Stress Detection via Sensor Translation

Sirat Samyoun
Department of Computer Science
University of Virginia
Charlottesville, USA
ss8hf@virginia.edu

Abu Sayeed Mondol
Department of Computer Science
University of Virginia
Charlottesville, USA
mm5gg@virginia.edu

John A. Stankovic
Department of Computer Science
University of Virginia
Charlottesville, USA
stankovic@virginia.edu

***Abstract*—Stress increases the risk of several mental and physical health problems like anxiety, hypertension, and cardiovascular diseases. Better guidance and interventions towards mitigating the impact of stress can be provided if stress can be monitored continuously. The recent proliferation of wearable devices and their capability in measuring several physiological signals related to stress have created the opportunity to measure stress continuously in the wild. Wearable devices used to measure physiological signals are mostly placed on the wrist and the chest. Though currently chest sensors, with/without wrist sensors, provide better results in detecting stress than using wrist sensors only, chest devices are not as convenient and prevalent as wrist devices, particularly in the free-living context. In this paper, we present a solution to detect stress using wrist sensors that emulate the gold standard chest sensors. Data from wrist sensors are translated into the data from chest sensors, and the translated data is used for stress detection without requiring the users to wear any device on the chest. We evaluated our solution using a public dataset, and results show that our solution detects stress with accuracy comparable to the gold standard chest devices which are impractical for daily use.**

***Index Terms*—Stress, Wrist, Chest, Sensors, Physiology**

I. Introduction

Stress is a major reason for several mental and physical health problems including anxiety, cardiovascular diseases, hypertension, and stroke. According to the British Health and Safety Executive, 37% of all work-related ill health cases are caused by stress [1]. It is possible to provide better guidance and interventions toward mitigating the impact if stress can be monitored continuously. However, stress detection is challenging, particularly in the free-living context. Facial expression is a common approach to detect stress, but it is not feasible to capture facial expressions continuously in the wild. In addition to privacy issues, such an approach might not work well for people who are good at hiding their emotions. On the other hand, different physiological signals are affected by stress and other affect states, and so these signals provide an alternative avenue to facial expression for stress detection. The physiological signals can be captured unobtrusively using wearable sensors on the chest or the wrist, and it is not possible to disguise such signals. This paper focuses on detecting stress using wearable sensors.

Stress detection using physiological signals is complex. One reason is because not all the modalities are good indicators of stress in all situations. Also, a non-stressful situation can also change the physiological parameters. So, it is difficult to set a standard accuracy for detecting stress. For example, in the WESAD dataset, authors detect stress with a maximum 80.34% accuracy using the gold standard chest devices.

Chest devices are widely used for stress detection in medical settings. However, wearing devices on the chest is not as convenient as a wrist device, particularly in daily life. As a result, any solution that requires wearing devices on the chest is likely to be less acceptable to the users. On the other hand, in the past wrist-worn devices proved to be far less accurate than the chest devices [2]. However, the recent proliferation of consumer grade wearable devices and their capability have opened the opportunity to monitor stress and affect states continuously in daily life and free-living contexts. But most recently, most of the state-of-the-art solutions use either chest-worn or wrist-worn devices for stress detection; none of these focus on achieving the performance as good as the chest sensors using only wrist sensors.

In this work, we emulate the chest sensors data using the data from the wrist sensors for stress detection. As a result, our solution allows the users to avoid wearing devices on the chest, but provides improved accuracy by using the emulated data. For the emulation, it is necessary to effectively learn a mapping or translation model between the physiological signals available from the wrist and the chest sensors. There are several challenges in developing and validating such a model. First, there is no standard translation model for raw physiological signals. Though several translation models have been used for audio signals translation [3], image translation [4] [5], and text translation [6] tasks, no research has been done to show how these models would perform on physiological signals. Second, unlike other translation models, where both the source and target domains are both either audio signals, texts or images, translation between wrist and chest modalities involves a variety of physiological signals such as Electrodermal activity (EDA), Blood Volume Pulse (BVP), Electrocardiogram (ECG), and Electromyography (EMG) that are often sampled at different frequencies. Third, a translation model needs to be evaluated in the application domain. For example, keyword spotting is used in the audio translation models, while image segmentation or object detection is used in the image translation models. There is no such solution available for stress detection. Fourth, stress detection is challenging because there exist many confounding variables. For example, physiological arousal that should be indicative of stress can be easily confounded by different non-stressful situations. So, it is challenging to develop a translation model that could properly learn the mapping and work well for stress detection.

In this paper, we address each of the above challenges and present a solution for stress detection using wearable sensors by incorporating sensor translation between the chest and wrist modalities. Our work emulates signals for the chest-worn sensors using signals from wrist-worn sensors. However, our goal is not to ensure the goodness of generated signal or itś properties. Rather we are interested in the features of the signal those are vital for stress detection, and those which translate well. Our solution reduces the burden of wearing chest devices while it improves the accuracy of stress detection. The main contributions of our work are summarized below.

