real-world data as it requires the immediate completion of the scale, whenever a change in emotions is experienced. While each of the models can be useful to capture different aspects of mental wellbeing, this work focuses on using a simplified version of SAM to measure categorical states of mental wellbeing as it provides the greatest opportunity for real-world reporting.

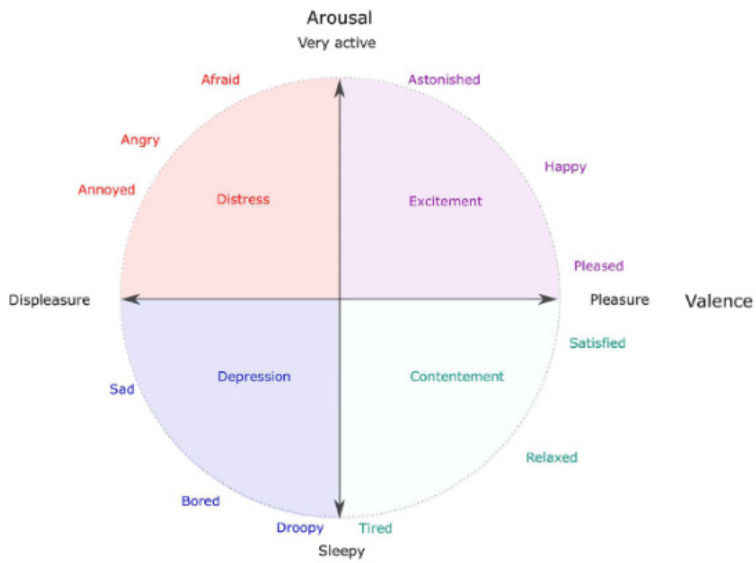

Fig. 1. Russell's circumplex model of affect.

2.2 Physiological Sensors to Monitor Mental Wellbeing

Numerous non-invasive physiological sensors can be used to assess real-world mental wellbeing including EDA and HR [72]. EDA correlates to the sympathetic nervous system [70] and can be used to detect mental wellbeing while HRV is the variation in time between heartbeats: as HRV is reduced the user is more likely to be stressed [29]. Previous work had investigated the use of embedded sensors within a wearable device that measured EDA and HRV during driving [29]. The device took five minute recordings of physiological sensor data enabling the model to predict stress at 97.4% accuracy. HRV and EDA were found to be very well statistically associated demonstrating these non-invasive sensors have the capability to accurately infer mental wellbeing.

Previous research has also developed a wearable device that measured ElectroCardioGram (ECG), EDA and ElectroMyoGraphy (EMG) of the trapezius muscles [92]. 18 participants wore the device while completing three stressor tasks with a perceived stress scale questionnaire completed before and after each task. Stressed and non-stressed states were classified with an average accuracy of almost 80%. However, as this study was conducted in a controlled environment it is not known how well the model would generalize in real-world environments, where physiological signals may be impacted by more than just stress.

Skin temperature has also been explored to infer stress as it indicates acute stressor intensity [30]. A further study investigated a wearable device that measured EDA, skin temperature and motion. The devices were provided to six participants with dementia for two months with the ground truth labels obtained from clinical notes [42]. Stress was then assigned into one of five integer levels where accuracy ranged from 9.9% to 89.4% between the levels while F1-Scores ranged from 1.4% to 26.8% demonstrating a high level of false positives and false negatives. The wide variation of accuracy is due to the low stress threshold as when the threshold was increased there were fewer classifications of stress thus increasing accuracy.

Once physiological data has been collected from the devices, different models must be explored to assign class labels. There are two main types of neural networks: Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). They are structurally different and are used for fundamentally different purposes.

CNNs have convolutional layers to transform data, whilst RNNs reuse activation functions from other data points. CNNs are feed forward networks that like RNNs do not require features to be first extracted although unlike RNNs they extract local and positional invariant features rather than different patterns across time.

Previously, RNN Long Short-Term Memory (LSTM) networks have been used to classify mental wellbeing as they capture long-term temporal dependencies. An LSTM network with a stack autoencoder to decompose the combined EEG signals was used to infer emotions from 32 participants. This approach of using the context correlations of the EEG feature sequences resulted in increased performance, achieving 81.1% accuracy [96]. Furthermore, previous work has fused raw EEG signals with videos of participants to improve model accuracy when inferring wellbeing [51]. The model achieved 74.5% accuracy by using temporal attention to ignore the redundant information. LSTM networks have also been used to classify EDA, skin temperature, motion and phone usage data to infer stress achieving 81.4% accuracy, outperforming other support vector machine and logistic regression models [83]. Raw EEG signals have been used to train a LSTM network achieving 85.45% in valence [4] and to classify stress in construction workers with 80.32% accuracy using a gaussian support vector machine model [37]. Dynamic Time Warping (DTW) has similarly been paired with K nearest neighbour machine learning classifier to infer emotions achieving between 65.6% and 94% across different datasets [3].

CNNs have also been used to infer mental wellbeing. A CNN has been trained to classify four emotions (relaxation, anxiety, excitement and fun) using EDA and blood volume pulse data [53]. DL algorithms were compared with standard feature extraction and selection approaches concluding DL outperformed manual ad-hoc feature extraction as it produced significantly more accurate affective models, even outperforming models that were boosted by automatic feature selection. Additionally, a CNN model using channel selection strategy has been trained using EEG data collected from 32 participants watching 40 1-minute excerpts of music videos to elicit emotions [65]. The channel selection strategy used the channels with the strongest correlation with valence to generate the training set. The model classified four possible emotions: (1) high arousal and high valence (2) high arousal and low valence (3) low arousal and high valence (4) low arousal and low valence. Using the EEG data, this channel selection approach achieved 87.27% accuracy, improving the accuracy by nearly 20%. A CNN and an RNN have been combined to allow raw data to be classified more accurately automating feature extraction and selection [39] [38]. Physiological, environmental and location data was used to train the model to infer emotions resulting in the combined model outperforming traditional DL models by over 20%. This work concluded that the CNN model matched or outperformed models with the features pre-extracted showing the benefits of DL.

Gramian Angular Fields have been used to measure emotions, however most of this work has focused on the use of EEGs and not wearable sensors such as HRV or EDA that can be used in the wild outside of controlled experiments [66], [100]. Furthermore, previous work shows that signal-image encoding techniques may not be best suited for fast-changing physiological data such as HRV or EDA [1]. Overall, while previous work has combined multiple physiological sensors for monitoring mental wellbeing [105], there has been little exploration of signal-image encoding techniques or the use of accelerometer data, which may hold the potential to further increase performance.

2.3 Transfer Learning

One of the biggest challenges in developing accurate DL models is the implicit practical requirement to collect a large *labelled* dataset. TL [61] is a common approach in machine learning to mitigate the problem occurring due to the scarcity of data. Caruana [16] introduced multi-task learning that uses domain information contained in the training signals of related tasks. It is based on the ability to learn new tasks relatively fast, alleviating the need for large datasets by relying on previous, similar data from related problems. TL capitalizes on a likely large dataset stemming from a related problem to pre-train a model, and subsequently adapt that model for the needs of a problem with a (potentially smaller) different dataset [85]. CNNs are commonly used in TL approaches, being initially trained on a vast dataset and then having the last fully-connected layer removed and further trained on a smaller target dataset. A pre-trained CNN alleviates the need for a large dataset while simultaneously decreasing the time required to train the model. The premise of TL is to improve the learning of a target task in three ways [80]: (1) improving initial performance, (2) producing sharp performance growth, (3) potentially resulting in higher training performance.

Inter-subject TL approaches using physiological signals have previously been used to detect driver status [46] and seizures [19]. Furthermore, an inter-subject TL approach has been used with ECG signals to infer mental state achieving 79.26% compared with a baseline of 67.90%, demonstrating the potential for TL to improve affective model performance with small physiological datasets [26]. Similarly, the possibility for TL to be used to personalise affective models has previously been explored and has helped personalise EEG signals, improving model accuracy by 19% [103] and 12.72% [48] using an inter-person approach while also reducing the amount of data required to train the models. However, relying on inter-subject TL approaches relies on an initial dataset with a vast number of users, which remains challenging to collect in real-world environments.

TL can be used to help alleviate scarce data as by using decision trees, data from similar subjects can be used to improve accuracy by around 10% although if data from dissimilar subjects is used it can have a negative impact on the model accuracy [54]. To ensure negative TL that degrades the performance of the model does not occur, a conditional TL framework has been developed that assesses an individual's transferability against individual's data within the dataset. The conditional TL model identified 16 individuals who could benefit from 18 individuals data within the EEG dataset, improving classification accuracy by around 15% [49].

TL has also been utilised in a subject-independent approach combining physiological features from ECG and EDA signals to train a CNN to classify stress [67]. The results show that TL resulted in an accuracy increase of between 1-4%. Transfer Learning has also been utilised with HR signals to improve the accuracy of stress recognition [2]. However, better results can often be obtained by combining multiple physiological signals for emotion recognition using a transfer learning approach [62]. This demonstrates previous work has explored the use of TL for affective classification using physiological sensors but does not show consideration of transforming time-series data to images and using a TL approach.

The inference of emotions from images and videos has also benefited from TL approaches. When using models pre-trained on the ImageNet dataset and testing using images of faces expressing seven emotions an accuracy of 55.6% was achieved compared with a baseline performance of 39.13% [59]. Additionally, audio and video have been explored to infer six emotions where the source task was gender classification and the target task was emotion classification as many features such as energy, frequency and spectral are similar across both gender classification and emotion classification. This TL approach improved base line accuracy by 16.73% [60].

