IEEE Sensors Council

# iFidgetCube: Tangible Fidgeting Interfaces (TFIs) to Monitor and Improve Mental Wellbeing

Kieran Woodward and Eiman Kanjo

Abstract—The ability to unobtrusively measure mental wellbeing states using non-invasive sensors has the potential to greatly improve mental wellbeing by alleviating the effects of high stress levels. Multiple sensors, such as electrodermal activity, heart rate and accelerometers, embedded within tangible devices pave the way to continuously and non-invasively monitor wellbeing in real-world environments. On the other hand, fidgeting tools enable repetitive interaction methods that may help to tap into individual's psychological need to feel occupied and engaged; hence potentially reducing stress. In this article, we present the design, implementation, and deployment of Tangible Fidgeting Interfaces (TFIs) in the form of computerised iFidgetCubes. iFidgetCubes embed non-invasive sensors along with fidgeting mechanisms to aid relaxation and ease restlessness. We take advantage of our labeling techniques at the point of collection to implement multiple subject-independent deep learning classifiers to infer wellbeing. The obtained performance demonstrates that these new forms of tangible interfaces combined with deep learning classifiers have the potential to accurately infer wellbeing in addition to providing fidgeting tools.

iFidgetCube: Tangible Fidgeting Interfaces (TFIs) to Improve Wellbeing

Fidgeting — Embedded fidgeting tools to aid relaxation and reduce stress

Sensors — Real-world data collected from HR, EDA and IMU sensors

Deep Learning Classifiers — 81.9% Average 1D CNN accuracy — 8 classifiers tested with 9 users' real-world labelled data

Index Terms—Mental wellbeing, fidgeting, emotion recognition, physiological sensors, deep learning, subject-independent, tangible user interfaces.
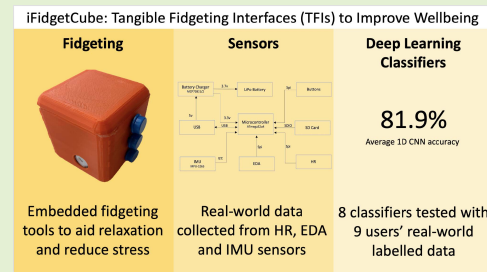
I. INTRODUCTION

PERSONAL health monitoring has the potential to measure and reduce stress resulting in significant improvements in mental wellbeing. Advances in non-invasive physiological sensors have created the potential to monitor real-world mental wellbeing and potentially improve quality of life by providing interventional feedback [1].

Measuring wellbeing is more important than ever with modern lifestyles contributing to increased daily stress as 59% of UK adults experience work-related stress [2], costing the economy £2.4 billion [2] each year. Furthermore, students studying higher degrees experience high stress levels [3] with 32% of college students suffering from mental health issues [4]. Majority of high-risk undergraduate students rate their mental health as poor or very poor [5] and less than 20% of college students with mental health issues received treatment [6].

Traditional mental wellbeing assessment methods require people to be aware of their mental health and seek help which can be challenging due to social stigma and lack of available resources [7]. The decreasing cost and increasing capability of sensors and edge computing is enabling new forms of interfaces which are more powerful and dynamic than traditional assessment technologies. A technological alternative that could actively monitor an individual's mental health state and provide wellbeing interventions would be extremely beneficial in improving accessibility to mental health tools for all [8].

Physiological and motion sensors present a more objective method to measure wellbeing. Recent developments in non-invasive sensors paired with deep learning classifiers introduce the possibility to quantify mental wellbeing in real-time. Deep learning enables models to be trained using raw sensor data unlike machine learning classifiers and with advances in deep learning architectures the accuracy in which mental wellbeing can be classified may be improved.

When stressed people commonly fidget with objects such as pens, as fidgeting is a natural response that demonstrates the potential to regulate stress [9], [10] and improve information retention [11]. Recently, fidgeting cubes have begun to increase in popularity; they are small plastic cubes whose sides provide sensory tools to facilitate fidgeting and help normalise stimming (self-stimulatory behavior such as tapping or clicking). Fidgeting cubes offer a variety of sensory actions catering for a wide range of needs in a small, unobtrusive design. An example of a fidget cube is shown in Figure 1.

Fig. 1. Example of traditional fidget cube.

Most previous affective computing approaches have focused on wearable devices however, wearables present many challenges such as poor battery life and limited space to embed additional sensors.In this article, we introduce Tangible Fidgeting Interfaces (TFIs) which are physical fidgeting devices that enable repetitive physical interaction while also enabling objective sensor measurement. To our knowledge this article presents the first device combining traditional fidgeting cubes with a microcontroller and non-invasive physiological sensors. Furthermore, we show how these fidgeting interfaces can be coupled with deep learning algorithms, paving the way for a new type of real-time interaction. TFIs can vary in form but by developing handheld interfaces that embed the necessary sensors, it is possible to develop devices that encourage engagement and improve wellbeing through fidgeting, acting as a preventative tool.

In the remainder of this article we explore the development of wellbeing TFIs, followed by a pilot study of users collecting real-world labelled data. Experimental results explores 8 deep learning classifiers trained using the real-world datasets collected to infer wellbeing, followed by evaluating their capabilities and comparing performance. In the future, these deep learning algorithms can be used to control actuated feedback and allow people to act on their current state of wellbeing.

II. BACKGROUND

A. Tangible Fidgeting Interfaces (TFIs)

We propose TFIs as Tangible User Interfaces (TUIs) that enable physical interaction with digital information while also providing fidgeting mechanisms powered by intelligent algorithms. TFIs can take any form enabling them to ubiquitously become part of everyday interactions. They can be of any shape (e.g. cube, ball or polytope) and can be made of hard or soft material (squeezable ball or clicking tool). By including fidgeting mechanisms within TUIs they can act as a distraction to any wellbeing challenges encountered, which is often used as a coping strategy to reduce stress [12], [13].

