Received January 12, 2021, accepted January 23, 2021, date of publication February 5, 2021, date of current version February 16, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3057578

Intelligent Stress Monitoring Assistant for First Responders

KENNETH LAI, (Student Member, IEEE), SVETLANA N. YANUSHKEVICH, (Senior Member, IEEE), AND VLAD P. SHMERKO

Biometric Technologies Laboratory, Department of Electrical and Computer Engineering, University of Calgary, Calgary, AB T2N 1N4, Canada

Corresponding author: Kenneth Lai (kelai@ucalgary.ca)

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) through grant "Biometric-enabled Identity management and Risk Assessment for Smart Cities," and in part by the Department of National Defence through the Innovation for Defence Excellence and Security (IDEaS) Program, Canada.

ABSTRACT This paper describes a prototype of an intelligent Stress Monitoring Assistant (SMA), – the next generation of stress detectors. The SMA is intended for the first responders and professionals coping with exposure to extreme physical and psychological stressors, e.g. firefighters, combat military personnel, explosive ordnance disposal operatives, law enforcement officers, emergency medical technicians, and paramedics. Stress impacts human behavior and decision-making, which can be propagated between the team members. The SMA is an integral part of the Decision Support System, it is a component of the decision support perception-action cycle. We model this cycle as a cognitive dynamic system. The intelligent part of the SMA is designed using *a*) a residual-temporal convolution network for learning data from sensors and detection of stress features, and *b*) a reasoning mechanism based on a causal network for fusion at various levels. The SMA prototype has been tested using a multi-factor physiological dataset WEarable Stress and Affect Detection (WESAD). In both modes, the stress *recognition* and stress *detection*, the SMA achieves an accuracy of 86% and 98% for the WESAD dataset, respectively. This performance is superior to the known results in satisfying the requirements of reliable decision support.

INDEX TERMS Soft biometrics, stress monitoring, intelligent decision support, causal inference, machine reasoning, affect recognition.

I. INTRODUCTION

Humans have three primary systems vital for survival: vision, cognitive processing, and motor skill (V-C-M). Under stress, all three can break down [1]. The goal of this paper is to create a tool to support professionals who respond to emergencies, e.g. firefighters, combat military personnel, explosive ordnance disposal operatives, law enforcement officers, emergency medical technicians, and paramedics. These groups are engaged in uniquely demanding and dangerous work involving regular exposure to both physical and psychological stressors.

Recent research roadmaps, e.g. [2] and reviews, e.g. [3], highlight a great demand for wearable stress detectors for these professionals. In response to this demand, we develop and prototype an intelligent Stress Management Assistant (SMA). The SMA is a dynamic system that is cognitive of emotional stress and is an integral part of the Decision Support System (DSS). The DSS learns the stress level in causal relations to human behavior and provides the required support. The DSS is one of the critical components of a user model to build an intelligent adaptive system.

Taxonomically, stress detection is a part of the affect recognition problem [3], [4]. Affect addresses various emotional states, e.g. sadness, happiness, unhappiness, etc. This problem is formalized using a pattern recognition theory and techniques with priorities of facial expressions. Stress is a particular and most troubled affective state that can jeopardize the subject's survival.

Prediction of heat stress, in particular, is a priority for firefighters and Explosive Ordnance Disposal operatives. In their missions, wearing heavy protective clothing is mandatory as there are dangers of heat stress in thermally harsh environments which can occur when the cooling required to maintain

The associate editor coordinating the review of this manuscript and approving it for publication was Donato Impedovo.

a steady thermal state is greater than the cooling capability of the environment [5].

Stress compromises the cognitive performance of firefighters [6]. During emergency scenarios, firefighters have to quickly respond to numerous stimuli under pressure, while the decisions made during these emergency situations tend to be based on information that is ambiguous, incomplete or unusual, further complicating the decision-making process [7]. Besides the physiological reactions, stress may also cause cognitive responses such as attentional processes which are an integral part of an adaption process for intelligent systems.

The focus of our study is stress detection for the purpose of assisting the first responders. This problem is significantly different from detecting a relatively low daily life stress. The specific occupational stresses are characterized as follows:

Requirement I: A Wearable Sensor Network (WSN) is preferable for stress detection and monitoring [8].