1) We present a novel solution for stress detection using wrist-worn sensors. Our solution improves accuracy by emulating the data for chest-worn sensors without imposing additional burden on users to wear any chest device.
2) We develop the first models for translation between physiological signals obtained from the wrist and chest modalities in the wearable stress detection domain. We also use a feature translation model by identifying the situation where the features can not be effectively generated from signal translation.
3) We conduct extensive experiments on a publicly available dataset using different translation models and demonstrate that our solution using only wrist sensors detects stress with accuracy comparable to the golden standard based on chest sensors for all the models.

II. Background and Problem Formulation

A. *Physiological Signals for Stress Detection:*

Several physiological signals can be captured using wearable sensors, particularly using wrist-worn and chest-worn sensors. As background, some of the important physiological signals for stress detection are discussed below.

1) **EDA:** An EDA sensor captures the electrodermal activity of the skin arising from emotional stimulation, or physical or cognitive activities. An EDA signal has two main components, a slowly varying tonic component that represents the current skin conductance level (SCL), and a rapidly changing phasic component, that represents skin conductance response (SCR). Both of these components are vitals for stress detection [7].
2) **ECG and BVP:** ECG refers to the Electrocardiogram and BVP refers to the Blood Volume Pulse that are usually measured from the chest and the wrist, respectively. The BVP is also known as Photoplethysmogram (PPG). It measures the blood volume changes of the heart by measuring light transmission or reflection. Both ECG and BVP signals are used to calculate the heart-rate and the heart rate variability (HRV). The HRV is an important indicator of stress [8].

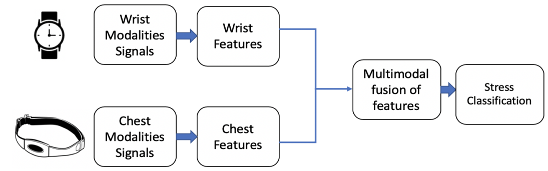

Fig. 1. Stress detection pipeline using both wrist and chest sensors

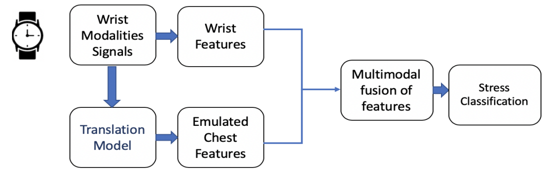

Fig. 2. Stress detection using wrist sensors with translated features for chest

3) **Respiration:** Different studies have reported the response of the human respiratory system to mental stress [9]. Moreover, respiration rate increases during stressful situations compared to a resting condition [9].
4) **EMG:** Electromyography (EMG) refers to the electrical activity produced by the skeletal muscles. Previous research have shown that stress can lead to increased EMG in specific muscles of the body [10].
5) **Temperature:** Body temperature can be measured using both the wrist and the chest sensors. It has been shown that the human skin temperature uniformly increases in response to stress. [11].

B. *Problem Formulation:*

Chest sensors in combination with wrist sensors provide better results for stress detection compared to using wrist sensors only [12]. A general pipeline for stress detection using chest and wrist sensors is shown in the figure 1. In this approach, features are extracted from the physiological signals of wrist and chest sensors, and then the features are combined and used for the stress classification task [12].

The focus of this paper is to detect stress using only the wrist modalities, and still get as good performance as that using the chest modalities, and thereby to eliminate the need of using the chest devices in stress detection. To solve this problem, we need to develop a translation model which can emulate the chest features using the data from wrist sensors, as shown in the figure 2. At first, the chest features which are important for stress detection needs to be determined. Next, the translation model has to learn the mapping between the wrist and the chest modalities. So, the key research goals of this paper are:

1) Finding the features from the chest modalities that are good indicators of stress.
2) Developing a translation model to emulate the chest features from Wrist Sensors.
3) Developing classification models for stress detection using the translated features as well as original features from the wrist sensors.

III. METHODOLOGY

In this section, we discuss the methodology followed for achieving the research goals identified in the previous section.

*1) **Feature Selection from Chest Modalities**:* A number of feature selection techniques can be used for selecting the best features from the chest modalities which are good indicators of stress. In this paper, we use the Recursive Feature Elimination (RFE, Guyon et al. (2002) [13]), a greedy algorithm that fits a model with the training data and removes the weakest features one by one until reaching the desired number of features.

*2) **Translation model for Generating Features**:* We use two kinds of translation to generate chest features from the wrist sensors, which are briefly discussed below.

- **Signal translation:** It is the process of emulating a signal (target signal) from other signal (source signal) using a translation model (Figure 3). Previous studies showed that neural networks can be effectively used for translating data sequences between the correlated sensors [14].
- **Feature Translation:** In feature translation, a feature is generated from a set of selected features (Figure 3). The source features are chosen on the basis of their relative importance to the target feature from the training data.