Previous fidgeting tools have only been used to self-report emotions through squeezing [14] and display emotions as different colours, enabling users to privately share their wellbeing [15]. Grasp [16] measured how much force was exhorted when participants squeezed the device to report stress. This enabled users to easily record their stress levels, providing an intuitive method of interaction. However, there has been less focus on additionally collecting objective sensor data that can be used to train machine learning algorithms to infer wellbeing [17].

A hybrid data collection approach is most suitable when collecting wellbeing data due to the subjectivity of the data.

Self-reporting of wellbeing can be combined with the passive collection of sensor data to collect a labelled dataset [18], which is vital due to the individualisation of emotions. TFIs are ideal to embed the necessary sensors to infer wellbeing along with a labelling mechanism while also taking advantage of embedded actuation. These digitally enabled fidgeting mechanisms will provide a distraction and unmet need for people with persistent anxiety or stress.

B. Mental Wellbeing Inference

In order to provide direct and intuitive feedback to users, TFIs need to take advantage of real-time algorithms to automatically infer mental wellbeing from sensor data. Electrodermal Activity (EDA) and Heart Rate (HR) sensors are especially beneficial as they directly correlate with the sympathetic nervous system [19]–[21]. EDA and Heart Rate Variability (HRV) have previously been used to measure stress over 5 minute time frames achieving 97.4% accuracy [22]. The results showed HRV and EDA are highly beneficial when inferring stress, making them ideal for mental wellbeing interfaces.

Neural networks have the potential to increase model performance by using the raw sensor data, removing the necessity for feature extraction. ElectroEncephaloGraphy (EEG) data was used to train a Long Short-Term Memory (LSTM) model to infer valence and arousal, achieving 81.1% accuracy [23]. Similarly, EDA, skin temperature, motion phone usage data has been used to train an LSTM network to infer stress achieving 81.4% accuracy, outperforming comparative machine learning models [24].

Convolutional Neural Networks (CNNs) have also been used to infer wellbeing. EDA and blood volume pulse data was used to train a one dimensional CNN (1D CNN) to classify relaxation, anxiety excitement [25] achieving accuracies between 70-75%. Additionally, EEG data was used to train a CNN to infer valence and arousal using channel selection strategy, where the strongest correlated channels generate the training set, achieving 87.27% accuracy, an increase of nearly 20% [26]. Furthermore, 1D CNNs have been used with a transfer learning approach to increase affective model personalisation, achieving 93.9% accuracy when tested with 3 users [27].

A combined CNN and LSTM model has been trained using raw physiological and environmental data to infer 5 emotions [28]. The combined model outperformed other deep learning models by around 20%. Neural networks have improved the accuracy in which mental wellbeing can be classified, although they require large training datasets which can be challenging to collect.

C. Sensors

Non-invasive sensors present the most significant opportunity to assess mental wellbeing as they can easily be embedded within TFIs and used in the real-world. Non-invasive physiological measures for mental wellbeing include EDA and HR due to their high correlation with the sympathetic nervous system [19].
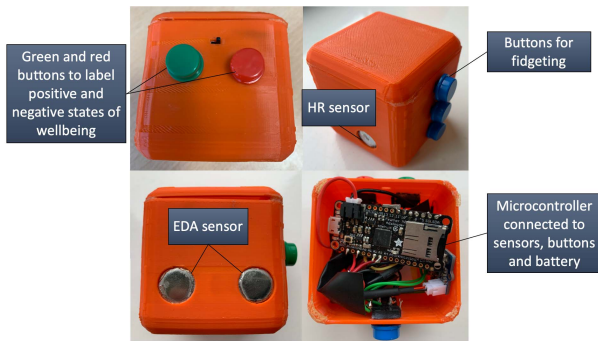
Fig. 2. iFidgetcube showing labelling buttons (left) and fidgeting buttons and HR sensor (right).

Fig. 3. Tangible fidgeting interface schematic showing how the battery, sensors, SD card and buttons connect to the microcontroller.

1) Heart Rate (HR): HR sensors are commonly used within wearable computing as they can be embedded within a wide range of devices due to their small footprint. Similarly, HRV is commonly used within affective computing as it is the variation in time between heartbeats, often indicating stress [29]. HRV can be accurately measured using Photo-PlethysmoGgraphy (PPG) which is easier and cheaper to use than ElectroCardioGraphy (ECG) as it only requires 1 contact point. There have been three main forms of PPG developed: transmitted, reflected and remote. Transmitted signals are most commonly used in medical monitoring, remote signals often utilise cameras to measure changes in skin colour and reflected signals measure the reflected light from an LED using photodiodes. Reflection PPG is the the smallest and most convenient method to measure HR and HRV within TFIs [30].

2) Elecotrodermal Activity (EDA): EDA is often used to train affective models to infer mental wellbeing as it directly correlates to the sympathetic nervous system [20]. Alternatively, near-infrared spectroscopy can be used to measure oxyhemoglobin and deoxyhemoglobin enabling the inference of stress with similar levels of accuracy as EDA [31]. However, near-infrared spectroscopy cannot be used to collect data in the real-world due to its large size and placement on the forehead.

3) Motion: Motion data collected through accelerometers, gyroscopes and magnetometers could be used in addition to physiological sensor data to infer wellbeing. Previous work has used motion data to infer emotions with 81.2% accuracy across 3 classes [32]. However, other work has reported lower levels of accuracy when inferring emotions from motion data alone ranging from 50% to 72% [33]–[35].