Requirement II: The standards for decision support for stressful occupations are being developed, and may become mandatory [2], [9]. In contrast, the stress detectors for daily lives (e.g. "Stress tracker" color bar on the Samsung Health app) are built on different principles; they rather approximately "estimate" the level of stress based on the heart rate measured by placing a finger on the sensor (if available on the phone) or via a Smartwatch.

Limitation: Face is a useful source of data for stress detection [10], [11]; however, it may be occluded by the protective equipment. In contrast, in public-centric applications, e.g. on iPhones, the face (mouth, pupil, skin temperature distribution) is the main source of information.

The remainder of this paper is organized as follows. Abbreviations and problem formulation are provided in Sections II and III. The contributions of this paper are explained in Section IV. Section V provides a survey of the most important related works, while the architecture of the proposed assistant is described in Section VI. The experimental settings and results are described in Sections VII and VIII. Conclusions and recommendations are given in Section IX.

II. ABBREVIATIONS

ACC – Accelerometer
BN – Bayesian Network
BVP – Blood Volume Pulse
CPT – Conditional Probability Table
DSS – Decision Support System
DL – Decision-level
ECG – Echocardiogram
EDA – Electrodermal activity
EEG – Electroencephalogram
EMG – Electromyogram
FL – Feature-level
ML – Machine Learning
Res-TCN – Residual-Temporal Convolution Network
RESP – Respiration
SL – Sensor-level
SMA – Stress Monitoring Assistant
TCN – Temporal Convolution Network
TEMP – Temperature
V-C-M – Vision, Cognition, and Motor skills
WSN – Wearable Sensor Network

III. PROBLEM FORMULATION AND APPROACH

The National Institute of Standards and Technology (NIST) has defined a research roadmap on technology and real-time analytics in fire fighting [2]. It identifies the key *conceptual* gaps, *technological* gaps, and *development* gaps.

A roadmap for first responders' coping is provided by the U.S. Homeland Security [9]. Challenges related to the stress in combat are defined in [12], [13].

Our study identifies the following gaps in the existing design and deployment of the SMA for first responders:

– Most WSNs provide the sensor data logging [14] and monitoring [8],
– Only a few attempts to implement the elements of decision support such as the fusion of multi-sensor data for generating a situational picture [15].

Thus, there is a *dissonance* in the development of the technology for stress detectors, and the decision support for stress coping; in our view, the stress detectors shall be able to perform the risk assessment that shall be instrumental for recovery from stress state, trust to the action of stressed combatant, and mitigation of stress propagation in a team.

This paper addresses the above problems and makes a further step in developing the SMA. We propose a system-level design approach that combines sensing arrays, stress detection, and decision support on a common platform. This platform account for the primary survival faculties, i.e. vision, cognition, and motor skills. This contribution distinguishes our work from the known results that focus on stand-alone solutions [14], [15].

A. STRESS MONITORING ASSISTANT

Fig. 1 explains a novel approach to the development of the wearable SMA and DSS. The schematic sketch shows the three information pathways that correspond to the primary human survival systems: Vision (V), Cognition (C), and Motor skill (M) [1].

The wearable sensing landscape is *decomposed* with respect to the V-C-M content; the closest to the V-C-M are the biometric projections [16]. The DSS function is formulated in the terms of the V-C-M: under stress, a first responder needs assistance in the vision domain (e.g. he/she cannot see), motion domain (e.g. he/she cannot move), and cognitive domain (e.g. he/she cannot make an adequate decision). This stress state should be recognized by the team members in order to prevent stress propagation and, consequently, the failure of the mission.

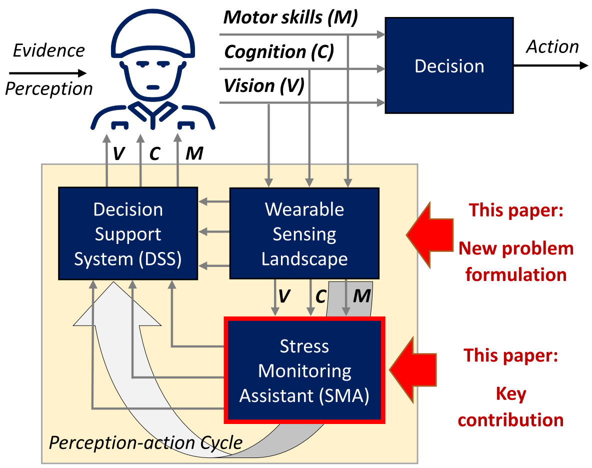

FIGURE 1. Summarizing the challenges for the first responders, and the solutions that this research offers.