*3) **Stress classification using translated features**:* For feature extraction, the sequences of the sensor signals are segmented into windows using a sliding window method. The chosen important features available from these signals are computed and extracted from these windows. In case of feature translation, the features are already generated via translation. The features from the signal translation and the feature translation models are then combined to form the emulated feature set. The emulated set of features is then used for the stress classification.

Primer on GAN: We use Generative Adversarial Networks (GAN) for signal translation between two homogeneous signals, because GANs are very effective for generating plausible transformation of any source data sequences via generative modeling, and is widely used for image translation [5]. A GAN consists of two neural networks competing with opposite goals, the generator and the discriminator. In each training iteration, the generator tries fool the discriminator by producing realistic fake samples from a random distribution, while the discriminator tries to correctly discriminate between the real and the fake samples. In this work, we have implemented a GAN similar to Pix2Pix [4], a type of a Conditional GAN, where the generation of the target data sequence is conditional on source data sequence.

Primer on bi-directional RNN: We also use bi-directional Recurrent Neural Networks (RNNs) for signal translation, as RNNs are very effective for modeling sequential data. This is because the RNNs have hidden states which allows the network to remember the contextual information about a sequence. Moreover, a bi-directional RNN works better for translation than the uni-directional RNN [6], because it can capture the contextual information from both sides.

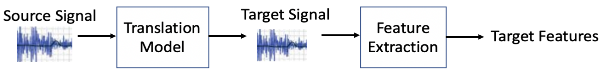

Fig. 3. Signal translation

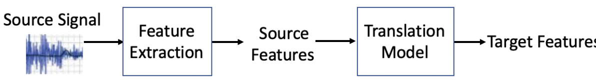

Fig. 4. Feature translation

IV. SOLUTION AND IMPLEMENTATION OF THE TRANSLATION MODELS

In this section, we focus on the 3 main solution steps for translation to emulate the top chest features.

1) A GAN-based translation model is used for translation between two homogeneous signals. For example, to generate a top chest EDA feature, we use the wrist EDA signal to chest EDA signal translation.
2) A RNN-based translation model is used for translation between two heterogenous signals, if the signals have some correlated underlying attributes. For example, to generate a top chest ECG feature, we use the wrist BVP signal to chest ECG signal translation. Because an ECG signal is unavailable on the wrist, however a wrist BVP signal shares some attributes similar to an ECG signal.
3) An MLP-based feature translation model is used otherwise. For example, to generate a top chest EMG feature, we use feature translation from the features available from the wrist sensors. Because neither an EMG signal, nor any other correlated signal is available on the wrist.

The implementation of the three translation models are discussed below.

A. GAN-based Signal Translation Model

*1) **Network Architecture**:* The generator (G) in the translation model takes a source data sequence as input, and outputs a target sequence. Inspired by the work of [15], we implemented G using a U-Net architecture. G consists a combination of layers of encoder and decoder. The skip connections were used to connect layers in the encoder with corresponding layers in the decoder, thus resulting a U-shape, which is shown in the figure 5. The encoder layers encodes the input sequence signal down to a bottleneck representation through a number of convolution layers. Batch normalization was also used both during the training and the testing to ensure that the statistical values were computed for each batch. A LeakyRelu activation function was used on the encoder part. The structure of the decoder part is opposite to that of the encoder. It takes the bottleneck representation as input and it uses the transpose of convolution layers to upscale to the required output sequence of the signal. A kernel size of 4 and stride size 2 were used in the convolution and it's transpose layers. Dropout layers were used in the encoder-decoder structure to introduce randomness during training.

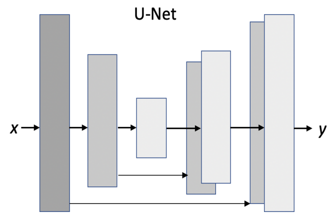

Fig. 5. U-Net Generator Architecture

The discriminator (D) of a translation model takes samples from both the source and target data sequence, and it aims to figure out whether the target sample (y) is a plausible transformation of the source sample (x). D was implemented by a convolutional PatchGAN classifier. It is designed in a way that each output of the model maps to a patch of the input data sequence. A combination of convolution layers along with Batch Normalization were used to implement D.

*2) **Training details**:* The generator G of the translation model is trained via the adversarial loss (L_C). L_C makes the generator generate samples of the target signal in such a way that those cannot be distinguished from samples of the real signal by D. G is also updated via L_1 loss. An L_1 loss is measured between the translated data generated by itself and the expected output data (y). This additional loss encourages G to move the distribution of the translated samples as close as possible to real distribution of the samples in the target domain. The loss function (G^*) of the translation model is shown below, where λ is the regularization parameter:

$$G^* = \arg\min_G \max_D L_C(G, D) + \lambda L_1(G)$$
$$L_C = E_{x,y}[logD(x, y)] + E_{x,z}[log(1 - D(x, G(x, z))]$$
$$L_1 = E_{x,y,z}[||y - G(x, z)||]$$

The discriminator (D) was trained on both real chest samples and fake chest samples for the corresponding real wrist signal samples. Next the generator (G) was trained on batch with real wrist and chest samples, so that it is able produce the fake (translated) chest samples. While training G, weights of D were kept frozen. D verified the output of G and tried to predict the goodness of the translated fake chest samples. In this way through each training iteration, G kept minimizing itś loss. Typically a GAN model does not converge, instead, an equilibrium is found between the G and D. Therefore, the training was stopped when more iterations did not significantly change the loss of the G or the D. The model was optimized using binary cross entropy optimizer, and a weighting is used in a way that the updates to the model have half (0.5) the usual effect. This weighting was chosen to decelerate the changes to D, relative to G.