III. PROPOSED ARCHITECTURE

A. iFidgetCube

In this article we introduce iFidgetCube, a TFI in the shape of a cube. This small plastic cube's various sides provide sensory tools such as buttons, as shown in figure 2. Unlike traditional fidgeting cubes, iFidgetCube embeds a microcontroller and non-invasive sensors. The size of the cube is 6cm × 6cm × 6cm making it handheld and ideal to embed all of the necessary sensors. Two buttons (1 green and 1 red) were also embedded within the interfaces to enable the real-time labelling of positive and negative states of wellbeing respectively [36].
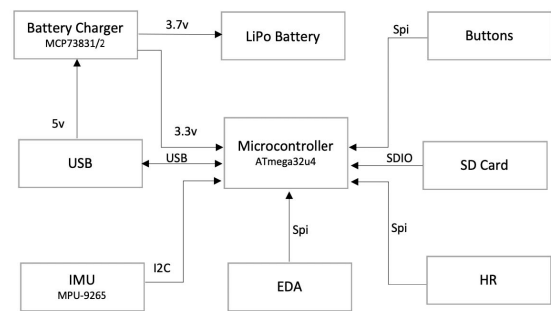
Multiple physiological sensors are included within the interfaces to measure wellbeing. The HR data is obtained via a PPG sensor embedded within an indent for thumb placement measuring Beats Per Minute (BPM), raw signal amplitude and HRV. Similarly, the EDA sensors are located on the opposite face of the PPG sensor where two fingers can be placed to comfortably hold the device while simultaneously recording the sensor data.

The iFidgetCube contains 3 buttons that can be used for fidgeting similar to fidgeting toys along with a 9 degree-of-freedom inertial motion unit (9-DOF IMU) to capture the fidgeting motion of interactions. The IMU is a MPU-9265 and consists of an accelerometer, a gyroscope, and a magnetometer operated at 3.3v, footprint of 22mm × 17mm.

The sensors are connected to a small Arduino-compatible ATmega32u4 based processor to process all of the data, with a footprint of 28.8mm × 33.1mm, powered by a small size 3.7v lithium polymer battery as shown in figure 3.

IV. METHODOLOGY AND EXPERIMENT

A. Experimental Setup

A total of 14 participants (8 Males and 6 females) were provided with an iFidgetCube containing HR, HRV and EDA physiological sensors in addition to the 9-DOF IMU. Furthermore, green and red buttons were present on each interface to enable the real-time self-labelling of positive (happiness, joy, relaxed) and negative (sadness, anger, stress) states of wellbeing. The addition a green and red button provides a simple way for users to label the data in real-time as it is not possible to label sensor data after the point of collection unlike other data types such as images [37]. Participants were instructed how to correctly label their emotions and to fidget with the cubes as frequently as possible throughout their normal daily life to collect the real-world labelled data required to train the classification models.

Nine of the participants have complex learning and physical disabilities resulting in mental wellbeing challenges often being diagnostically overshadowed. Therefore, a device that could simultaneously monitor wellbeing in addition to providing fidgeting tools could be extremely beneficial. However, the data collected from 5 of these participants was inadequate to train the models due to an overwhelming majority of the recorded labels being positive emotions resulting in limited biased datasets. The data from five participants with no
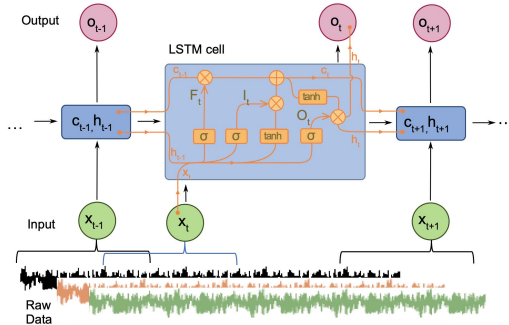
Fig. 4. LSTM cell showing input vector (X), cell output (h), cell memory (c) and input (I), forget (f) and output (o) gates.

disabilities (users 1, 2, 3, 6 and 7) and four participants with learning and physical disabilities (users 4, 5, 8 and 9) has been used to examine deep learning affective models.

All participants used the cubes during their daily life to collect real-world labelled wellbeing data over a period of 2 weeks. This differs from many previous studies that collected controlled experimental data or include specified activities during the data collection period to artificially impact wellbeing.

B. Classifiers

The real-world data consisting of HR (BPM), HRV, HR (amplitude), EDA and motion collected from the cubes was cleaned and normalised before being used to train 8 deep learning classifiers. Four of the tested networks (ResNet, TWIESN, Encoder and MCDCNN) have been previously explored using time series data [38] however, their effectiveness for affective modeling has not been explored. These four models have been compared with an additional 4 models (1D CNN, LSTM, CapsNet and Inception) to evaluate their affective modelling performance.

1) Long-Short Term Memory (LSTM): Recurrent Neural Networks (RNNs) utilise the temporal correlations between neurons. LSTM [39] cells are often combined with RNNs where they serve as the memory units through gradient descent. LSTM cells use input (I), forget (f) and output (o) gates to regulate the flow of information, as shown in figure 4, helping remove the vanishing gradient problem faced by traditional RNNs.

The LSTM network comprises of 2 hidden layers with batch normalization being performed to prevent the internal covariate shifting across one mini-batch training set [40]. A SoftMax layer follows using a cross entropy loss function to produce a predicted label from the available classes.

2) 1 Dimension CNN (1D CNN): A convolution involves sliding a filter over the time series data although unlike images where CNNs are traditionally used, the filters exhibit only one dimension instead of two dimensions. A general form of applying the convolution for a time stamp t is given in the following equation:

$$C_t = f(\omega X_{tl/2:t+l/2} + b)|t[1, T]$$

where C denotes the result of a convolution applied on a time series X of length T with filter ω, bias b and a non-linear

function f such as Rectified Linear Unit (ReLU). Weight sharing enables the same convolution to be used to find the result for all time stamps, allowing filters that are invariant across the time dimension to be learned.