Another TL approach trained a DBN structure on a large database designed for acoustic phoneme recognition and transferred the knowledge learned for PTSD diagnosis. This helped increase the classification accuracy of PTSD using speech by 13.5% [7]. Similarly, a sparse autoencoder-based feature TL approach has been developed to infer emotions from speech using the FAU AiboEmotion Corpus dataset [10]. The autoencoder approach to find a common structure in a small target base dataset and apply the structure to source data improved unrated average recall from 51.6% to 59.9% with only 50 data instances used [22]. Whispered speech has also been explored to infer emotions applying three TL approaches; denoising autoencoders, shared-hidden-layer autoencoders, and extreme learning machines autoencoders. Extreme learning machines autoencoders provide good generalisation extremely fast [34], enhancing the prediction accuracy on a range of emotion tasks achieving up to 74.6% arousal [21]. Speech has also been explored to improve PTSD diagnosis using TL and deep belief networks. The TL approach improved model accuracy from 61.53 to 74.99% [8]. Furthermore, deep belief TL networks have been used to improve the accuracy of emotion recognition through speech cross-language [47]. TL for emotion recognition has also been used to infer wellbeing from text [73]. As no large text-based affective datasets existed it was hypothesised that the social media domain, specifically the large amount of public tweets from Twitter, would be similar enough to transfer knowledge of content, style, and structure to the mental health domain. Therefore, an RNN with full weight transfer where the base model was trained using a Twitter dataset to classify tweets as positive or negative valence achieved an overall accuracy of 78% for all four classes where the standard RNN achieved 72%.

Hitherto, TL has most commonly been used to train images as large ImageNets have been used to developed pre-trained models such as VGGNet [74], Inceptionv3 [76] and mobileNetv3 [33] that contain pre-trained object classification models. The pre-trained CNN models were employed to compute mid-level image representations for object classification in PASCAL VOC images [25], leading to significantly improved results. TL has facilitated training new models in the visual domain using pre-trained CNNs [71]. However, modelling emotions using time series data such as HRV, HR, EDA or acceleration cannot be visually interpreted. Sensor data must first be transformed to translate the raw sensor data to images, for example using techniques such as GASF, GASF and MTF.

Recently TL has been utilised with signal-image encoding approaches such as GASF and GADF for post-stroke rehabilitation assessment, where it was shown that TL helped further increase model performance beyond that of only signal-image encoding [13] to 98.53% accuracy. Similarly, a drought-based prediction system combined signal-image encoding with TL to further improve results [79] as well as a water pollutants classifier [56] and an occupancy prediction system which achieved 99.42% accuracy, outperforming the comparative 1D CNN. This demonstrates the benefits of both signal-image encoding as well as TL to further improve accuracy. Overall, TL shows promise to improve classification performance when using small datasets, however there has been little consideration of the use of TL with signal-image encoding techniques to improve performance of mental wellbeing recognition.

3 DATA COLLECTION: ENVBODYSENS

Experiment setup: EnvBodySens is a dataset that has been previously collected by [53] that consists of 26 data files collected from 26 healthy female participants (average age of 28) walking around the city centre in Nottingham, UK on specific routes. The participants were asked to spend no more than 45 minutes walking in the city center. Data was collected in similar weather conditions (average 20°C), at around 11am.

Participants were asked to continuously report how they felt based on a 5-point predefined emotion scale as they walked around the city centre experiencing general daily life stressors such as loud environmental noises and crowded environments. The 5-step SAM Scale for Valence from Banzhaf et al. [9] was adopted using a smartphone app developed for the study, simplifying the continuous labelling process. This allowed participants to report their subjective state of mental wellbeing using five on-screen buttons to represent the 5-point Likert scale ranging from very positive, positive, neutral to poor and very poor. The screen auto sleep mode on the mobile devices was disabled, so the screen was kept on during the data collection process. Data from six users were excluded due to logging problems. For example, one user was unable to collect data due to a battery problem with the mobile phone and another user switched the application off accidentally.

Sensors: The dataset is composed of non-invasive physiological data (HR, EDA, body temperature, acceleration) sampled at 8Hz, environmental data (noise levels, Ultra Violet (UV) and air pressure) also sampled at 8Hz, time stamps and self reports. The data was logged by the EnvBodySens mobile application on Android phones, connected wirelessly to a Microsoft wrist Band 2 [55] that was provided to participants to collect the physiological and environmental data.

The EnvBodySens dataset resulted in 29965 samples for state 1, 35333 samples for state 2, 106210 samples for state 3, 77103 samples for state 4 and 106478 samples for state 5. Figure 2 shows the EDA (mean 1455.2kΩ, SD 2870.5kΩ) and HR (mean 74.5BPM, SD 11.8BPM) for all participants when experiencing each of the five self-reported states of valence from 1 being most positive to 5 being most negative.

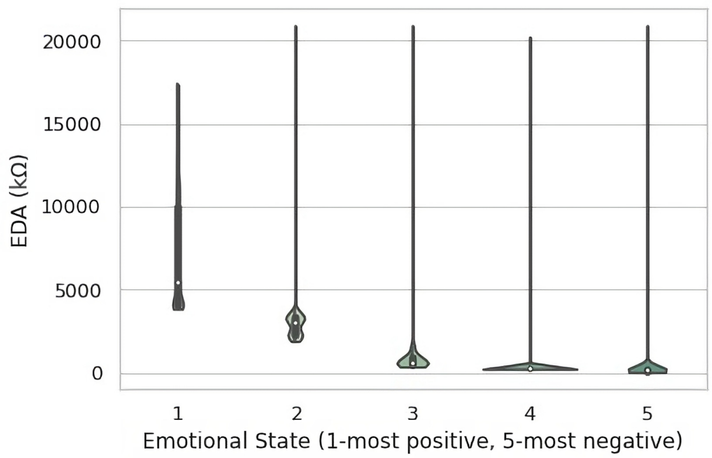

Fig. 2. EnvBodySens EDA data for reported emotional states from 1 (positive) to 5 (negative).

The distribution in Figure 2 demonstrates that as users record poorer states of wellbeing, the average EDA value decreased. The EDA data collected behaves as expected with the median EDA value gradually decreasing as users experience worsening wellbeing.

Fig. 3. EnvBodySens HR data for reported emotional states from 1 (positive) to 5 (negative).

However, Figure 3 shows wellbeing levels do not impact the distribution of HR like EDA; instead the distribution of HR remains relatively similar for all wellbeing states. Reported wellbeing state 2 has the highest distribution of HR reaching over 120 Beats Per Minute (BPM) even though this is the second most relaxed state. As users experienced worsening wellbeing the upper adjacent values are reduced, which is unexpected as when users experienced poor wellbeing they are more likely to have increased HR [77]. The outlier HR data in states 1 and 2 that go beyond 180BPM are most likely artifacts of the data due to sensor error, demonstrating that there is little change in HR over the 5 states of wellbeing. This demonstrates that HR alone as used in most commercial wearables, may not be sufficient to monitor affective state, requiring additional data modalities such as HRV and EDA. Overall, the EDA data behaves as expected while there is little to distinguish HR during the different states of wellbeing.

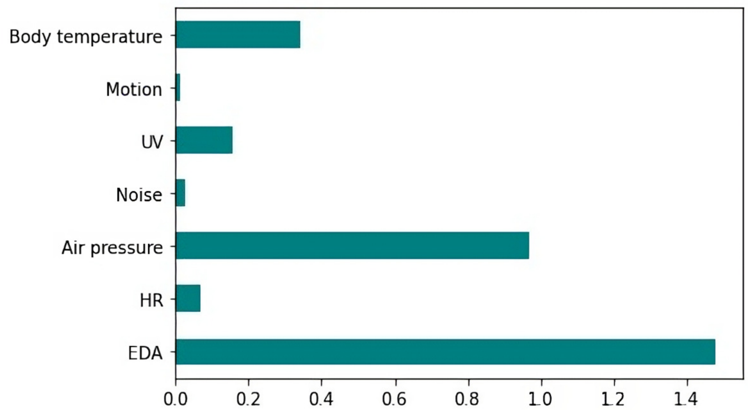

Fig. 4. Information Gain of each physiological and environmental feature.

Figure 4 shows the Information gain for each of the physiological and environmental features. Information gain is a feature selection algorithm that calculates the reduction in entropy from the transformation of the dataset. It can be used for feature selection by evaluating the Information gain of each variable in the context of the self-reported wellbeing state. EDA is shown to be a strong feature followed by air pressure and body temperature and should be used for the classification of wellbeing. UV and HR are also shown to be beneficial features with noise and motion being the least beneficial. This demonstrates in line with previous work the benefits of combining multiple data streams when classifying mental wellbeing [94], [38] and highlights environmental factors may be beneficial in affective monitoring.

Figure 5 shows a correlation matrix between the variables EDA, HR, air pressure, noise, UV, motion, body temperature and the self-reported label. The figure shows there is a statistically strong correlation (|R|>=0.3 [81]) between the label and EDA (R=-0.52) showing EDA has the highest statistical association towards mental wellbeing. Body temperature and HR follow with lower correlations of 0.2 and -0.11 respectively. This shows that physiological data has the largest correlation with wellbeing with environmental data having a lower correlation suggesting the physiological data will be most beneficial for classification. There is also a statistically strong correlation between air pressure and body temperature suggesting there is may be possible to train using only one of these features although as the correlation is R=-0.48 it may still prove beneficial to explore training using both modalities.