*B. **RNN-based Signal Translation Model:***

We used a bi-directional LSTM network, a kind of RNN for translation between the heterogenous sensing modalities. A sequence of sensor signals are provided as input to the bi-directional LSTM that generates output for each of the input samples. The output of the LSTM network is the emulated signals for the target chest sensor. For implementing the architecture, a bi-LSTM layer with 32 units was used, followed by a dense layer with a ReLu activation function. Moreover, the return sequences option was enabled, meaning the RNN layer returned the full sequence as the output instead of returning only the last hidden state output.

*C. **MLP-based Feature Translation Model:***

We argue that any random signal should not be emulated using some other signals by neural network models. Some correlation must be present between the source and target signals, particularly over the sequence. It is possible to build a neural network model that memorizes all the training data, but such a model would perform very poorly on test data. So, signal translation would not perform well where the correlation between the signals is not sufficient enough to achieve acceptable accuracy. For the top features which could not be generated via signal translation, we used feature translation. The intuition behind this is a feature on a chest physiological signal might have a strong or weak correlation with one or a number of features from the wrist physiological signal. So, using feature selection on the training data, we try to select a feature or set of features from the wrist data which are relatively important for predicting that chest feature. We use the feature(s) to train a multilayer perceptron (MLP). A multilayer perceptron (MLP) is a simple feedforward artificial neural network with several fully connected layers that takes a number of features as input and generates a set of outputs. In this paper, an MLP was created with two dense layers, followed by ReLU activation function, and was trained with the features selected from the wrist modalities.

V. EXPERIMENTS

We present a thorough evaluation to highlight the performance improvement achieved by our solution. The key research questions that drive our evaluations are:

1) How accurate are the emulated features compared to the original chest features?
2) Does the translation model improve the stress and affect detection performance using the wrist sensors?
3) Can the translation model eliminate the need of placing sensors on the chest (which is not practical in the wild) for detecting stress?

*A. **Data Description***

We have evaluated our solution using WESAD [12], a publicly available dataset for stress and affect detection. We use WESAD because it is the only publicly available dataset that contains synchronized data from both the wrist sensors and the chest sensors required for emulating chest features on the wrist. The dataset has three kinds of labels for stress classification for a total of 15 subjects, namely, the baseline (neutral) condition, the amusement condition, and the stressed

TABLE I
WESAD DATASET STATS

Features	Modalities	Sampling rate	No. of samples
Wrist Modalities	EDA	4 Hz	347472
	BVP	64 Hz	5559552
	Acceleration	32 Hz	2779776
	Temperature	4 Hz	347472
Chest Modalities	EDA	700 Hz	60807600
	ECG	700 Hz	60807600
	EMG	700 Hz	60807600
	Respiration	700 Hz	60807600
	Acceleration	700 Hz	60807600
	Temperature	700 Hz	60807600

condition. Both the baseline and amusement conditions represent the non-stressful situations. Table I presents the signals in the WESAD dataset along with the sampling frequencies.

B. Data Pre-Processing

First the synchronized raw data files were loaded for all subjects. To map between the wrist and the chest samples, downsampling, a commonly used signal processing technique was used. The EDA signals were passed through a low pass filter corresponding to the EDA sampling rate, and then were passed through cvxEDA [16], an algorithm which uses different convex optimization methods for the analysis of EDA signals. The phasic and the tonic, the two important EDA components for stress detection, were extracted using this algorithm. An additional component called SMNA (Sudomotor Nerve Activity) was extracted from the signal. This component is another vital element for stress detection. Some pre-processing was also needed to remove the noisy artifacts from the signals. The raw accelerometer data was passed through a FIR (Finite Impulse Response) filter to filter high-frequency vibration noises. The net acceleration was measured on the accelerations along the X, Y and Z axes.

C. Signal and Feature Translation

Being homogeneous signals, a translation between the wrist EDA and chest EDA signal was done. Among the heterogeneous signals, the wrist BVP to chest ECG, and the wrist BVP to chest Respiration signal translations were carried out. Because both the wrist BVP and chest ECG signals contains some stress indicative heart-rate related parameters [8]. Moreover, previous studies used BVP/PPG signals to estimate the respiratory rate [17], which is a strong indicator of stress. Feature translation was used for the top features coming from the signals not involved in the signal translation.