Pooling is then performed, this can include local pooling such as max pooling where a sliding window aggregates the input data reducing it by length T. Alternatively, global pooling can be performed where the data is aggregated over the entire dimension resulting in a single value. In addition to pooling layers, normalization layers have been used to help the network converge quicker and batch normalization has also been performed. The final layer takes the result of the convolutions and outputs a probability distribution using the SoftMax activation function and a cross entropy loss function.

3) Capsule Network (CapsNet): CapsNets [41] are comprised of capsules, where each capsule encompasses a group of neurons in a layer which perform internal computations to predict the presence and instantiation parameters. CapsNets aim to preserve hierarchical spatial relationships helping to learn faster and use fewer samples per class with a 1 dimensional CapsNet ideal for timeseries data, being recently introduced [42].

The first 3 layers within the CapsNet are to encode, and the second 3 are to decode. The first layer is a traditional convolutional layer followed by a PrimaryCaps layer. The PrimaryCaps layer contains primary capsules who take basic features detected by the convolutional layer and produce combinations of the features. Next, the DigitCaps layer accepts inputs from all of the capsules in the previous layer. Non-linear activations at both the Primary and DigitCaps layer are provided by the squash function. Connections between these two layers are dynamic and are governed by dynamic routing.

$$L_k = T_k max(0, m^+ - ||v_k||)^2 + (1 - T_k)max(0, ||v_k|| - m^-)^2$$

Dynamic routing allows weights to decide which higher level capsule the current capsule will send it's output to [43]. This is done by lower level capsules sending their input to higher level capsules that "agree" with the input. Two CapsNet loss functions are then used, as shown above, to equivariance between capsules and calculate the correct DigiCap. Finally, three fully connected layers decode the vector from the correct DigitCap and provide the output of the network as a vector.

4) Residual Network (ResNet): A proposed architecture from [44] is ResNet. The network is composed of three residual blocks followed by a global average pooling layer and a final SoftMax classifier using a cross entropy loss function, whose number of neurons is equal to the number of classes. Batch normalisation and ReLU activation function follow. Each residual block is comprised of three convolutions whose output is added to the residual block's input and then fed to the next layer. The fundamental feature of a ResNet is a linear shortcut to link the output of a residual block to its input enabling the direct flow of the gradient through the connections, removing the vanishing gradient problem [45].

5) Time Warping Invariant Echo State Network (TWIESN): TWIESN [46] is the second RNN explored. For each element in an input time series, the reservoir space is used to project this element into a higher dimensional space. Then for

TABLE I

COMPARISON OF DEEP LEARNING MODELS' ACCURACY TESTED USING LOOCV ON INDIVIDUAL USERS' PHYSIOLOGICAL DATA

	1D CNN	LSTM	CapsNet	ResNet	TWIESN	Encoder	Inception	MCDCNN
User 1	**74.1%**	64.2%	53.1%	67.8%	59.8%	47.9%	68.8%	47.9%
User 2	**85.1%**	68.4%	58.5%	59.6%	54.4%	47.1%	59.3%	56%
User 3	**72.5%**	66.7%	63.2%	35.9%	25.4%	36.8%	36%	36.8%
User 4	**85.5%**	56.4%	67.7%	33.5%	52.5%	9.4%	36.4%	9.5%
User 5	**86.2%**	74.7%	53.1%	51.1%	52.7%	53%	45.8%	53%
User 6	**83%**	55%	81%	24.8%	22.7%	21.6%	25.6%	78.4%
User 7	81.1%	50%	17.4%	**83.3%**	21%	21%	81.6%	21%
User 8	**85.3%**	68.3%	54.8%	40.2%	43.2%	45.3%	43.2%	54.7%
User 9	83.8%	59.9%	**88%**	47.4%	87.6%	87.7%	43.5%	87.7%
Average	**81.9%**	62.6%	58.5%	49.1%	46.9	41.1%	54.8%	40.7%

each element, a ridge classifier [47] is trained to predict the class of each time series element. During testing, the ridge classifier outputs a probability distribution over the classes in a dataset. Then the posteriori probability for each class is averaged assigning a label for each test set where the averaged probability is highest.

6) Encoder: An Encoder network [48] is a hybrid CNN [44] but where the global average pooling layer is replaced with an attention layer, enabling invariance across all layers. The first three layers are convolutional. Each convolution is followed by batch normalisation and then Parametric Rectified Linear Unit (PReLU) [49] activation function. The output of PReLU is followed by a dropout layer and a final max pooling layer. The third convolutional layer is fed to an attention layer [50] that enables the network to learn the most important aspects for classification. Finally, a SoftMax layer with a cross entropy loss function is used to produce a predicted label from the available classes.

7) InceptionTime: InceptionTime [51] is an ensemble of CNN models, inspired by the Inception-v4 architecture. The Inception network uses a cross entropy loss function and contains two different residual blocks comprising of three Inception modules. Each residual block's input is transferred via a shortcut linear connection added to the next block's input, enabling a direct flow of the gradient and removing the vanishing gradient problem.

The first component within the Inception module is the "bottleneck" layer. This layer performs an operation of sliding filters of length 1 with a stride equal to 1, allowing for longer filters than ResNet. The next layer involves sliding multiple filters of different lengths simultaneously on the same input data. The output of the sliding MaxPooling window is then calculated and the output of each independent parallel MaxPooling layer is concatenated. By training the weights of multiple inception models using filters of varying lengths, the network is able to extract latent hierarchical features.

8) Multi Channel Deep Convolutional Neural Network (MCD-CNN): MCDCNN is a CNN where the convolutions are applied independently on each dimension [52]. Each dimension of input data passes through two convolutional layers with ReLU as the activation function followed by a MaxPooling operation. The output of the second convolutional layer for all dimensions is concatenated over the channels axis and then fed to a fully connected layer with ReLU as the activation function before the SoftMax classifier using a cross entropy loss function.