B. STRESS MONITORING MODEL

Design requirements to stress monitoring can be met using the principles of a cognitive dynamic system. This system is based on *Haykin's* model [17]:

Perception-action cycle implies that there are perceptor and actuator;

Memory for the purpose to learn from the environment and store knowledge;

Attention – ability to prioritize the allocation of available resources; and

Intelligence– a function that enables the control and decision-making mechanism to help identify intelligent choices.

In this model (Fig. 1), the problem of stress monitoring should be understood as follows. Stressful events cause dynamic changes in the human body, human behavior, and cognition, in other words, the *V-C-M survival dimensions*. The subject's *stress perceptors* perceive the environment by processing the input stimuli called *stressors* that threaten the homeostasis (the ability to maintain a relatively stable internal state of the subject). A stress response is the body's reaction to regain homeostasis using the *perception-action cycle*. The V-C-M changes can be observed by monitoring the body response signals, caused involuntarily by the autonomic nervous system. These changes can be measured by the SMA. An intelligent SMA shall not only process these measurements, but also assess the risk of stress, and then pass the assessment to the DSS that generates a recommendation to the subject to support his/her decisions regarding the stressful events.

C. APPROACH

The goal of this study is to develop and prototype an intelligent SMA for the first responders. The SMA is seen as a wearable system that satisfies a number of specific requirements, in particular:

Ad Hoc: The SMA must be integrated into first responder networks, e.g. [2], [9].

Compatibility: The SMA must be compatible with the first responder decision support tools; the risk of stress should be taken into account in the decision-making and decision support process.

Intelligence: The SMA must be an integral part of the cognitive processes of a first responder; the SMA shall be able to learn the subject's personal physiological and behavioral data, and then conduct the respective reasoning.

In our approach, the SMA is a component of decision support built on the principle of stress perception-action cycle known as a General Adaptation Syndrome (GAS) [18]. This model provides the necessary conditions for the incorporation of computational intelligence in the form of Machine Learning (ML) and machine reasoning. Note that in this paper, we used a controlled dataset for experimentation. Should the data change, for example, real-life scenario data becomes available, the learning and reasoning mechanism will make it possible to adjust to the new data.

IV. CONTRIBUTION

While our work complies with the general stress detector development doctrine [12], it addresses a particular niche, the first responders' support. There are necessary and sufficient conditions to satisfy the specific requirements and limitations of this application.

The key contribution of this paper to satisfy the necessary conditions are as follows:

1) The SMA for first responders' support is based on the cognitive dynamic models in the V-C-M survival dimensions.
2) The SMA design is approached using this new concept.

Our work also partially covers sufficient conditions as follows:

1) The SMA design uses the advances in machine learning, in particular, the Residual-Temporal Convolution Network (Res-TCN).
2) The SMA intelligence is realized by probabilistic reasoning; it also performs a fusion of the stress-related features. The reasoning mechanism is implemented on a causal network via a forward inference for fusion, and via a backward one for diagnosis.

V. RELATED WORK AND RESEARCH GAPS

The key paradigm of stress as a psychological phenomenon is well understood but far from practical use. This is because stress is unique for each individual. Contemporary technologies for stress detection can cope, however, with multiple factors and their causal relationships that trigger the stress. These factors are determined by the physiological, emotional, behavior, and social profiles of an individual, as explained in a brief review below.