To make the model general and not be overfit to any training samples, we performed leave-one-subject-out (LOSO) cross validation in both translation models. Previous studies have shown that LOSO increases the generalization capabilities of the models against the real-world unseen data [18]. Using the LOSO scheme, a model was trained on data from all subjects except one, data from that one subject was used as test data. The parameters were optimized for each of the models. Following the translation, the generated signals and the original chest and wrist signals were pre-processed and features were extracted for classification.

D. Feature Extraction

We windowed the pre-processed data with a window size of 60 seconds, and a sliding length of 5 seconds per window. This window size was picked according to the paper in which the dataset was originally published [12]. Once the windows were generated, we extracted different features from the windows. The mean, standard deviation, minimum and maximum values of a window were found for the modalities. The features from the modalities were then fused together to form a combined set of features. A total of 36 and 44 features were derived from the original wrist and chest signals, respectively. Top 15 features from the chest sensors, as listed in Table IV, were derived through translation. Twelve of the features are extracted from the translated signals, and the remaining features are derived through feature translation.

E. Translation Results

Our translation models effectively learnt the mapping between the wrist and the chest modalities. However, any loss in the translation process will result in error in the translated signals and the features extracted from the translated signals when compared to the original signals or the features extracted from the original signals. We demonstrate both these errors by computing Mean Average Error (MAE), a commonly used performance metrics. The MAE between the original signal ($X^{originial}$) and the translated signal ($X^{translated}$) is:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|X_i^{original} - X_i^{translated}| \tag{1}$$

Table II shows the performance of the two translation models across different signals. The feature level translation errors, that is, the MAEs of the translated features compared to the original features, are shown in Table III. For space limitation, we opt to show the average MAE for all the subjects. The features are ranked in the order of their importance to detecting stress. The MAEs for the emulated top 15 features are shown in Table IV.

Results show that the GAN translation works better than the RNN when the signals are homogeneous, that is, the data are of similar kind. For example, the wrist EDA to chest EDA translation using RNN incurs a MAE of 0.944, whereas that using GAN incurs a MAE of 0.842, leading to a 10% reduction of error rate. This is because GANs are very effective for producing fake samples that resembles the actual data, as it uses generative modeling. However, RNN model provides slight improvements for heterogeneous signals. So, for stress classification, we opt to use GAN translated signals for homogeneous translations, and RNN for the heterogeneous ones. It must be also mentioned that the claim of this paper is not to ensure the goodness of generated signal or itś properties. Rather we are interested in the features of the translated signal those are vital for stress detection, and those which were translated well, so that we get the best translated chest features

TABLE II
MAE OF THE TRANSLATED SIGNALS

Source Signal	Target Signal	MAE - GAN	MAE - RNN
Wrist EDA	Chest EDA	0.8420	0.9445
Wrist BVP	Chest ECG	0.5130	0.5026
Wrist BVP	Chest Resp	0.7271	0.7100

TABLE III
MAE OF THE TRANSLATED FEATURES

Source Feature	Target Feature	MAE - MLP
Wrist- All Features	EMG_Raw_mean	0.0007
Wrist- All Features	EMG_Raw_std	0.0046
Wrist- All Features	Temp_Raw_std	0.0909
Wrist- All Features	Temp_Raw_mean	0.0794

for stress detection. Therefore, the features translated with high error, such as, EDA_Tonic_max, and EDA_smna_mean were not considered for stress classification.

F. Classification Results

We evaluated the performance of both the original features and the translated features for stress detection. Classification for the stress detection was carried out with different feature sets as mentioned below.

1) **Wrist Only:** Features extracted from the original wrist signals are used.
2) **Chest Only:** Features extracted from the original chest signals are used.
3) **Wrist and Chest:** Features extracted from both wrist and chest signals are used.
4) **Wrist using Translation:** Features extracted using the translation model on the wrist are used (our solution).

We used some of the most widely used machine learning algorithms for classification. The WESAD dataset is labeled with three main classes, namely, baseline, stress, and amusement. Similar to the paper that presents WESAD [12], we carried out three-class (stress vs baseline vs amusement) and two-class classification (stress vs non-stress). We detected stress using Random Forest (RF), Extra Trees (EXT), Decision Trees (DT), Linear discriminant analysis (LDA), Logistic Regression (LR) and Multi-Layer Perceptron (MLP). We evaluated the performance of our solution using Accuracy and F1-score, the most commonly used performance metrics for classification.