V. EXPERIMENTAL RESULTS

A. Multivariate Models

Initially 6 frequently used machine learning classifiers were trained using the physiological sensor data to enable comparisons with the deep learning classifiers. The machine learning classifiers included: Naive Bayes, Logistic Regression, Fast Large Margin, Decision Tree, Random Forest and Support vector machine along with a CNN for direct comparison. Hold-out validation was used with a 30% test split and the features were selected on an individual basis for each model using multi-objective evolutionary automatic feature engineering. The average result for the models along with the extracted features are shown in table III.

The best performing machine learning model was Decision tree achieving an average of 80.3% accuracy using HRV, EDA, $EDA/HR-EDA$ and $EDA/HR/HR$ as the features. The CNN achieved 88.1% accuracy demonstrating its ability to improve upon the traditional machine learning classifiers while also enabling the raw sensor data to be used to train models. As the CNN showed, deep learning classifiers have the capability to increase the accuracy of real-world affective modelling.

The real-world physiological data (HR (BPM), HRV, HR (amplitude) and EDA) was then used to train each of the aforementioned deep learning networks. Each model was tested using Leave-One-Out Cross Validation (LOOCV) on a subject-independent basis where the model was tested with each user's data separately after being first trained with the remaining users' data over 50 epochs. This method of testing accurately measures model performance on an individual basis and better simulates real-world performance. The results for each model tested using LOOCV is shown in table I where the highest accuracy for each user is highlighted.

The results show that for the physiological model the 1D CNN outperformed all other models for the majority of users, achieving between 72.5% and 86.2% accuracy with an average of 81.9%. The average accuracy when testing against users was 6.2% lower than when using hold-out validation but demonstrated consistent results across all users, closer simulating real-world usage. CapsNet outperformed the 1D CNN by 4.2% for user 9, achieving the highest overall accuracy of 88%, while ResNet outperformed the 1D CNN for user 7 by 2.2%.

Motion data was also collected from the cubes, however the motion data from 3 of the devices was corrupted therefore not usable to train models. Table II shows the results for each

TABLE II
COMPARISON OF DEEP LEARNING MODELS' ACCURACY TESTED USING LOOCV ON INDIVIDUAL USERS' PHYSIOLOGICAL AND MOTION DATA

	1D CNN	LSTM	CapsNet	ResNet	TWIESN	Encoder	Inception	MCDCNN
User 4	**87%**	52.6%	9.4%	79.7%	54.5%	62.2%	28.9%	49.7%
User 5	**79%**	57.8%	53.1%	47.4%	53%	52.8%	50.3%	53.9%
User 6	**81.1%**	72.3%	81%	45.6%	21.7%	24.2%	41.7%	78.4%
User 7	72%	29.2%	**82.6%**	40.2%	25.2%	21%	79.2%	79%
User 8	**54.8%**	49.2%	45.2%	36.7%	45.2%	44.9%	39.9%	36.9%
User 9	50%	59.9%	12%	42.3%	**87.6%**	80%	42.8%	80.4%
Average	**70.7%**	53.5%	47.2%	48.7%	47.9%	47.5%	47.1%	63%

TABLE III
COMPARISON OF COMMON MACHINE LEARNING MODELS' ACCURACY TESTED USING HOLD-OUT VALIDATION

Model	Extracted features	Average accuracy
Naive Bayes	$EDA * EDA$	72.2%
Logistic Regression	$EDA * EDA, HR$	71.8%
Fast Large Margin	$EDA/HR, EDA, HR$	62.5%
Decision Tree	$HRV, EDA/HR - EDA, EDA, EDA/HR/HR$	80.3%
Random Forest	$HRV, EDA, HR - EDA, \sqrt{(EDA)}$	77.9%
Support vector machine	$EDA, HRV, HR, \exp(HRV)$	79.5%
1D CNN	EDA, HR, HRV	88.1%

TABLE IV
COMPARISON OF UNIVARIATE 1D CNNs ACCURACY TESTED USING LOOCV ON INDIVIDUAL'S HR (BPM), HRV, EDA OR MOTION DATA

User	HR	HRV	EDA	Motion
User 1	60%	66%	**75%**	N/A
User 2	65%	62%	**86%**	N/A
User 3	63%	56%	**86%**	N/A
User 4	50%	56%	72%	**86%**
User 5	78%	78%	77%	**80%**
User 6	74%	75%	**77%**	73%
User 7	69%	66%	**75%**	28%
User 8	74%	75%	**80%**	27%
User 9	68%	**71%**	70%	69%
Average Accuracy	66.8%	67.2%	**77.6%**	60.5%

of the deep learning models when trained using physiological and motion data. The 1D CNN model again outperformed all other models for the majority of users, achieving an average accuracy of 70.7%. However, the CapsNet model achieved the highest accuracy for user 7 and TWIESN for user 9. Combining the physiological data with motion resulted in overall reduced performance although increased accuracy for user 4 by 1.5%.

B. Univariate Models

As the 1D CNN outperformed all other models for the majority of users, this model was further explored to examine the impact training using each individual data source has on performance. The 1D CNN was again tested on a subject-independent basis using LOOCV for each of the 9 users with either the HR (BPM), HRV, EDA or motion data used to train the models, as shown in table IV.

The results show that high model accuracy of up to 86% can be achieved using only 1 data source with EDA being the most accurate univariate model for the majority of users, achieving an average of 77.6% accuracy. This demonstrates

the importance of using EDA sensors when inferring wellbeing. Motion achieved the lowest average accuracy of 60.5% although the motion model was the most accurate univariate model for users 4 and 5 demonstrating the possibility of inferring wellbeing from motion data alone.