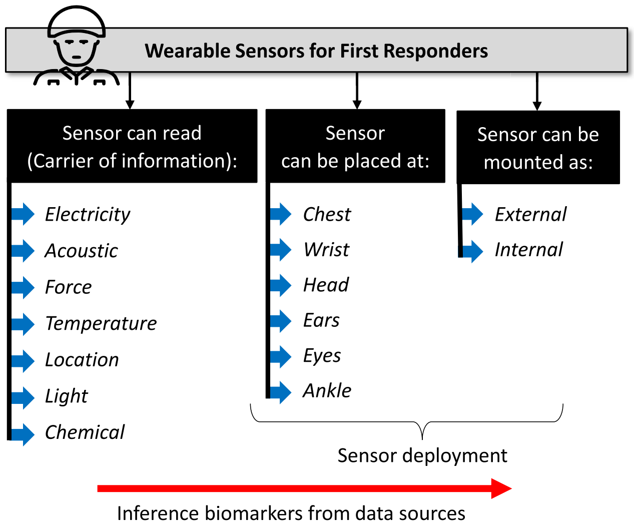

FIGURE 2. **Wearable sensor taxonomy for the first responders with respect to the carrier of information and sensor deployment.**

A. MULTI-FACTOR STRESS SOURCES

According to the classical model [19], "stress results from the appraisal that environmental or internal demands exceed or tax the coping resources of the individual''. Medical and psychiatry studies consider stress in relation to multiple physiological factors such as age and gender, as well as social factors, e.g. life conditions, foot, training, and education.

B. TAXONOMICAL PROJECTIONS OF WEARABLE SENSOR NETWORKS

The sensing landscape should satisfy the fundamental requirements of the V-C-M survival [1], that is, predict or timely detect the breaking down of at least one of the survival systems. The V-C-M are also considered to be biometrics; a comprehensive taxonomy of wearable sensors for biometric purposes is provided in [16]. Based on this taxonomy, we formulated the SMA projections for the first responders as follows (Fig. 2):

– a carrier of information that sensors read,
– a body part appropriate for acquiring the corresponding source of information,
– how the sensor can be mounted.

This sensing landscape is dependent on the efficiency of the inferred technology. For example, sweat, saliva, and tears are bio-fluids that contain multiple physiologically relevant chemical constituents. They can be detected using electrochemical sensors or/and electronic lenses, and require the appropriate inferring and reasoning approaches.

This taxonomical view allows for a new formulation of stress detection and recognition problem in terms of the SMA cognition, reliability of detection, as well as the credibility of the information. In particular, the following questions are of importance:

– How reliable are the physiological factors provided by the internal sensors, for example, electrochemical tattoo [20] for stress detection?
– Where they should be placed to achieve maximum credibility of information?
– Does electronic contact lenses, e.g. [21], provide essential information for stress detection?

C. INTELLIGENT GARMENT

Intelligent garment (or smart clothing) is a form of a WSN. The key components of the intelligent garment include: sensors that monitor physiological and environmental signals; actuators that control the clothing movements and display some visual effects; a microcontroller that processes the measured data; a communication system; and an energy supplying unit. In [22], an intelligent garment was adopted for firefighters protection with several extensions toward intelligent functions such as detecting the relationship between the wearer's health state (fatigue, stress) and the physiological measures. In [23], stress detection for military combatants was conducted using the wearable sensorized garments and gloves. The most informative data sources happen to be the ECG, heart rate, and respiration rate. A similar work on stress monitoring using a wearable patch with integrated sensors such as skin temperature, skin conductance, and pulse-wave was reported in [24].

D. SENSOR PLACEMENT AND COMMUNICATION

The sensor can be placed in a WSN in four ways:

– *In-body*; sensors are placed inside the body;
– *On-body*; sensors are placed on the skin surface;
– *Off-body*; sensors are placed between a few centimeters and up to a few meters away from the body; and
– *Mixed*.

Note that choosing the communication means for the SMA significantly impacts the overall design process. For communication, short-range radio frequency solutions based on widespread commercial standards such as Bluetooth are typically used. These solutions have limitations such as energy consumption, interferences, and vulnerability to attacks. An alternative is intra-body communication, a technique that uses the human body as a transmission channel for electrical signals to interconnect devices in the WSN. These devices can be both on-body and implanted (in-body), and communicate with each other and with a central device through the low-power and low-data-rate body channels [25], [26]. The ultrasonic band is one of the possible approaches to intra-body communication. For example, in [27], a wearable sensor network uses ultrasonic communication for minimal message exchange, asynchronous updates, and distributed (without centralized control) resource allocation algorithms.