TABLE IV
TOP RANKED CHEST FEATURES AND MAE ON TRANSLATION

Rank	Features	Modality	MAE	Translation Type
1	ECG_Raw_std	ECG	0.0526	Signal-Signal
2	EDA_Phasic_std	EDA	2.4750	Signal-Signal
3	EDA_Tonic_max	EDA	35.0474	Signal-Signal
4	EDA_Tonic_std	EDA	2.2260	Signal-Signal
5	EDA_smna_min	EDA	0.0001	Signal-Signal
6	EMG_Raw_mean	EMG	0.0005	Feature-Feature
7	EMG_Raw_std	EMG	0.0046	Feature-Feature
8	RESP_Raw_max	RESP	0.0152	Signal-Signal
9	Temp_Raw_std	TEMP	0.0909	Feature-Feature
10	EDA_smna_mean	EDA	29.3640	Signal-Signal
11	Temp_Raw_mean	TEMP	0.0794	Feature-Feature
12	RESP_Raw_std	TEMP	1.1252	Signal-Signal
13	EDA_Tonic_max	EDA	0.5381	Signal-Signal
14	EDA_Phasic_mean	EDA	1.7358	Signal-Signal
15	ECG_Raw_max	ECG	1.1372	Signal-Signal

The results of the 3-class classification and the 2-class classification are presented in the tables V and VI, respectively. We make the following observations from the results:

1) In general, the chest sensors work better than the wrist sensors for stress detection by about 4-5% for accuracy, and the combination of both chest and wrist sensors works better than using sensors on the wrist or chest only. Chest sensors are less affected by motion artifacts than wrist sensors, and also most physiological signals are better captured at the chest than the wrist due to the internal structure (e.g., location of heart) of the human body. These results justify the use of chest sensors for stress detection in terms of classification performance if user convenience is not considered.
2) Classification performance is improved when original data from wrist are combined with translated data instead of using data from the wrist only. The improvement in the performance is achieved without sacrificing user convenience as users need to wear a device on the wrist only, not on the chest. It should be noted that data from both wrist and chest are used for training purposes. Once the translation models are built, our solution only needs data from the wrist.
3) For 3-class classification, our solution improves stress detection accuracy and F1-score by an average of 5-6% compared to the wrist only modality. The improvement is smaller for the 2-class classification than 3-class classification, because the latter is relatively less complex, and benefits more from the extra information available in the translated data.
4) Using both modalities in the original form, that is original data from both the wrist and the chest sensors, gives the best performance for all the models. However, our solution provides accuracy comparable to that in most cases. For example, the best stress detection accuracy using chest only sensors in 80.5%, that with the same classifier using both the wrist and chest sensors is 81.6%, while our solution achieves 81.4%. It validates that our solution ensures user convenience with very little compromise on accuracy.
5) The ensemble based classifiers, the Random Forest and the Extra Trees gives the best performance for both the classification tasks. Because the ensemble methods reduce variance by combining results from a number of underlying weak classifiers, and thus are less prone to overfitting. Among the other classifiers, Linear Discriminatory Analysis performs better in both the classification tasks. We also observe that the Decision Tree shows different results than those of the other classifiers. This is because it is highly prone to overfitting to the training data, and performs poorly on the testing data.

TABLE V
CLASSIFICATION ACCURACY AND F1-SCORE, 3 CLASS (BASELILNE/STRESS/NON-STRESS)

Classifier	**Accuracy**				**F1-Score**			
	Wrist Only	**Chest Only**	**Wrist + Chest**	**Wrist using Translation**	**Wrist Only**	**Chest Only**	**Wrist + Chest**	**Wrist using Translation**
Random Forest	**75.6**	**80.5**	81.6	**81.4**	**71.5**	**67.1**	**75.8**	**74.5**
Extra Tree	74.1	79.1	**83.4**	80.6	66.8	61.2	72.8	65.2
Decision Tree	72.8	76.4	71.5	75.3	66.7	63.8	63.3	64.5
Linear Discriminatory Analysis	70.5	73.0	75.1	74.1	64.8	58.5	69.6	67.5
Logistic Regression	68.6	71.9	72.2	72.3	64.3	56.2	65.1	65.0
Multi-Layer Perceptron	70.2	70.9	72.9	71.6	63.2	62.7	67.8	65.8

TABLE VI
CLASSIFICATION ACCURACY AND F1-SCORE, 2 CLASS (STRESS/NON-STRESS)

Classifier	**Accuracy**				**F1-Score**			
	Wrist Only	**Chest Only**	**Wrist + Chest**	**Wrist using Translation**	**Wrist Only**	**Chest Only**	**Wrist + Chest**	**Wrist using Translation**
Random Forest	**89.9**	90.0	**94.7**	**92.1**	**87.6**	87.5	**93.4**	**89.7**
Extra Tree	88.6	**91.1**	93.7	90.6	86.4	**90.2**	92.2	91.4
Decision Tree	85.1	88.1	82.9	85.8	82.4	85.9	80.2	85.8
Linear Discriminatory Analysis	89.2	90.2	90.7	90.6	86.8	87.8	88.4	88.1
Logistic Regression	84.9	87.0	85.9	86.2	82.0	84.3	82.5	83.8
Multi-layer Perceptron	83.0	89.1	91.6	88.3	79.6	86.3	89.6	86.5

Overall, the translation results show that our translation models using only the wrist sensors are able to generate the best features of the chest which are indicators of stress. The best stress detection performance on the wrist device (75.6% accuracy, obtained from the Random Forest classifier) was beaten by a 5%-6% improvement by our solution by using the same classifier (81.4% accuracy). By doing do, our solution also achieves a stress detection performance on the wrist essentially the same as using chest sensors too (81.6% accuracy, using the same classifier). For the other classifiers too, the stress detection performance using our wrist-based solution are very close to that using the chest sensors.