C. Reflection

The real-world data collected from the interfaces was used to train a variety of deep learning models. The results show that when inferring wellbeing using physiological data, the 1D CNN model outperformed all other models for the majority of users with an average accuracy of 81.9%, 19.3% higher than the next best performing model, LSTM. It is surprising the LSTM model performed significantly worse than the 1D CNN as RNNs are most commonly used with time series data. The 1D CNN achieved accuracies between 72.5% and 86.2% with a standard deviation of 5.1. This shows that wellbeing is highly personal as there was a 13.7% variation in accuracy between users when tested using the same 1D CNN model.

While there is a 13.7% range of accuracy for the 1D CNN, the two next best performing physiological models, LSTM and CapsNet show much wider variation of 24.7% for LSTM (SD 7.86) and 70.6% for CapsNet (SD 20.1). The model accuracy for CapsNet ranges between 17.4% to the highest accuracy achieved of 88% and while the 17.4% accuracy is an outlier, all other models achieved higher performance with the same user's data. This demonstrates that while CapsNet can be used to infer wellbeing, its high volatility results in inadequate subject-independent models.

When testing the same models with combined physiological and motion data the 1D CNN again outperformed all other models for the majority of users, although there was an 11.3% reduction in average accuracy compared with the physiological 1D CNN model. All users other than user 4 achieved lower performance with the combined motion and physiological data. When comparing the best performing model for each user between the physiological and combined physiological and motion models, it shows there is a wide variation between a 30.5% decrease and a 1.5% increase in performance with an average decrease of 6.5% for the combined motion and physiological models. This demonstrates that for the majority of users the inclusion of motion data has a detrimental impact on the inference of mental wellbeing.

The results show a wide variation in accuracy between the models on the same data. For the physiological models the standard deviation of the average model accuracy was 13.6 ranging between 10 and 30.7 for each model. Seven users

experienced higher standard deviance between models than the overall average, with user 7 experiencing the greatest deviation of 30.7 with a 65.9% variance in model accuracy. Similarly the combined motion and physiological models also demonstrated high deviations between models with the average standard deviation between models being 19.4 ranging from 8.9 to 27.1. The high deviation between models when tested on the same data shows the importance model selection has on performance.

As the 1D CNN significantly outperformed all other models, the model was further trained using data from individual sensors. EDA univariate models outperformed the other univariate models for 6 out of 9 users, with the highest average accuracy of 77.6%, demonstrating its importance in inferring wellbeing. HRV was the highest performing univariate model for user 9, whereas motion was the highest performing model for users 4 and 5, which was unexpected as motion frequently degraded performance in the multivariate models.

Surprisingly the EDA univariate models for users 1, 2 and 3 all outperformed the comparative multivariate physiological 1D CNNs by 0.9%, 0.9% and 13.5% respectively. However, the average overall accuracy for the univariate models does not outperform the average accuracy of 81.9% for the multivariate model, demonstrating multivariate physiological models are most applicable for the majority of users. Overall, these results show that while multivariate physiological models provide the highest accuracy for the majority of users, univariate affective models can improve performance for individual users.

VI. DISCUSSION AND FUTURE WORK

When stressed people often fidget with physical objects to provide a distraction, in turn improving their mental wellbeing. We have presented Tangible Fidgeting Interfaces as a technological alternative to traditional fidgeting tools in the shape of cubes (i.e. iFidgetCube). Each cube embedded sensors measuring EDA, HR, HRV and motion along with a tangible method to label the real-time data in additional to fidgeting tools.

After using the cubes all users stated they enjoyed using them and found them easy to handle. As users only recorded their emotion before fidgeting it is not possible to use the labelled data to explore the impact fidgeting had. However, user's described the devices as *"calming"* and *"relaxing"* due to being able to fidget by moving the cubes as well as pressing the buttons. Overall, users enjoyed using the interfaces and while TFIs are not as ubiquitous as wearable computing, they provide new opportunities to help relax users through fidgeting and embed sensors that are often not included within wearables, such as EDA.

The real-world data collected from iFidgetCubes was used to train 8 deep learning subject-independent classifiers to infer mental wellbeing including CNN, CapsNet, ResNet, LSTM, TWIESN, Encoder, Inception and MCDCNN. The results showed that the 1D CNN outperformed all other models for the majority of users, achieving an average accuracy of 81.9%. Univariate 1D CNN models, trained using a single data source were also explored, demonstrating EDA alone can achieve high performing models with an average accuracy of 77.6%.

Surprisingly the univariate models for 3 users outperformed their comparative multivariate physiological model showing additional physiological sensors do not necessarily increase wellbeing inference for all users.

Overall, iFidgetCubes have demonstrated their ability to aid labelled wellbeing sensor data collection, while simultaneously providing a fidgeting interface as a preventative mechanism of worsening mental wellbeing. The data collected from the devices for most users was sufficient to test a range of deep learning models where a 1D CNN demonstrated the highest overall performance.

In the future, TFIs should be trialled with more users over a longer period of time to collect additional data and further explore the impact of fidgeting on wellbeing. In addition, there is a need to incorporate different forms of actuation (e.g. vibration) to enable real-time physical feedback and allow people to act on their current state of wellbeing.

REFERENCES

[1] K. Woodward, E. Kanjo, M. Umair, and C. Sas, "Harnessing digital phenotyping to deliver real-time interventional bio-feedback," in *Proc. ACM Int. Joint Conf. Pervas. Ubiquitous Comput. Proc. ACM Int. Symp. Wearable Comput. (UbiComp/ISWC)*, 2019, pp. 1206–1209.