E. REASONING MECHANISM

One of the first works on machine reasoning for stress detection and propagation in combat soldiers is a master thesis by

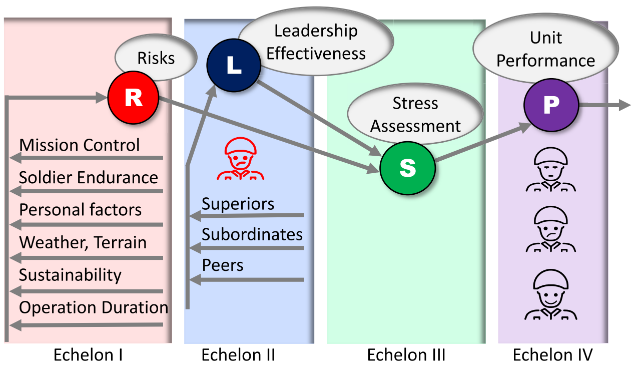

FIGURE 3. **Multi-echelon causal model of a combat unit. Stress *S* is a moderating factor between the leadership effectiveness *L* and the performance of their unit *P* biased by the risk and trust factors such as mission control, soldier endurance and selection, weather, terrain, sustainability, and planning.**

US Army Captain Therrien [28]. It studies the stress states of combat soldiers and their performance under the impact of various factors using the reasoning mechanism of Bayesian networks. The stress states were considered in *causal relations* with the leadership effectiveness, risks of human factors (e.g. exertion, endurance), conditions (e.g. training, sleeping, fitness), environment (e.g. terrain, weather), etc. Fig. 3 provides more detail of this work that considered stress assessment of combat soldier as a monitoring and learning process.

F. COMBAT AND OPERATIONAL STRESS CONTROL

In military operations [12], two kinds of stress are distinguished: *combat stress* (the experience of lethal force or its aftermath) and operational stress (the experience or consequences of military operations other than combat). Combat and Operational Stress Control (COSC) model suggests four stress levels [12]: the *Green zone* (or Ready), the *Yellow zone* (or Reacting), the *Orange zone* (or Injured), and the *Red zone* (diagnosable mental disorders). Automated support is needed at each of the stress levels. Our study addresses rather the Yellow and Orange stress zones.

Example 1: Accordingly to the COSC, each member of a military unit should become comfortable and confident in the use of the following procedures needed to calm down, set feelings aside, and restore mental focus: 1) deep breathing, and 2) grounding (refocus your thinking). Both procedures should be included in the DSS protocol and partially monitored, e.g. deep breathing shall be monitored using the respiration sensor.

Example 2: The following behavior warning signs (Orange zone) can be detected automatically. For a period of time, the individual:

– seems unable to stop laughing, crying, or screaming;
– could not move parts of his or her body; freezes or appears to be moving very slowly;
– stutters, repeats words or phrases, or has a shaky or squeaky voice.
– exhibits physical reactions, such as sweating, shaking, or heart pounding.

Example 3: Combat and Operational Stress First Aid consists of several actions which should be supported by automated tools:

– Check (identify the need for stress first aid),
– Coordinate (inform others who need to know),
– Cover (get individual to safety as soon as possible),
– Calm (reduce heart rate and emotional intensity),
– Connect (promote peer support, prevent stressed individuals from isolating themselves),
– Competence (restore mental and physical capabilities),
– Confidence (restore self-confidence, self-esteem, and hope).

G. RESEARCH GAPS

Stress detection, as a part of the first responder decision-making support, must be coherent with the overall system requirements and limitations. These includes WSN, communication paradigms, and intelligence of decision support. While the first two components have been addressed extensively, the current gap is in the area of intelligent decision support. In our study, we address this gap by applying deep ML to detect individual stress factors in the perception-action cycle. The recent advances in ML made it possible to address the psychological factors and projection of the problem, i.e. the facts indicate that stress and its physiological manifestation are unique for each individual. The SMA learning of the unique stress features for a given individual (first responder), i.e. SMA "personalization" enabled by deep ML, is the core of our design approach.

Accepting that the SMA is a personalized tool, our work aims at bridging the following research gaps: how to apply ML and machine reasoning with respect to the stress states, and how to solve the related problems of fusion and porting into the overall DSS protocols.