VI. RELATED WORKS

Stress Detection using Wearables: Several papers [19] [12] detected stress using different physiological signals recorded from the chest modalities, such as, ECG, respiration, acceleration, EMG, etc. Authors in [20] detected stress with high accuracy from chest ECG and respiration signals using a DNN-based framework. On the other hand, authors in [21] [7] used different physiological parameters obtained from the wrist modalities such as, BVP, EDA, skin temperature, and acceleration to detect stress. Authors in [12] combined both the wrist and chest modalities to achieve higher stress detection accuracy than the independent modalities. Authors in [22] developed a new wearable capable of measuring a person's cortisol levels from their sweat to detect stress. However, none of the works focused on bridging the gap between the stress detection performance on the chest and wrist wearables.

Other Works on Stress Detection: Authors in [23] summarizes stress and other affective states detected from different sensor modalities. Authors in [24] used wristbands to measure stress among dementia patients. Authors in [25] studied the effect of heart-rate related features in stress detection.

Data Translation Models: Pix2Pix [4] is a general purpose image-to-image translation model, while the CycleGAN [5] model is used to perform translation between domains with unpaired images. On the other hand, the Recurrent Neural Networks (RNNs) are widely used for text translation models [6]. However, for the acceptability of these translation models, the translated data from these models were evaluated on the application domain. For example, spotting the keywords in translated audio signals [3], detecting and counting the objects in the translated images [4]. However, there is no standard translation model for translating data sequences across different kinds of physiological signals, and no research has been done to show how these state-of-the-art models could perform in stress detection domain.

Works on the Physiological Signals: RespNet [17] is a deep learning based framework that performs the task of extracting a respiration signal from a PPG signal. Several other works [18] [26] focuses on extracting or estimating the physiological parameters such as, heart-rate, respiratory rate from the signals, such as PPG or ECG. None of the works provide any solution for mapping physiological signals or features from one domain to another domain. Moreover, none of these works are targeted for stress detection.

VII. DISCUSSION AND LIMITATIONS

We argue that translation should not be done between any two random physiological signals. Only the signals that are similar or have correlated underlying attributes should be selected for translation. Because most of the signals for stress detection from the chest and wrist modalities are either the same, or they share similar physiological parameters, our translation models were able to effectively translate the wrist signals into the chest signals.

Our solution with translation works better than using the original wrist data, because we chose the best features of the chest data that represent stress, generated those features from the related wrist data, and then picked only the features with low translation error for classifying stress. Moreover, our solution uses both GAN, which is very popular for generative modeling, and RNN which is very effective for modeling sequential data.

An alternative solution without translation is to train a classifier for detecting stress directly from the wrist data. But to match the performance of the chest sensors, the classifier would need large amount of training data to automatically learn the best features from the wrist. However, this data is not available among the existing datasets. On the other hand, our solution generates the best stress indicator features for the chest modalities via translation.

The wrist and chest sensors are subject to real-world noise and motion artifacts. Although some preprocessing were done to handle these motion artifacts, this paper does not focus on handling the noise and motion artifacts of the data. Further research is needed to improve the robustness of the translation model against these problems.

This paper focused on how the translation models improve the stress detection performance using the wrist sensors only. Based on application requirements and resource availability, the translation models can be deployed in different platforms like wrist devices, smartphones, and servers in the cloud. Future works include implementation and resource requirement analysis for the translation tasks on the different platforms both for online (real-time) and offline stress detection.

VIII. Conclusion

Smart wrist devices are becoming more prevalent for wellness and healthcare. In this work, we leveraged the stress detection performance on the wrist by emulating the chest features from the wrist signals using a novel collection of 3 translation schemes. Our overall translation solution is the first of its kind that tackles the problems of achieving stress detection performance using the wrist sensors close to the level as using the chest sensors.

Acknowledgement

This work was supported, in part, by the NSF grant CNS-1646470.

References

[1] Health and safety at work, "Hse on work-related ill health," 2016. [Online]. Available: http://www.hse.gov.uk/statistics/overall/hssh1516.pdf

[2] "Wristband vs chest sensors," https://www.acc.org/about-acc/press-releases/2017/03/08/14/02/wrist-worn-heart-rate-monitors-less-accurate-than-standard-chest-strap.