[2] Perkbox. (2018). *The 2O18 UK Workplace Stress Survey*. [Online]. Available: https://www.perkbox.com/uk/resources/library/interactive-the-2018-uk-workplace-stress-survey

[3] The Physiological Society. (2017). *Stress in Modern Britain Making Sense of Stress*. [Online]. Available: https://static.physoc.org/app/uploads/2020/02/20131612/Stress-in-modern-Britain.pdf

[4] D. Eisenberg, J. Hunt, and N. Speer, "Mental health in American colleges and universities: variation across student subgroups and across campuses," *J. Nervous Mental Disease*, vol. 201, no. 1, pp. 60–67, 2013.

[5] B. Sheaves *et al.*, "Insomnia, nightmares, and chronotype as markers of risk for severe mental illness: Results from a student population," *Sleep*, vol. 39, no. 1, pp. 173–181, Jan. 2016.

[6] C. Blanco *et al.*, "Mental health of college students and their non–college-attending peers: Results from the national epidemiologic study on alcohol and related conditions," *Arch. Gen. Psychiatry*, vol. 65, no. 12, pp. 1429–1437, 2008.

[7] S. Clement *et al.*, "What is the impact of mental health-related stigma on help-seeking? A systematic review of quantitative and qualitative studies," *Psychol. Med.*, vol. 45, no. 1, pp. 11–27, Jan. 2015.

[8] K. Woodward *et al.*, "Beyond mobile Apps: A survey of technologies for mental well-being," *IEEE Trans. Affect. Comput.*, early access, Aug. 7, 2020, doi: 10.1109/TAFFC.2020.3015018.

[9] C. Mohiyeddini and S. Semple, "Displacement behaviour regulates the experience of stress in men," *Stress*, vol. 16, no. 2, pp. 163–171, Mar. 2013.

[10] B. F. Hudson, J. Ogden, and M. S. Whiteley, "Randomized controlled trial to compare the effect of simple distraction interventions on pain and anxiety experienced during conscious surgery," *Eur. J. Pain*, vol. 19, no. 10, pp. 1447–1455, Nov. 2015.

[11] J. Farley, E. F. Risko, and A. Kingstone, "Everyday attention and lecture retention: The effects of time, fidgeting, and mind wandering," *Frontiers Psychol.*, vol. 4, p. 619, Sep. 2013.

[12] V. Carr, "Patients' techniques for coping with schizophrenia: An exploratory study," *Brit. J. Med. Psychol.*, vol. 61, no. 4, pp. 339–352, Dec. 1988.

[13] J. Joormann, M. Siemer, and I. H. Gotlib, "Mood regulation in depression: Differential effects of distraction and recall of happy memories on sad mood," *J. Abnormal Psychol.*, vol. 116, no. 3, p. 484, 2007.

[14] C. Fuentes, I. Rodríguez, and V. Herskovic, "EmoBall: A study on a tangible interface to self-report emotional information considering digital competences," in *Ambient Intelligence for Health. AmIHEALTH* (Lecture Notes in Computer Science), vol. 9456, J. Bravo, R. Hervás, and V. Villarreal, Eds. 2015.

[15] M. Balaam, G. Fitzpatrick, J. Good, and R. Luckin, "Exploring affective technologies for the classroom with the subtle stone," in *Proc. 28th Int. Conf. Hum. Factors Comput. Syst. (CHI)*, 2010, p. 1623.

[16] F. Guribye and T. Gjøsæter, "Tangible interaction in the dentist office," in *Proc. 12th Int. Conf. Tangible, Embedded, Embodied Interact.*, Mar. 2018, pp. 123–130.

[17] E. M. G. Younis, E. Kanjo, and A. Chamberlain, "Designing and evaluating mobile self-reporting techniques: Crowdsourcing for citizen science," *Pers. Ubiquitous Comput.*, vol. 23, pp. 329–338, Mar. 2019.

[18] F. Sarzotti, "Self-monitoring of emotions and mood using a tangible approach," *Computers*, vol. 7, no. 1, p. 7, Jan. 2018.

[19] N. Sharma and T. Gedeon, "Objective measures, sensors and computational techniques for stress recognition and classification: A survey," *Comput. Methods Programs Biomed.*, vol. 108, no. 3, pp. 1287–1301, Dec. 2012.

[20] C. Setz *et al.*, "Discriminating stress from cognitive load using a wearable EDA device" *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 2, pp. 410–417, Mar. 2010.

[21] N. Alajmi, E. Kanjo, N. El Mawass, and A. Chamberlain, "Shopmobia: An emotion-based shop rating system," in *Proc. Humaine Assoc. Conf. Affect. Comput. Intell. Interact.*, Sep. 2013, pp. 745–750.

[22] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 2, pp. 156–166, Jun. 2005.

[23] X. Xing, Z. Li, T. Xu, L. Shu, B. Hu, and X. Xu, "SAE+LSTM: A new framework for emotion recognition from multi-channel EEG," *Frontiers Neurorobotics*, vol. 13, p. 37, Jun. 2019.

[24] T. Umematsu, A. Sano, S. Taylor, and R. W. Picard, "Improving Students' daily life stress forecasting using LSTM neural networks," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Informat. (BHI)*, May 2019, pp. 1–4.

[25] H. P. Martinez, Y. Bengio, and G. N. Yannakakis, "Learning deep physiological models of affect," *IEEE Comput. Intell. Mag.*, vol. 8, no. 2, pp. 20–33, May 2013.

[26] R. Qiao, C. Qing, T. Zhang, X. Xing, and X. Xu, "A novel deep-learning based framework for multi-subject emotion recognition," in *Proc. 4th Int. Conf. Inf., Cybern. Comput. Social Syst. (ICCSS)*, Jul. 2017, pp. 181–185.

[27] K. Woodward, E. Kanjo, D. Brown, and T. McGinnity, "On-device transfer learning for personalising psychological stress modelling using a convolutional neural network," in *Proc. On-Device Intell. Workshop (MLSys)*, Austin, TX, USA, 2020, pp. 46–56.