During the data collection process, 5345 self-report responses rated from 1 (most positive) to 5 (most negative), where sufficient samples for each rating were collected (1-8.44%, 2-9.95%, 3-29.91%, 4-21.71%, 5-29.99%). Data was successfully collected from all classes but class imbalance from an individual's dataset may impact the performance of the model. The number of samples collected by each user for each class was explored to ensure there wasn't significant bias. Each user successfully collected data for each of the five classes showing similar patterns to the complete dataset, with the majority of users collecting more data for classes 3, 4 and 5. Therefore, the percentage of the data each user collected per class was calculated with an average standard deviation was 0.15 showing that while there is a small class imbalance, no user has a significant class imbalance that would

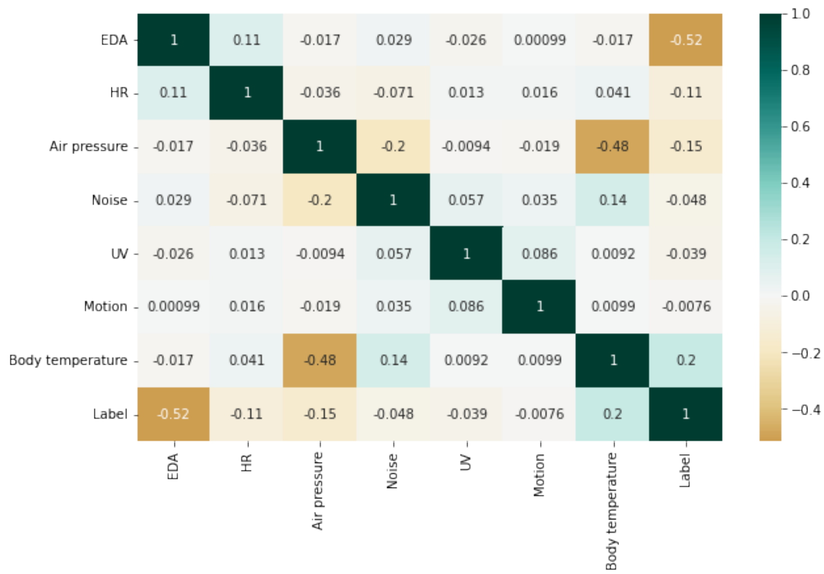

Fig. 5. Correlation matrix showing the correlation between each of the features and the self-reported wellbeing label.

impact the classification model. No single user collected significantly more or less data than the average user with the number of samples collected between users having a standard deviation of 0.01 with the size of each user's dataset ranging from 3.1% to 6.8% of the total dataset.

4 METHODS

Using the aforementioned dataset, we propose a DL approach to help improve accuracy when classifying mental wellbeing using signal-image encoding to transform accelerometer data into images and then applying a CNN-TL-based approach combined with a separate 1D CNN trained using the remaining physiological sensor data.

4.1 Modality Transformation

An image is comprised of pixels which can be conveniently represented in a matrix with a colour image containing three channels; red, green and blue for each pixel, compared with grayscale images that contain only one channel. Transforming time series data into images can help extract multi-level features [87] and improve classification accuracy [97], [40].

This study aims to explore the use of signal-image encoding with time series data for mental wellbeing classification. Therefore the continuous, fast changing datastream of accelerometer data must first be transformed into images. It is not plausible to transform the physiological data into images due to its static nature where HR and EDA can often remain constant for several seconds resulting in no data being encoded. Therefore, this physiological data will be used to train a separate 1D CNN and the two models will be concatenated. Three methods of modality transformation using accelerometer data are utilised: GADF, GASF and MTF.

Wang and Oates transformed time series data into images using Gramian Angular Field (GAF) [87]. First, the data was normalised between -1 and 1 by applying 1. The normalised data is then encoded using the value as the angular cosine and the time stamp as the radius r with 2, where ϕ is the angle polar coordinates, ti is the time stamp, N is a constant factor to regularize the span of the polar coordinate system and $\tilde{X}$ represents the re-scaled time series data [97].

$$\tilde{x}^i_{-1} = \frac{(x_i - \max(x)) + (x_i - \min(x))}{\max(x) - \min(x)} \tag{1}$$

$$\begin{cases} \phi = \arccos\left(\tilde{x}_i\right), -1 \leq \tilde{x}_i \leq 1, \tilde{x}_i \in \tilde{X} \\ r = \frac{t_i}{N}, t_i \in \mathbb{N} \end{cases} \tag{2}$$

$$GASF = \left[\cos\left(\emptyset_i + \emptyset_j\right)\right] \tag{3}$$

$$= \tilde{X}' \cdot \tilde{X} - \sqrt{I - \tilde{X}^2}' \cdot \sqrt{I - \tilde{X}^2} \tag{4}$$

$$GADF = \left[\sin\left(\emptyset_i - \emptyset_j\right)\right] \tag{5}$$

$$= \sqrt{I - \tilde{X}^2}' \cdot \tilde{X} - \tilde{X}' \cdot \sqrt{I - \tilde{X}^2} \tag{6}$$

The normalized data is then transformed into polar coordinates instead of the typical Cartesian coordinates. After transformation, the vectors are transformed into a symmetric matrix called the Gramian Matrix. There are two ways to transform the vectors into a symmetric matrix: GASF and GADF as shown from 3 to 6 where $\emptyset$ is the angle polar coordinates. These methods preserve the temporal dependency, with the position moving from top-left to bottom-right with time.

$$M = \begin{bmatrix} w_{ij|x_1 \in q_i, x_1 \in q_j} & \cdots & w_{ij|x_1 \in q_i, x_n \in q_j} \\ w_{ij|x_2 \in q_i, x_1 \in q_j} & \cdots & w_{ij|x_2 \in q_i, x_n \in q_j} \\ \vdots & \ddots & \vdots \\ w_{ij|x_n \in q_i, x_1 \in q_j} & \cdots & w_{ij|x_n \in q_i, x_n \in q_j} \end{bmatrix} \tag{7}$$

Alternatively, images can be generated using MTF where the Markov matrix is built and the dynamic transition probability is encoded in a quasi-Gramian matrix as defined in 7. Given a time series x and its q quantile bins each x_i is assigned to the corresponding bins, q_j $(j \in [1, q])$. A q X q Markov transition matrix (w) is created by dividing the data into q quantile bins. The quantile bins that contain the data at time stamp i and j (temporal axis) are q_i and q_j. The information of the inter-relationship is preserved by extracting the Markov transition probabilities to encode dynamic transitional fields in a sequence of actions [87]. A comparison of identical X, Y, Z and total acceleration data transformed as GASF, GADF and MTF can be seen in figure 6.

4.2 Transfer Learning

This work proposes the novel combination of a 2D CNN utilising TL with signal-encoded images and a 1D CNN model to improve the accuracy of mental wellbeing classification from the EnvBodySens dataset. TL first requires a pre-trained model; for this work multiple pre-trained object recognition networks have been explored. Given that the majority of pre-trained models for TL have been trained on images, it is beneficial to train these networks using signal encoded images from the continually changing motion data and not the physiological data which can often remain static resulting in little data being encoded.

The general process of transfer learning is to pre-train the model in the source domain with sufficient data, and then fine-tune the parameters in the target domain with sparse data to achieve the purpose of reducing training data, enhancing the generalization ability of the model, and improving accuracy. This process works when the features are generic, meaning suitable to both base and target tasks, instead of specific to the base task [84].

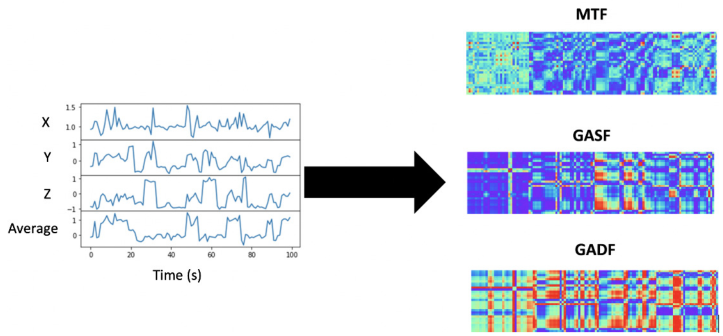

Fig. 6. An example of raw accelerometer data (X, Y, Z and average motion) transformed using MTF, GASF and GADF.

Given the relatively limited sample size for this experiment, it is easy to cause over-fitting problems resulting in poor model accuracy. Transfer learning is ideally suitable to utilise state of the art pre-trained networks that have been optimised on large amounts of data for long periods of time. The lower layers which are trained on the large ImageNet are retained, and the top layers of the network are retrained for the smaller target dataset, in this case to classify mental wellbeing. When training our own data set, we only need to add our own classification layer to the last layer of the extracted model, and only need to adjust the parameters of the last layer to realize migration learning.

We leverage pre-trained object recognition models to the images encoding the sensory data streams. Even though the pre-trained models were designed for object recognition rather than time-series classification, the vast datasets the models were trained with makes them ideal for detecting patterns within images. Previous research has explored time-series image-encoding and transfer learning has similarly used pre-trained models designed for object recognition [14], [86], [12]. However, previous work has not explored image encoding for mental well-being recognition where increases in accuracy can have a significant real-world impact.

4.3 Image Encoding Model

To generate the images for the signa;-image encoding approach, the accelerometer data from the EnvBodySens dataset was transformed into images using GADF, GASF and MTF, resulting in a total of 17,750 images for each encoding technique. When training using TL, the source pre-trained model is imported without its last layer, with dense layers then added to enable the new model to learn more complex functions from the new data. Therefore, the generated images are used as input to train the 2D CNN consisting of 2 convolutional layers, polling layer, dropout layer and fully connected layer over 10 epochs to classify 5 states of wellbeing by exploring 7 pre-trained models (Xception, VGG19, ResNet, NasNet, DenseNet, DenseNet V2 & MobileNet) to apply the TL approach.