[3] A. Mathur, A. Isopoussu, F. Kawsar, N. Berthouze, and N. D. Lane, "Mic2mic: using cycle-consistent generative adversarial networks to overcome microphone variability in speech systems," in *Proceedings of the 18th International Conference on Information Processing in Sensor Networks*. ACM, 2019, pp. 169–180.

[4] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

[5] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[6] K. Cho, B. Van Merriënboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," *arXiv preprint arXiv:1409.1259*, 2014.

[7] C. Setz, B. Arnrich, J. Schumm, R. La Marca, G. Tröster, and U. Ehlert, "Discriminating stress from cognitive load using a wearable eda device," *IEEE Transactions on information technology in biomedicine*, vol. 14, no. 2, pp. 410–417, 2009.

[8] H.-G. Kim, E.-J. Cheon, D.-S. Bai, Y. H. Lee, and B.-H. Koo, "Stress and heart rate variability: A meta-analysis and review of the literature," *Psychiatry investigation*, vol. 15, no. 3, p. 235, 2018.

[9] D. Widjaja, M. Orini, E. Vlemincx, and S. Van Huffel, "Cardiorespiratory dynamic response to mental stress: a multivariate time-frequency analysis," *Computational and mathematical methods in medicine*, vol. 2013, 2013.

[10] R. Luijcks, H. J. Hermens, L. Bodar, C. J. Vossen, J. van Os, and R. Lousberg, "Experimentally induced stress validated by emg activity," *PloS one*, vol. 9, no. 4, p. e95215, 2014.

[11] C. H. Vinkers, R. Penning, J. Hellhammer, J. C. Verster, J. H. Klaessens, B. Olivier, and C. J. Kalkman, "The effect of stress on core and peripheral body temperature in humans," *Stress*, vol. 16, no. 5, pp. 520–530, 2013.

[12] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing wesad, a multimodal dataset for wearable stress and affect detection," in *Proceedings of the 2018 on International Conference on Multimodal Interaction*. ACM, 2018, pp. 400–408.

[13] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine learning*, vol. 46, no. 1-3, pp. 389–422, 2002.

[14] M. A. S. Mondol and J. A. Stankovic, "Neural sensor translation," in *Proceedings of the International Conference on Internet of Things Design and Implementation*. ACM, 2019, pp. 279–280.

[15] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[16] A. Greco, G. Valenza, A. Lanata, E. P. Scilingo, and L. Citi, "cvxeda: A convex optimization approach to electrodermal activity processing," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 4, pp. 797–804, 2015.

[17] V. Ravichandran, B. Murugesan, V. Balakarthikeyan, S. M. Shankaranarayana, K. Ram, J. Joseph, M. Sivaprakasam *et al.*, "Respnet: A deep learning model for extraction of respiration from photoplethysmogram," *arXiv preprint arXiv:1902.04236*, 2019.

[18] A. Reiss, I. Indlekofer, P. Schmidt, and K. Van Laerhoven, "Deep ppg: Large-scale heart rate estimation with convolutional neural networks," *Sensors*, vol. 19, no. 14, p. 3079, 2019.

[19] F.-T. Sun, C. Kuo, H.-T. Cheng, S. Buthpitiya, P. Collins, and M. Griss, "Activity-aware mental stress detection using physiological sensors," in *International conference on Mobile computing, applications, and services*. Springer, 2010, pp. 282–301.

[20] W. Seo, N. Kim, S. Kim, C. Lee, and S.-M. Park, "Deep ecg-respiration network (deeper net) for recognizing mental stress," *Sensors*, vol. 19, no. 13, p. 3021, 2019.

[21] A. Sano and R. W. Picard, "Stress recognition using wearable sensors and mobile phones," in *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*. IEEE, 2013, pp. 671–676.

[22] O. Parlak, S. T. Keene, A. Marais, V. F. Curto, and A. Salleo, "Molecularly selective nanoporous membrane-based wearable organic electrochemical device for noninvasive cortisol sensing," *Science advances*, vol. 4, no. 7, p. eaar2904, 2018.

[23] P. Schmidt, A. Reiss, R. Duerichen, and K. Van Laerhoven, "Wearable affect and stress recognition: A review," *arXiv preprint arXiv:1811.08854*, 2018.

[24] B. Kikhia, T. Stavropoulos, S. Andreadis, N. Karvonen, I. Kompatsiaris, S. Sävenstedt, M. Pijl, and C. Melander, "Utilizing a wristband sensor to measure the stress level for people with dementia," *Sensors*, vol. 16, no. 12, p. 1989, 2016.

[25] P. Melillo, M. Bracale, and L. Pecchia, "Nonlinear heart rate variability features for real-life stress detection. case study: students under stress due to university examination," *Biomedical engineering online*, vol. 10, no. 1, p. 96, 2011.

[26] M. Pirhonen, M. Peltokangas, and A. Vehkaoja, "Acquiring respiration rate from photoplethysmographic signal by recursive bayesian tracking of intrinsic modes in time-frequency spectra," *Sensors*, vol. 18, p. 1693, 05 2018.