[28] E. Kanjo, E. M. G. Younis, and C. S. Ang, "Deep learning analysis of mobile physiological, environmental and location sensor data for emotion detection," *Inf. Fusion*, vol. 49, pp. 46–56, Sep. 2019.

[29] U. R. Acharya, K. P. Joseph, N. Kannathal, C. M. Lim, and J. S. Suri, "Heart rate variability: A review," *Med. Biol. Eng. Comput.*, vol. 44, no. 12, pp. 1031–1051, Dec. 2006.

[30] Y. Maeda, M. Sekine, and T. Tamura, "The advantages of wearable green reflected photoplethysmography," *J. Med. Syst.*, vol. 35, no. 5, pp. 829–834, Oct. 2011.

[31] M. Tanida, M. Katsuyama, and K. Sakatani, "Relation between mental stress-induced prefrontal cortex activity and skin conditions: A near-infrared spectroscopy study," *Brain Res.*, vol. 1184, pp. 210–216, Dec. 2007.

[32] Z. Zhang, Y. Song, L. Cui, X. Liu, and T. Zhu, "Emotion recognition based on customized smart bracelet with built-in accelerometer," *PeerJ*, vol. 4, p. e2258, Jul. 2016.

[33] J. C. Quiroz, M. H. Yong, and E. Geangu, "Emotion-recognition using smart watch accelerometer data: Preliminary findings," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput. 2017 ACM Int. Symp. Wearable Comput.*, 2017, pp. 805–812.

[34] A. F. Olsen and J. Torresen, "Smartphone accelerometer data used for detecting human emotions," in *Proc. 3rd Int. Conf. Syst. Informat. (ICSAI)*, Nov. 2016, pp. 410–415.

[35] R. B. Hossain, M. Sadat, and H. Mahmud, "Recognition of human affection in smartphone perspective based on accelerometer and user's sitting position," in *Proc. 17th Int. Conf. Comput. Inf. Technol. (ICCIT)*, Dec. 2014, pp. 87–91.

[36] K. Woodward, E. Kanjo, A. Oikonomou, and A. Chamberlain, "LabelSens: Enabling real-time sensor data labelling at the point of collection using an artificial intelligence-based approach," *Pers. Ubiquitous Comput.*, vol. 24, no. 5, pp. 709–722, Jun. 2020.

[37] Google. (2019). *reCAPTCHA: Easy on Humans, Hard on Bots*. [Online]. Available: https://www.google.com/recaptcha/intro/v3.html

[38] H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller, "Deep learning for time series classification: A review," *Data Mining Knowl. Discovery*, vol. 33, no. 4, pp. 917–963, Jul. 2019.

[39] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[40] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn. (ICML)*, Lile, France, vol. 1, Jul. 2015, pp. 448–456.

[41] G. E. Hinton, A. Krizhevsky, and S. D. Wang, "Transforming autoencoders," in *Artificial Neural Networks and Machine Learning—ICANN* (Lecture Notes in Computer Science), vol. 6791, T. Honkela, W. Duch, M. Girolami, and S. Kaski, Eds. 2011.

[42] K. Suri and R. Gupta, "Continuous sign language recognition from wearable IMUs using deep capsule networks and game theory," *Comput. Electr. Eng.*, vol. 78, pp. 493–503, Sep. 2019.

[43] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 3856–3866.

[44] Z. Wang, W. Yan, and T. Oates, "Time series classification from scratch with deep neural networks: A strong baseline," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 1578–1585.

[45] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[46] P. Tanisaro and G. Heidemann, "Time series classification using time warping invariant echo state networks," in *Proc. 15th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2016, pp. 831–836.

[47] A. E. Hoerl and R. W. Kennard, "Ridge regression: Applications to nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 69–82, Feb. 1970.

[48] J. Serrà, S. Pascual, and A. Karatzoglou, "Towards a universal neural network encoder for time series," in *Proc. Int. Conf. Catalan Assoc. Artif. Intell. (CCIA), Frontiers Artif. Intell. Appl.*, vol. 308, Oct. 2018, pp. 120–129.

[49] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1026–1034.

[50] D. Bahdanau, K. H. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, 2015.

[51] H. I. Fawaz *et al.*, "Inceptiontime: Finding alexnet for time series classification," *Data Mining Knowl. Discovery*, vol. 34, pp. 1936–1962, Sep. 2019.

[52] Y. Zheng, Q. Liu, E. Chen, Y. Ge, and J. L. Zhao, "Time series classification using multi-channels deep convolutional neural networks," in *Web-Age Information Management. WAIM* (Lecture Notes in Computer Science), vol. 8485, F. Li, G. Li, S. Hwang, B. Yao, and Z. Zhang, Eds. 2014.

Kieran Woodward received the B.Sc. (Hons.) degree in information and communications technology from Nottingham Trent University (NTU) in 2016 and the M.Sc. degree in computing systems in 2017. He is currently pursuing the Ph.D. degree with NTU researching the use of tangible user interfaces and on-device machine learning to infer mental wellbeing in real-time.

Eiman Kanjo is currently an Associate Professor of Mobile Sensing Pervasive Computing with Nottingham Trent University. She is a Technologist, a Developer and an Active Researcher in the area of mobile sensing, smart cities, spatial analysis, and data analytics, who worked previously at the University of Cambridge, Mixed Reality Laboratory, University of Nottingham, and the International Centre for Computer Games and Virtual Entertainment, Dundee. She has authored some of the earliest articles in the area of mobile sensing and currently carries out work in the area of digital photonotyping smart cities, mental health and the Internet of Things for behaviour change in collaboration with many industrial partners and end-user organizations.