While much work on affective computing uses recurrent neural networks such as LSTMs, the CNN used in this approach is indicated to enable the TL approach as CNNs are particularly appealing towards learning spatial features. CNNs have traditionally been used to classify images and speech, however their application has been expanded to classify raw sensor data [53], [41]. Furthermore LSTMs are often not successful for short-time, frequently changing, and non-periodical data [43], where CNNs outperform recurrent networks while demonstrating longer effective memory [6] and run faster than LSTMs [90]. Previous work shows one-dimensional CNN outperforming LSTM networks for mental wellbeing recognition from physiological sensor data [94] and when detecting emotions from EEG signals with the LSTM models being less stable, less accurate and taking longer to train [102]. For those reasons, we decided not to explore further the use of LSTMs in this study

4.4 Physiological Data Model

An additional 1D CNN model was trained using the remaining physiological sensor data from the EnvBodySens dataset (HR, EDA, body temperature, acceleration, noise, UV and air pressure). The proposed 1D CNN has three layers; an input layer, an output layer and a hidden layer. An overlapping sliding window strategy has been adopted to segment the time series data with a window size of 100 and a step of 20 chosen experimentally, by trying different window sized from 10 to 400. The training input data is represented as $x = [x_1, x_2, \ldots\ldots x_j]$, where the number of training samples is j and y is the output vector [28]. When σ is the sigmoid activation function, w1 and w2 are weight matrices between the input and hidden layer and the hidden and output layer respectively. Finally, b_1 and b_2 represent the bias vectors of the hidden and output layer respectively [101]:

$$h = \sigma(w_1 x + b_1] \tag{8}$$

$$y = \sigma(w_2 h + b_2) \tag{9}$$

Batch normalisation has been used within the network to normalise the inputs of each layer so they have a mean of 0 and standard deviation of 1 this enables the models to train quicker, allows for higher learning rates and makes the weights easier to initialise [35]. A dropout layer with a rate of 0.5 was added before the maxpooling layer to prevent overfitting by randomly ignoring selected neurons during training [75]. The pooling layers then subsample the data, reducing the number of weights within that activation. Finally, the fully-connected layers where each neuron is connected to all the neurons in the previous layer are used to calculate class predictions from the activation.

4.5 Concatenated Model

The two models (2D CNN trained using accelerometer signal encoded images & 1D CNN trained using physiological sensor data) are frozen and then the concatenated feature vector is fed into two fully-connected layers to classify the 5 states of wellbeing as shown in figure 7.

Hold-out validation using a 20% test split has been used to test the model using around 284,000 sensor data samples for training and 71,000 for testing. Additionally, Leave-One-participant-Out Cross-Validation (LOOCV) has also been utilised to test the signal-image encoding approach on a subject-independent basis. This is where the model is trained with 19 users' data then tested on the remaining user's data (19926 average data samples) to better simulate how the model would be used in the real-world to infer an individual's wellbeing.

5 RESULTS

The EnvBodySens dataset has been used to explore the multi-class problem of classifying five emotional states using the signal-image encoding model. Seven pre-trained models (Xception, VGG19, ResNet, NasNet, DenseNet, DenseNet V2 & MobileNet) were used to explore the TL approach for the three methods of signal-image transformation (GADF, GASF & MTF). This approach transformed the motion data from the EnvBodySens dataset to images to train a 2D CNN which was then paired with a 1D CNN trained using the remaining time series data from the EnvBodySens dataset. The final testing accuracy using the 20% test data split are reported for each model in Table 1.

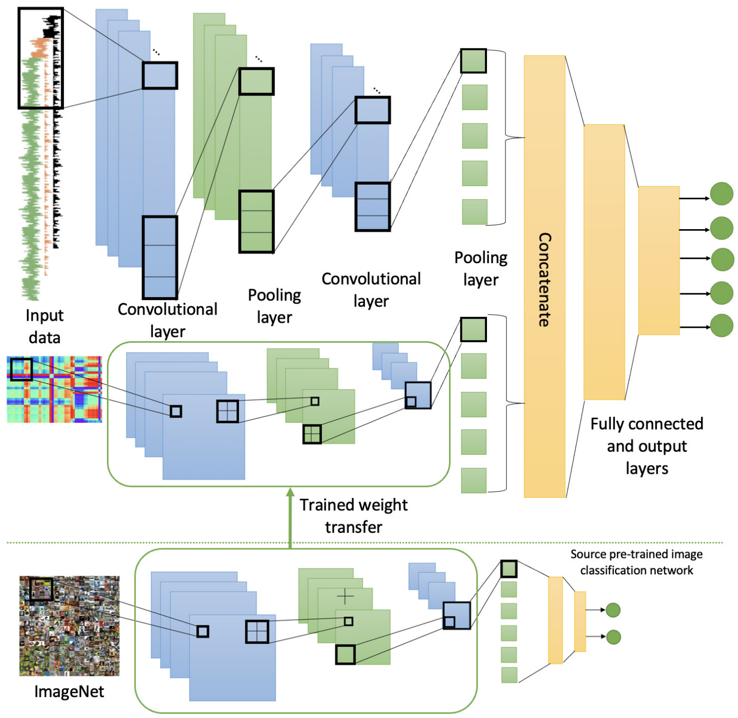

Fig. 7. Combinatory model consisting of 1D CNN trained using raw physiological sensor data (top) and a 2D CNN using a transfer learning approach trained using accelerometer encoded images (bottom).

Table 1. Comparison of accuracy for different pre-trained deep learning models adapted for mental wellbeing classification through TL.

	GASF		GADF		MTF	
	Physiological	All	Physiological	All	Physiological	All
Xception	0.975	0.96	0.977	0.971	0.972	0.956
VGG19	**0.984**	0.952	0.98	0.94	0.964	0.95
ResNet	0.963	0.955	**0.978**	0.937	0.964	0.972
NasNet	0.977	0.965	**0.983**	0.963	0.967	0.964
DensetNet	0.975	0.977	**0.985**	0.971	0.97	0.977
MobileNetV2	**0.981**	0.97	**0.981**	0.954	0.979	0.97
MobileNet	**0.98**	0.967	0.968	0.955	0.974	0.959
No TL	**0.98**	0.974	0.974	0.963	0.975	0.968

5.1 Comparison of Data Modalities

The data modalities were investigated to explore which modalities most contributed towards the classification of mental wellbeing. When all sensor data (HR, EDA, UV, body temperature, air pressure and noise) was used to train the 1D CNN combined with the signal-image transformed motion data, the model achieved accuracies between 93.7% and 97.7% as shown in Table 8. The 1D CNN was also trained using only physiological data (HR & EDA) to examine the impact not including environmental data has on model performance. When using the signal-image encoding approach for motion data and a 1D CNN trained using only physiological data, the model accuracy increased to the highest achieved accuracy of 98.5% when using GADF to transform the motion data and DenseNet to perform TL, as shown in figure 8. Furthermore, when comparing the highest accuracy for each pre-trained CNN the physiological model consistently outperformed the model trained using all modalities. This demonstrates the importance of physiological data when determining wellbeing state, unlike environmental data which resulted in more misclassification errors, in particular class 5 the poorest mental wellbeing state.

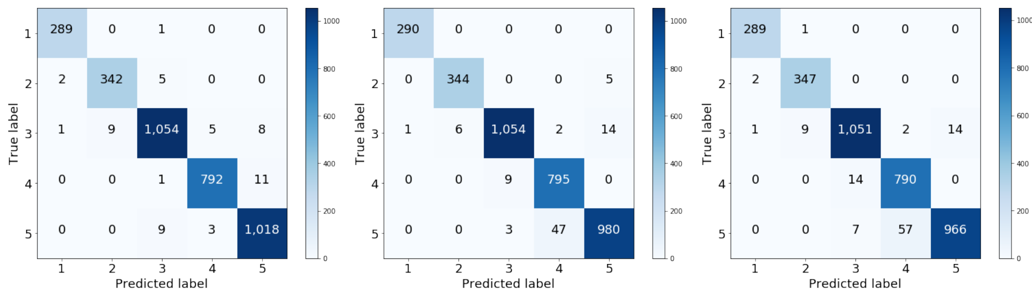

Fig. 8. Confusion matrix for DenseNet model trained using HR, EDA and GADF (left), GASF (middle) and MTF (right) encoded motion data.

To evaluate whether the signal-image encoding approach improves model performance all sensor data (HR, EDA, noise, UV, body temperature, air pressure and accelerometer data) was used to train the 1D CNN model without performing TL or signal-image encoding. The 1D CNN achieved 93% accuracy, an overall reduction in accuracy compared with the signal-image encoding model. Furthermore, when the same 1D CNN was trained again using only physiological data, the model achieved 94% accuracy, a 4.5% reduction in accuracy. This demonstrates that image encoding can increase overall model accuracy but performance is highly dependent on the additional sensor modalities used to train the network. Additionally, to explore whether simpler models can classify wellbeing a generalised linear model, naive bayes and logistic regression model were all trained with the sensor data. The models were trained using automatic feature engineering and tested using a 20% test split. The generalised linear model achieved the worst accuracy of 55% compared with Naive bayes achieving 83% accuracy and Logistic regression achieving 84%. This highlights the benefits of deep learning to achieve the highest accuracy compared with simple classifiers likely due to the complex nature of wellbeing states.

5.2 Comparison of Pre-trained Models

The results show TL has little impact on performance but to explore whether the high accuracy achieved was influenced by the pre-trained model used in the TL approach, other pre-trained CNNs were tested using the same GASF, GADF and MTF transformed images. As shown in table 1 DensNet achieved the highest accuracy for the GADF transformed data although VGG19 achieved the highest accuracy for GASF data and MobileNetV2 for MTF data. This demonstrates that the pre-trained model selected has little impact on performance with the average variance between the best and worst performing model for all 3 image encoding techniques being only 1.77% for the physiological models and 2.87% for the models trained using all sensor data. Figure 9 shows the F1-scores and error bars for each of the TL approaches using GASF encoded images for models trained using only physiological data and all data modalities. While the signal-image encoding had much greater performance gains than TL, the use of pre-trained models has not sufficiently demonstrated the potential to further increase accuracy.

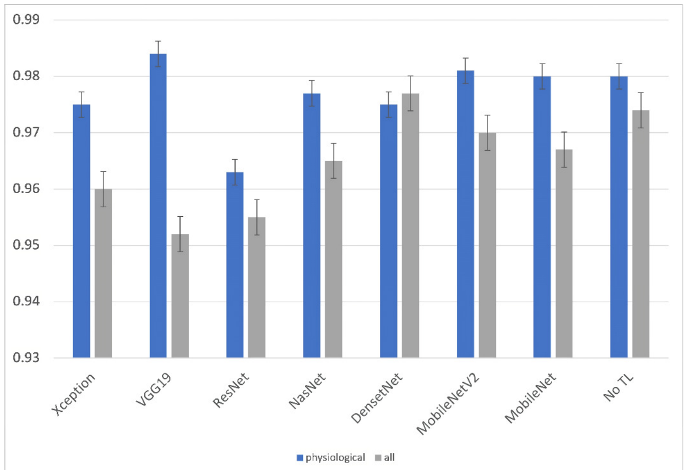

Fig. 9. Chart showing the GASF physiological and all data models F1-Score with error bars.

5.3 Comparison of Signal-Image Encoding Techniques

The use of signal-image encoding demonstrated the capability to increase affective modelling beyond traditional deep learning classifiers where the technique used also impacted model performance. GASF and GADF outperformed MTF for each pre-trained model, where GADF achieved the highest performance for four of the pre-trained models and GASF for the remaining three. The average accuracy for the GADF physiological model was 97.9% compared with 97.6% for GASF and 97% for MTF showing negligible variations in performance between the different techniques.

5.4 Subject-Independent Models

As the GADF signal-encoding technique slightly outperformed the other encoders, it was used to explore subject-independent physiological models. Table 2 shows the accuracy achieved for each of the 20 users when the model was tested using LOOCV with each individual's physiological data. The accuracies range between 36.4% for user 1 and 77.7% for users 16 and 17. The outlier low accuracy for user 1 is due to corrupt EDA data which continually recorded null readings. The remaining users demonstrate more consistent accuracies and while lower than when tested using hold-out validation, they demonstrate the possibility of inferring wellbeing on an individual basis.

The subject-independent models were also trained without the TL approach while still transforming signals into images to explore whether performance improvements were due to the TL approach. A 2D CNN was implemented to train the signal encoded images which was concatenated with the 1D CNN trained using the physiological data. The results show TL increased average accuracy by 0.55% for all users which falls within the margin of error, demonstrating no overall performance improvement. However, the TL approach never degraded the performance of individuals' models and achieved up to a 4% increase in accuracy.

6 DISCUSSION

A new CNN-TL-based approach towards affective state classification has been introduced that goes beyond previous signal-image encoding frameworks by incorporating TL in addition to a separate 1D CNN. This research demonstrates that a signal-image encoding approach can improve the performance in which five affective states can be classified, achieving up to 98.5% accuracy using hold-out validation and an average of 72.3% using LOOCV.

Table 2. Comparison of subject-independent classification accuracy.

User	Accuracy	User	Accuracy
1	0.364	11	0.709
2	0.698	12	0.734
3	0.702	13	0.723
4	0.683	14	0.738
5	0.666	15	0.749
6	0.752	16	0.777
7	0.736	17	0.777
8	0.706	18	0.753
9	0.737	19	0.690
10	0.636	20	0.763

This outperforms many previous real-world affect recognition systems [17], [58], [32], [99] including a previous stacked machine learning approach using the same EnvBodySens dataset which achieved an accuracy of 86% [38] and a combined CNN and RNN using the same dataset that achieved 94.9% accuracy [39].

The results have demonstrated that the integration of signal-image encoding as part of the newly proposed methodology, extending standard deep learning algorithms, can improve the classification of affective state. In particular, the combinatory approach of encoding accelerometer data as images using GADF, GASF and MTF and subsequently combining this model with a 1D CNN trained using physiological data, has improved the overall model accuracy. The proposed framework increased model accuracy by 4.5%, which is similar to related research work that has used signal-image encoding. A related study exploring the classification of human activity recognition increased by 4.5% using the signal-imaging TL approach [14],similarly coal-rock interface recognition increased by 7.1% [78] and eucalyptus region classification accuracy increased by up to 4.2% when compared with state-of-the-art models in the research literature [23]. Collectively, these empirical findings demonstrate that signal-image encoding and TL can be used to further improve deep learning classification models suggesting its use is most beneficial to increase the accuracy of well-performing models as used in this work.

While transfer learning marginally impacted overall performance by 0.04% -0.11% compared to the non-TL signal image encoding models, the signal-image encoding demonstrated the ability to further reliably increase affective modelling accuracy compared to the standard 1D CNN without signal-image encoding. As the signal-image encoding increased accuracy to 98% there is little room for improvement due to it being difficult to about bringing further improvement above 98%. Furthermore, the use of TL frequently outperformed the non-TL models showing its benefits even if marginal. The signal-image encoding was shown to provide greater gains in accuracy with the encoding technique used having a minor impact in model accuracy demonstrating GADF was most effective for the majority of the models. Similarly, the pre-trained model used to perform TL had a limited impact on model performance with an average difference of 2.32% between the different models. In comparison, TL slightly improved performance by an average of 0.55% when testing using subject-independent models, demonstrating the transformed images had a greater impact on model performance than TL.

Furthermore, solely using physiological and motion data resulted in the highest accuracy (98.5%), outperforming models additionally trained using environmental data. This suggests that environmental factors such as noise and UV are more challenging to use for affect recognition even when paired with physiological data. The reduced performance may be due to the intricate information in the environmental data already being captured inherently in the physiological and motion data for example poor weather having a negative impact on mood.

When testing using LOOCV the subject-independent accuracies are lower than subject-dependent accuracies. The average accuracy of the subject-independent physiological models excluding user 1 was 72.3% (SD 0.038), compared with 98.5% for the subject-dependent model both using GADF to transform the signals and a DenseNet pre-trained model. This likely reflects that different individuals have different patterns of physiology when experiencing the same state of wellbeing and that similar levels of activity are perceived differently in terms of valence [82] demonstrating similar results as other studies [36], [50], [5].

Classifying mental wellbeing is a challenging proposition that usually requires large real-world datasets that can be challenging to collect. Signal-image encoding and transfer learning resulted in the best prediction performance of mental wellbeing. This developed approach helps further increase classification accuracy beyond traditional deep learning methods resulting in more reliable real-world inference. This approach demonstrates the ability to accurately classify 5 emotional states outperforming traditional deep learning networks such as LSTM [96], [51], [83] and CNN [65]. This approach outperforms other approaches to classify mental wellbeing using ECG signals which achieved 79.26% [26]. Similarly, this framework outperforms previous work using EDA and HRV sensors that achieved accuracies between 70-75% when classifying 4 emotions [53], 74% when classifying 5 emotions from EDA[89] and 95% when measuring stress [20]. This demonstrates the benefits of using the signal-image encoding approach to classify 5 emotional states with high accuracy. This increase in performance is highly beneficial for affective computing as the outputs from the model can be used to initiate interventions to help improve mental wellbeing.

Overall, this work demonstrates that by using the proposed approach it is possible to capitalise on two modalities to accurately classify wellbeing on a 5-point Likert scale. The results have demonstrated that signal-image encoding approaches are appropriate for modeling affective states especially when training data is scarce. This approach has outperformed previous classification models using the same EnvBodySens dataset built on ad-hoc extracted features [38] and 2 dimensional CNNs [39]. These findings showcase the potential for signal-encoded images to improve affective multimodal modeling. In this work we have addressed research challenges 1 (User availability, incentivisation and willingness to participate in longitudinal studies) by helping to demonstrate the potential to develop highly accurate classification model with a limited number of participants and 6 (Time consuming nature of real-time self labelling) by leveraging pre-trained models and data reducing the amount of labelled data required to train a model. We believe that this approach has the potential to make longitudinal studies more accessible and appealing to participants, as it reduces the burden of participation. The remaining challenges will need to be considered and investigated in further research work that is beyond the scope of this study.

7 CONCLUSION AND FUTURE WORK

Recent developments are producing sensory datasets as people are going about their daily activities. However, accurately classifying these limited datasets can be a challenging proposition. A scenario of wellbeing classification using small multimodal datasets has been presented. Although these types of time series datasets can help us understand people's wellbeing, current recognition techniques are not efficient enough to tackle data scantiness. This research has demonstrated the advantages of employing a combinatory TL, signal-image encoding approach for raw multimodal sensor data modelling. We demonstrate how this approach can be practically implemented on a small affective dataset.

We have performed a comparative study of methods based on the combination of image-based time series representations and deep transfer learning models. In particular, we assessed, for the first time in the context of the affective recognition, the effectiveness of GASF, GADF and MTF representations combined with a TL approach. The proposed framework using signal encoded images with a 1D CNN. Accelerometer data was transformed into RGB images using GADF, MTF and GASF and was subsequently used to train a 2D CNN using pre-trained models to apply a TL approach. This model was concatenated with a 1D CNN architecture trained using raw physiological sensor data. This novel framework expands upon previous work by applying signal-image encoding to increase model performance, in particular for affective computing where lies great potential as signal-image encoding has had little exploration and collecting large labelled datasets can be extremely challenging.

There are several future directions to further study the signal-image encoding approach used in this research. First, this work has only explored 1D CNNs, in the future it would be worth evaluating whether other architectures could further improve classification accuracy. Additionally, this framework could be evaluated beyond the classification of mental wellbeing using alternative datasets to evaluate performance in other time-series domains such as human activity recognition.

REFERENCES

[1] Abdullah Ahmed, Jayroop Ramesh, Sandipan Ganguly, Raafat Aburukba, Assim Sagahyroon, and Fadi Aloul. 2023. Evaluating Multimodal Wearable Sensors for Quantifying Affective States and Depression with Neural Networks. *IEEE Sensors Journal* (2023). https://doi.org/10.1109/JSEN.2023.3303436

[2] Mariano Albaladejo-González, José A. Ruipérez-Valiente, and Félix Gómez Mármol. 2023. Evaluating different configurations of machine learning models and their transfer learning capabilities for stress detection using heart rate. *Journal of Ambient Intelligence and Humanized Computing* 14, 8 (aug 2023), 11011–11021. https://doi.org/10.1007/S12652-022-04365-Z/TABLES/6

[3] Amani Albraikan, Diana P. Tobon, and Abdulmotaleb El Saddik. 2019. Toward User-Independent Emotion Recognition Using Physiological Signals. *IEEE Sensors Journal* 19, 19 (oct 2019), 8402–8412. https://doi.org/10.1109/JSEN.2018.2867221

[4] Salma Alhagry, Aly Aly, and Reda A. 2017. Emotion Recognition based on EEG using LSTM Recurrent Neural Network. *International Journal of Advanced Computer Science and Applications* 8, 10 (2017). https://doi.org/10.14569/ijacsa.2017.081046

[5] Viet Hoang Anh, Manh Ngo Van, Bang Ban Ha, and Thang Huynh Quyet. 2012. A real-time model based Support Vector Machine for emotion recognition through EEG. In *2012 International Conference on Control, Automation and Information Sciences, ICCAIS 2012*. https://doi.org/10.1109/ICCAIS.2012.6466585

[6] Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. 2018. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. (mar 2018). https://doi.org/10.48550/arxiv.1803.01271 arXiv:1803.01271

[7] Debrup Banerjee, Kazi Islam, Gang Mei, Lemin Xiao, Guangfan Zhang, Roger Xu, Shuiwang Ji, and Jiang Li. 2017. A deep transfer learning approach for improved post-traumatic stress disorder diagnosis. In *Proceedings - IEEE International Conference on Data Mining, ICDM*.

[8] Debrup Banerjee, Kazi Islam, Keyi Xue, Gang Mei, Lemin Xiao, Guangfan Zhang, Roger Xu, Cai Lei, Shuiwang Ji, and Jiang Li. 2019. A deep transfer learning approach for improved post-traumatic stress disorder diagnosis. *Knowledge and Information Systems* 60, 3 (sep 2019), 1693–1724.

[9] Ellen Banzhaf, Francisco De La Barrera, Annegret Kindler, Sonia Reyes-Paecke, Uwe Schlink, Juliane Welz, and Sigrun Kabisch. 2014. A conceptual framework for integrated analysis of environmental quality and quality of life. *Ecological Indicators* (2014). https://doi.org/10.1016/j.ecolind.2014.06.002

[10] A Batliner, S Steidl, and E Nöth. 2008. Releasing a thoroughly annotated and processed spontaneous emotional database: the FAU Aibo Emotion Corpus. *Proc. of Workshop on Corpora for Research on Emotion and Affect LREC* (2008).

[11] Thomas Boraud. 2020. *How the Brain Makes Decisions*. https://doi.org/10.1093/oso/9780198824367.001.0001

[12] Issam Boukhennoufa, Zainab Altai, Xiaojun Zhai, Victor Utti, Klaus D. McDonald-Maier, and Bernard X.W. Liew. 2022. Predicting the Internal Knee Abduction Impulse During Walking Using Deep Learning. *Frontiers in Bioengineering and Biotechnology* 10 (may 2022). https://doi.org/10.3389/FBIOE.2022.877347/FULL

[13] Issam Boukhennoufa, Xiaojun Zhai, Klaus D. McDonald-Maier, Victor Utti, and Jo Jackson. 2021. Improving the activity recognition using GMAF and transfer learning in post-stroke rehabilitation assessment. *SAMI 2021 - IEEE 19th World Symposium on Applied Machine Intelligence and Informatics, Proceedings* (jan 2021), 391–397. https://doi.org/10.1109/SAMI50585.2021.9378670

[14] Issam Boukhennoufa, Xiaojun Zhai, Klaus D. McDonald-Maier, Victor Utti, and Jo Jackson. 2021. Improving the activity recognition using GMAF and transfer learning in post-stroke rehabilitation assessment. *SAMI 2021 - IEEE 19th World Symposium on Applied Machine Intelligence and Informatics, Proceedings* (jan 2021), 391–397. https://doi.org/10.1109/SAMI50585.2021.9378670

[15] Margaret M. Bradley and Peter J. Lang. 1994. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry* (1994). https://doi.org/10.1016/0005-7916(94)90063-9

[16] Rich Caruana. 1997. Multitask Learning. *Machine Learning* 28, 1 (1997), 41–75. https://doi.org/10.1023/A:1007379606734

[17] R. Castaldo, W. Xu, P. Melillo, L. Pecchia, L. Santamaria, and C. James. 2016. Detection of mental stress due to oral academic examination via ultra-short-term HRV analysis. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology*

Society, EMBS. https://doi.org/10.1109/EMBC.2016.7591557
[18] Antonio R. Damasio. 1998. Emotion in the perspective of an integrated nervous system. In *Brain Research Reviews*, Vol. 26. https://doi.org/10.1016/S0165-0173(97)00064-7
[19] Thomas De Cooman, Kaat Vandecasteele, Carolina Varon, Borbála Hunyadi, Evy Cleeren, Wim Van Paesschen, and Sabine Van Huffel. 2020. Personalizing Heart Rate-Based Seizure Detection Using Supervised SVM Transfer Learning. *Frontiers in Neurology* (2020). https://doi.org/10.3389/fneur.2020.00145
[20] Alberto De Santos Sierra, Carmen Sanchez Avila, Javier Guerra Casanova, Gonzalo Bailador Del Pozo, and Vicente Jara Vera. 2010. Two stress detection schemes based on physiological signals for real-time applications. In *Proceedings - 2010 6th International Conference on Intelligent Information Hiding and Multimedia Signal Processing, IIHMSP 2010*. https://doi.org/10.1109/IIHMSP.2010.95
[21] Jun Deng, Sascha Fruhholz, Zixing Zhang, and Bjorn Schuller. 2017. Recognizing Emotions from Whispered Speech Based on Acoustic Feature Transfer Learning. *IEEE Access* (2017), 1–1. https://doi.org/10.1109/ACCESS.2017.2672722
[22] Jun Deng, Zixing Zhang, Erik Marchi, and Björn Schuller. 2013. Sparse autoencoder-based feature transfer learning for speech emotion recognition. In *Proceedings - 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, ACII 2013*. 511–516. https://doi.org/10.1109/ACII.2013.90
[23] Danielle Dias, Ulisses Dias, Nathalia Menini, Rubens Lamparelli, Guerric Le Maire, and Ricardo Da S. Torres. 2020. Image-Based Time Series Representations for Pixelwise Eucalyptus Region Classification: A Comparative Study. *IEEE Geoscience and Remote Sensing Letters* 17, 8 (aug 2020), 1450–1454. https://doi.org/10.1109/LGRS.2019.2946951
[24] Paul Ekman and Wallace V. Friesen. 1971. Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology* (1971). https://doi.org/10.1037/h0030377
[25] M Everingham, L Van Gool, C K I Williams, J Winn, and A Zisserman. 2012. DLPASCAL The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Development Kit. *Pattern Analysis, Statistical Modelling and Computational Learning, Tech. Rep* (2012).
[26] Fatemeh Fahimi, Zhuo Zhang, Wooi Boon Goh, Tih Shi Lee, Kai Keng Ang, and Cuntai Guan. 2019. Inter-subject transfer learning with an end-to-end deep convolutional neural network for EEG-based BCI. *Journal of Neural Engineering* (2019). https://doi.org/10.1088/1741-2552/aaf3f6
[27] Ian Goodfellow, Yoshua Benigo, and Courville Aaron. 2016. Deep Learning (Adaptive Computation and Machine Learning).
[28] Md Junayed Hasan, Muhammad Sohaib, and Jong Myon Kim. 2019. 1D CNN-based transfer learning model for bearing fault diagnosis under variable working conditions. In *Advances in Intelligent Systems and Computing*, Vol. 888. Springer Verlag, 13–23. https://doi.org/10.1007/978-3-030-03302-6_2
[29] J.A. Healey and R.W. Picard. 2005. Detecting Stress During Real-World Driving Tasks Using Physiological Sensors. *IEEE Transactions on Intelligent Transportation Systems* 6, 2 (jun 2005), 156–166.
[30] Katherine A. Herborn, James L. Graves, Paul Jerem, Neil P. Evans, Ruedi Nager, Dominic J. McCafferty, and Dorothy E.F. F McKeegan. 2015. Skin temperature reveals the intensity of acute stress. *Physiology & Behavior* 152, Pt A (dec 2015), 225–230. https://doi.org/10.1016/J.PHYSBEH.2015.09.032
[31] M A Hogg, Dominic Abrams, and G N Martin. 2010. Social cognition and attitudes. *Psychology* (2010), 646–677.
[32] Karen Hovsepian, Mustafa Al'absi, Emre Ertin, Thomas Kamarck, Motohiro Nakajima, and Santosh Kumar. 2015. CStress: Towards a gold standard for continuous stress assessment in the mobile environment. In *UbiComp 2015 - Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. https://doi.org/10.1145/2750858.2807526
[33] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, Quoc V Le, Hartwig Adam, Google Ai, and Google Brain. [n. d.]. *Searching for MobileNetV3*. Technical Report. arXiv:1905.02244v5
[34] Guang Bin Huang, Hongming Zhou, Xiaojian Ding, and Rui Zhang. 2012. Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* (2012). https://doi.org/10.1109/TSMCB.2011.2168604
[35] Sergey Ioffe and Christian Szegedy. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *32nd International Conference on Machine Learning, ICML 2015*, Vol. 1. International Machine Learning Society (IMLS), 448–456.
[36] Noppadon Jatupaiboon, Setha Pan-Ngum, and Pasin Israsena. 2013. Real-time EEG-based happiness detection system. *The Scientific World Journal* (2013). https://doi.org/10.1155/2013/618649
[37] Houtan Jebelli, Sungjoo Hwang, and Sang Hyun Lee. 2018. EEG-based workers' stress recognition at construction sites. *Automation in Construction* (2018). https://doi.org/10.1016/j.autcon.2018.05.027
[38] Eiman Kanjo, Eman M.G. Younis, and Nasser Sherkat. 2018. Towards unravelling the relationship between on-body, environmental and emotion data using sensor information fusion approach. *Information Fusion* (2018).
[39] Eiman Kanjo, Eman M G Younis, and Chee Siang Ang. 2019. Deep learning analysis of mobile physiological, environmental and location sensor data for emotion detection. *Information Fusion* 49 (2019), 46–56.
[40] Saeed Karimi-Bidhendi, Faramarz Munshi, and Ashfaq Munshi. 2019. Scalable Classification of Univariate and Multivariate Time Series. In *Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018*. https://doi.org/10.1109/BigData.2018.8621889

[41] Reza Khosrowabadi, Chai Quek, Kai Keng Ang, Sau Wai Tung, and Michel Heijnen. 2011. A Brain-Computer Interface for classifying EEG correlates of chronic mental stress. In *The 2011 International Joint Conference on Neural Networks*. IEEE, 757–762. https://doi.org/10.1109/IJCNN.2011.6033297

[42] Basel Kikhia, Thanos G Stavropoulos, Stelios Andreadis, Niklas Karvonen, Ioannis Kompatsiaris, Stefan Sävenstedt, Marten Pijl, and Catharina Melander. 2016. Utilizing a Wristband Sensor to Measure the Stress Level for People with Dementia. *Sensors (Basel, Switzerland)* 16, 12 (nov 2016). https://doi.org/10.3390/s16121989

[43] Nobuaki Kimura, Ikuo Yoshinaga, Kenji Sekijima, Issaku Azechi, and Daichi Baba. 2019. Convolutional Neural Network Coupled with a Transfer-Learning Approach for Time-Series Flood Predictions. *Water* 12, 1 (dec 2019), 96.

[44] Dustin C. Krutsinger, Kuldeep N. Yadav, Elizabeth Cooney, Steven Brooks, Scott D. Halpern, and Katherine R. Courtright. 2019. A pilot randomized trial of five financial incentive strategies to increase study enrollment and retention rates. *Contemporary Clinical Trials Communications* (2019). https://doi.org/10.1016/j.conctc.2019.100390

[45] Jennifer R. Kwapisz, Gary M. Weiss, and Samuel A. Moore. 2011. Activity recognition using cell phone accelerometers. *ACM SIGKDD Explorations Newsletter* (2011). https://doi.org/10.1145/1964897.1964918

[46] Lan lan Chen, Ao Zhang, and Xiao guang Lou. 2019. Cross-subject driver status detection from physiological signals based on hybrid feature selection and transfer learning. *Expert Systems with Applications* (2019). https://doi.org/10.1016/j.eswa.2019.02.005

[47] Siddique Latif, Rajib Rana, Shahzad Younis, Junaid Qadir, and Julien Epps. 2018. Transfer learning for improving speech emotion classification accuracy. In *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*. https://doi.org/10.21437/Interspeech.2018-1625 arXiv:1801.06353

[48] Jinpeng Li, Shuang Qiu, Yuan-Yuan Shen, Cheng-Lin Liu, and Huiguang He. 2019. Multisource Transfer Learning for Cross-Subject EEG Emotion Recognition. *IEEE Transactions on Cybernetics* 50, 7 (2019), 3281 – 3293.

[49] Yuan Pin Lin and Tzyy Ping Jung. 2017. Improving EEG-based emotion classification using conditional transfer learning. *Frontiers in Human Neuroscience* (2017).

[50] Yuan Pin Lin, Chi Hong Wang, Tzyy Ping Jung, Tien Lin Wu, Shyh Kang Jeng, Jeng Ren Duann, and Jyh Horng Chen. 2010. EEG-based emotion recognition in music listening. *IEEE Transactions on Biomedical Engineering* (2010). https://doi.org/10.1109/TBME.2010.2048568

[51] Yun Liu and Siqing Du. 2018. Psychological stress level detection based on electrodermal activity. *Behavioural brain research* 341 (2018), 50–53.

[52] Irene Lopatovska. 2011. Researching emotion: challenges and solutions. In *Proceedings of the 2011 iConference on - iConference '11*. ACM Press, New York, New York, USA. https://doi.org/10.1145/1940761

[53] Hector P. Martinez, Yoshua Bengio, and Georgios Yannakakis. 2013. Learning deep physiological models of affect. *IEEE Computational Intelligence Magazine* 8, 2 (2013), 20–33.

[54] Alban Maxhuni, Pablo Hernandez-Leal, L. Enrique Sucar, Venet Osmani, Eduardo F. Morales, and Oscar Mayora. 2016. Stress modelling and prediction in presence of scarce data. *Journal of Biomedical Informatics* 63 (oct 2016), 344–356.

[55] Microsoft. 2019. Microsoft Band 2 features and functions. support.microsoft.com/en-gb/help/4000313

[56] M. Molinara, L. Ferrigno, A. Maffucci, P. Kuzhir, R. Cancelliere, A. Di Tinno, L. Micheli, and M. Shuba. 2022. A Deep Transfer Learning Approach to an Effective Classification of Water Pollutants From Voltammetric Characterizations. *MELECON 2022 - IEEE Mediterranean Electrotechnical Conference, Proceedings* (2022), 255–259. https://doi.org/10.1109/MELECON53508.2022.9842896

[57] O. Hobart Mowrer. 1960. *Learning theory and behavior.* John Wiley & Sons Inc. https://doi.org/10.1037/10802-000

[58] Amir Muaremi, Bert Arnrich, and Gerhard Tröster. 2013. Towards Measuring Stress with Smartphones and Wearable Devices During Workday and Sleep. *BioNanoScience* (2013). https://doi.org/10.1007/s12668-013-0089-2

[59] Hong Wei Ng, Viet Dung Nguyen, Vassilios Vonikakis, and Stefan Winkler. 2015. Deep learning for emotion recognition on small datasets using transfer learning. In *ICMI 2015 - Proceedings of the 2015 ACM International Conference on Multimodal Interaction*. Association for Computing Machinery, Inc, New York, New York, USA, 443–449. https://doi.org/10.1145/2818346.2830593

[60] Xi Ouyang, Shigenori Kawaai, Ester Gue Hua Goh, Shengmei Shen, Wan Ding, Huaiping Ming, and Dong Yan Huang. 2017. Audio-visual emotion recognition using deep transfer learning and multiple temporal models. In *ICMI 2017 - Proceedings of the 19th ACM International Conference on Multimodal Interaction*, Vol. 2017-Janua. Association for Computing Machinery, Inc, New York, New York, USA, 577–582. https://doi.org/10.1145/3136755.3143012

[61] Sinno Jialin Pan and Qiang Yang. 2010. A survey on transfer learning. https://doi.org/10.1109/TKDE.2009.191

[62] Fatma Patlar Akbulut. 2022. Hybrid deep convolutional model-based emotion recognition using multiple physiological signals. *Computer Methods in Biomechanics and Biomedical Engineering* 25, 15 (2022), 1678–1690. https://doi.org/10.1080/10255842.2022.2032682

[63] Amorette Perkins, Joseph Ridler, Daniel Browes, Guy Peryer, Caitlin Notley, and Corinna Hackmann. 2018. Experiencing mental health diagnosis: a systematic review of service user, clinician, and carer perspectives across clinical settings. https://doi.org/10.1016/S2215-0366(18)30095-6

[64] Rosalind W. Picard. 2003. Affective computing: challenges. *International Journal of Human-Computer Studies* 59, 1-2 (jul 2003), 55–64. https://doi.org/10.1016/S1071-5819(03)00052-1

[65] Rui Qiao, Chunmei Qing, Tong Zhang, Xiaofen Xing, and Xiangmin Xu. 2017. A novel deep-learning based framework for multi-subject emotion recognition. In *ICCSS 2017 - 2017 International Conference on Information, Cybernetics, and Computational Social Systems*. IEEE, 181–185.

[66] Jie Lin Qiu, Xin Yi Qiu, and Kai Hu. 2018. Emotion recognition based on gramian encoding visualization. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 11309 LNAI (2018), 3–12.

[67] K. Radhika and Ramana V. Murthy Oruganti. 2020. Transfer Learning for Subject-Independent Stress Detection using Physiological Signals. *2020 IEEE 17th India Council International Conference, INDICON 2020* (dec 2020). https://doi.org/10.1109/INDICON49873.2020.9342505

[68] James A. Russell. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology* 39, 6 (dec 1980), 1161–1178. https://doi.org/10.1037/h0077714

[69] Björn Schuller, R A Calvo, S Dmello, J Gratch, and A Kappas. 2015. Multimodal affect databases: Collection, challenges, and chances. *Handbook of Affective Computing* (2015), 323–333.

[70] Johannes Schumm, Eth Zurich, Ulrike Ehlert, Cornelia Setz, Bert Arnrich, Roberto La Marca, and Gerhard Tröster. 2010. Discriminating stress from cognitive load using a wearable EDA device. *IEEE Transactions on Information Technology in Biomedicine* 14, 2 (2010).

[71] Monit Shah Singh, Vinaychandran Pondenkandath, Bo Zhou, Paul Lukowicz, Marcus Liwicki, and Tu Kaiserslautern. [n. d.]. Transforming Sensor Data to the Image Domain for Deep Learning-an Application to Footstep Detection. ([n. d.]). https://doi.org/10.1109/IJCNN.2017.7966182 arXiv:1701.01077v3

[72] Nandita Sharma and Tom Gedeon. 2012. Objective measures, sensors and computational techniques for stress recognition and classification: A survey. *Computer Methods and Programs in Biomedicine* 108, 3 (dec 2012), 1287–1301.

[73] Benjamin Shickel, Martin Heesacker, Sherry Benton, and Parisa Rashidi. 2017. Hashtag Healthcare: From Tweets to Mental Health Journals Using Deep Transfer Learning. (aug 2017).

[74] Karen Simonyan and Andrew Zisserman. 2015. Very deep convolutional networks for large-scale image recognition. In *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. arXiv:1409.1556

[75] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, and Ruslan Salakhutdinov. 2014. *Dropout: A Simple Way to Prevent Neural Networks from Overfitting*. Technical Report. 1929–1958 pages.

[76] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. 2016. Rethinking the Inception Architecture for Computer Vision. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/CVPR.2016.308 arXiv:1512.00567

[77] Joachim Taelman, S. Vandeput, A. Spaepen, and S. Van Huffel. 2008. Influence of mental stress on heart rate and heart rate variability. In *IFMBE Proceedings*, Vol. 22. Springer, Berlin, Heidelberg, 1366–1369. https://doi.org/10.1007/978-3-540-89208-3_324

[78] Muqin Tian, Qianqian Li, Chunyu Xv, Yubo Yang, and Zhehua Li. 2021. Coal-rock Interface Recognition Method Based on GAF-deep Learning. *5th IEEE Conference on Energy Internet and Energy System Integration: Energy Internet for Carbon Neutrality, EI2 2021* (2021), 4029–4033. https://doi.org/10.1109/EI252483.2021.9712901

[79] Wan Tian, Jiujing Wu, Hengjian Cui, and Tao Hu. 2021. Drought Prediction Based on Feature-Based Transfer Learning and Time Series Imaging. *IEEE Access* 9 (2021), 101454–101468. https://doi.org/10.1109/ACCESS.2021.3097353

[80] Tatiana Tommasi, Francesco Orabona, and Barbara Caputo. 2010. Safety in numbers: Learning categories from few examples with multi model knowledge transfer. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/CVPR.2010.5540064

[81] A. Tsanas, M. A. Little, and P. E. McSharry. 2013. A methodology for the analysis of medical data. *Handbook of Systems and Complexity in Health* (jan 2013), 113–125. https://doi.org/10.1007/978-1-4614-4998-0_7/FIGURES/00073

[82] A. Tsanas, K.E.A. Saunders, A.C. Bilderbeck, N. Palmius, M. Osipov, G.D. Clifford, G. Goodwin, and M. De Vos. 2016. Daily longitudinal self-monitoring of mood variability in bipolar disorder and borderline personality disorder. *Journal of Affective Disorders* 205 (nov 2016), 225–233.

[83] Terumi Umematsu, Akane Sano, Sara Taylor, and Rosalind W. Picard. 2019. Improving Students' Daily Life Stress Forecasting using LSTM Neural Networks. Institute of Electrical and Electronics Engineers (IEEE), 1–4.

[84] Jenny Venton, Philip J. Aston, Nadia A.S. Smith, and Peter M. Harris. 2020. Signal to Image to Classification: Transfer Learning for ECG. *2020 11th Conference of the European Study Group on Cardiovascular Oscillations: Computation and Modelling in Physiology: New Challenges and Opportunities, ESGCO 2020* (jul 2020). https://doi.org/10.1109/ESGCO49734.2020.9158037

[85] Jindong Wang, Yiqiang Chen, Vincent W. Zheng, and Meiyu Huang. 2018. Deep Transfer Learning for Cross-domain Activity Recognition. In *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3265689.3265705 arXiv:1807.07963

[86] Weinan Wang, Pedram Mohseni, Kevin L. Kilgore, and Laleh Najafizadeh. 2022. Cuff-Less Blood Pressure Estimation From Photoplethysmography via Visibility Graph and Transfer Learning. *IEEE Journal of Biomedical and Health Informatics* 26, 5 (may 2022), 2075–2085. https://doi.org/10.1109/JBHI.2021.3128383

[87] Zhiguang Wang and Tim Oates. 2015. Imaging time-series to improve classification and imputation. In *IJCAI International Joint Conference on Artificial Intelligence*. arXiv:1506.00327

[88] B Weiner and S Graham. 1985. An attributional approach to emotional development. In *Emotions, cognition, and behavior*.

[89] Wanhui Wen, Guangyuan Liu, Nanpu Cheng, Jie Wei, Pengchao Shangguan, and Wenjin Huang. 2014. Emotion recognition based on multi-variant correlation of physiological signals. *IEEE Transactions on Affective Computing* (2014). https://doi.org/10.1109/TAFFC.2014.2327617

[90] Hans Weytjens and Jochen De Weerdt. 2021. Process Outcome Prediction: CNN vs. LSTM (with Attention). *Lecture Notes in Business Information Processing* 397 (apr 2021), 321–333. https://doi.org/10.1007/978-3-030-66498-5_24

[91] WHO and World Health Organisation. 2014. WHO | Mental health: a state of well-being. https://www.who.int/features/factfiles/mental_health/en/

[92] J. Wijsman, B. Grundlehner, Hao Liu, H. Hermens, and J. Penders. 2011. Towards mental stress detection using wearable physiological sensors. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 1798–1801.

[93] Kieran Woodward, Eiman Kanjo, David Brown, T. M. McGinnity, Becky Inkster, Donald J Macintyre, and Athanasios Tsanas. 2020. Beyond Mobile Apps: A Survey of Technologies for Mental Well-being. *IEEE Transactions on Affective Computing* (2020).

[94] Kieran Woodward, Eiman Kanjo, David J. Brown, and T.M. McGinnity. 2021. Towards Personalised Mental Wellbeing Recognition On-Device using Transfer Learning "in the Wild". In *IEEE International Smart Cities Conference 2021*.

[95] Kieran Woodward, Eiman Kanjo, Andreas Oikonomou, and Alan Chamberlain. 2020. LabelSens: enabling real-time sensor data labelling at the point of collection using an artificial intelligence-based approach. *Personal and Ubiquitous Computing* 24, 5 (jun 2020), 709–722.

[96] Xiaofen Xing, Zhenqi Li, Tianyuan Xu, Lin Shu, Bin Hu, and Xiangmin Xu. 2019. SAE+LSTM: A new framework for emotion recognition from multi-channel EEG. *Frontiers in Neurorobotics* 13 (2019).

[97] Chen Yi Chao Lung Yang, Chen Yi Chao Lung Yang, Zhi Xuan Chen, and Nai Wei Lo. 2019. Multivariate Time Series Data Transformation for Convolutional Neural Network. In *Proceedings of the 2019 IEEE/SICE International Symposium on System Integration, SII 2019*. Institute of Electrical and Electronics Engineers Inc., 188–192. https://doi.org/10.1109/SII.2019.8700425

[98] Xin Yu and Panos D. Prevedouros. 2013. Performance and Challenges in Utilizing Non-Intrusive Sensors for Traffic Data Collection. *Advances in Remote Sensing* (2013). https://doi.org/10.4236/ars.2013.22006

[99] Alexandros Zenonos, Aftab Khan, Georgios Kalogridis, Stefanos Vatsikas, Tim Lewis, and Mahesh Sooriyabandara. 2016. HealthyOffice: Mood recognition at work using smartphones and wearable sensors. In *2016 IEEE International Conference on Pervasive Computing and Communication Workshops, PerCom Workshops 2016*. https://doi.org/10.1109/PERCOMW.2016.7457166

[100] Jing Zhang, Yixin Wang, and Guiyan Wei. 2023. Global Spatial Representation: EEG Correcting for Subject-Independent Emotion Recognition. In *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*. Springer, Cham, 385–396. https://doi.org/10.1007/978-3-031-36819-6_34

[101] Ran Zhang, Hongyang Tao, Lifeng Wu, and Yong Guan. 2017. Transfer Learning with Neural Networks for Bearing Fault Diagnosis in Changing Working Conditions. *IEEE Access* 5 (2017), 14347–14357.

[102] Yaqing Zhang, Jinling Chen, Jen Hong Tan, Yuxuan Chen, Yunyi Chen, Dihan Li, Lei Yang, Jian Su, Xin Huang, and Wenliang Che. 2020. An Investigation of Deep Learning Models for EEG-Based Emotion Recognition. *Frontiers in Neuroscience* 14 (dec 2020), 1344. https://doi.org/10.3389/FNINS.2020.622759/XML/NLM

[103] Wei-Long Zheng and Bao-Liang Lu. 2016. Personalizing EEG-Based Affective Models with Transfer Learning. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16)*.

[104] Jan Henrik Ziegeldorf, Oscar Garcia Morchon, and Klaus Wehrle. 2014. Privacy in the internet of things: Threats and challenges. *Security and Communication Networks* (2014). https://doi.org/10.1002/sec.795

[105] M. Sami Zitouni, Cheul Young Park, Uichin Lee, Leontios J. Hadjileontiadis, and Ahsan Khandoker. 2023. LSTM-Modeling of Emotion Recognition Using Peripheral Physiological Signals in Naturalistic Conversations. *IEEE Journal of Biomedical and Health Informatics* 27, 2 (feb 2023), 912–923. https://doi.org/10.1109/JBHI.2022.3225330