VI. SMA ARCHITECTURE

The key architectural components of the proposed SMA prototype are:

1) A *WSN* for extracting physiological data of the subject.
2) *Res-TCN*, – a Residual-Temporal Convolution Network for learning the subject's data from wearable sensors. Each sensory data is used to train a separate Res-TCN;

$$\underbrace{\texttt{Stress phenomenon}}_{\textit{Psychology}} \Rightarrow \left\{\begin{array}{c}\texttt{Machine}\\ \texttt{Learning and}\\ \texttt{Reasoning}\end{array}\right\} \Leftarrow \underbrace{\texttt{Stress factors}}_{\textit{Physiology}}$$

TABLE 1. Sources of physiological data for stress recognition and detection.

Channel	Input	Output	Setup	Intelligent Processing
ACC → TCN → Prediction	Input is the estimated motion of the body measured in terms of acceleration in the Cartesian coordinate system.	Output is the pattern of body movements used for inferring emotional state.	Accelerometer sensors attached to the chest and non-dominant hand. Signal sampled at 700 Hz for the chest and 32 Hz for the hand.	Provide self-feedback based on the level of movement as well as the current state of emotions.
ECG → TCN → Prediction	Input is the electrical impulses measuring the pace of heart rate, the rhythm of the heart-beat, and the strength of each pulse.	Output is the heart rate pattern, used for finding the level of emotional response.	A standard three-point ECG sensor is attached to the chest. Signal sampled at 700 Hz.	Detection of abnormality of heart rate as well as provide an analysis of the heart rate level and current emotional state. Contacts emergency unit if heart-rate is too low or too high.
EDA → TCN → Prediction	Input is the measure of skin conductance in terms of tonic (conductance level) and phasic (conductance response).	Output the skin conductance patterns, used for finding the level of emotional response.	EDA sensors are attached to the rectus abdominis and non-dominant hand. Signal sampled at 700 Hz for the chest and 4 Hz for the hand.	Provide self-support based on the current emotional state.
EMG → TCN → Prediction	Input is electrical impulses measuring the muscle activity and nerve response.	Output is the change in muscle response to be used for predicting the corresponding emotion.	An EMG sensor is attached to the chest. Signal sampled at 700 Hz.	Detection of abnormal skeletal-muscle behavior and provide warning and/or contact emergency unit based on the current emotional state
RESP → TCN → Prediction	Input is the measure of the movement of the chest and abdominal wall.	Output is the estimated breathing rate that is further analyzed to estimate the emotion via patterns of breathing.	A respiration inductive plethysmograph sensor is attached to a subject's chest. Signal sampled at 700 Hz.	Detection of abnormal breathing rate and provide real-time breathing rate feedback. Contact the emergency unit if no chest movement (not breathing) is detected.
TEMP → TCN → Prediction	Input is the temperature measured at selected regions of interest.	Output is the change in temperature resulting from the transition between emotions.	Temperature sensors attached to the sternum and non-dominant hand. Signal sampled at 700 Hz for chest and 4 Hz for hand.	Detection of abnormal temperature and provide a warning if a spike in temperature is detected.
BVP → TCN → Prediction	Input is the measure of heart rate based on the volume of blood that passes through a region given each heartbeat.	Output is heart rate used to analyze how emotional state affects the heart rate.	A BVP sensor attached to the non-dominant hand Signal sampled at 64 Hz.	Detection of abnormal heart rate value and provide support based on the current emotional state.

3) A *Fusion Profiler* that chooses a preferable fusion procedure;
4) A *reasoning mechanism* for making decisions regarding the stress based on the Res-TCN reports. A causal network is used for this purpose.

These components are incorporated in the perception-action cycle defined in Fig. 1.

A. WEARABLE SENSOR NETWORK

The goal of a WSN is the inference from a body physiological data and transmit them for processing. According to taxonomy Fig. 2, the WSN must satisfy a wide spectrum of requirements such as sensor type, placement, and distribution on the body. Table 1 provides a detailed specification of the WSN data fed into the Res-TCN input (columns 'Channel' and 'Input').

B. RESIDUAL TCN

In our approach to building the individual-centric SMA, we deploy ML to learn the stress indicators from physiological data of first responders during both the rest and action time, and thus, classify the emotion. We deploy machine reasoning to amalgamate the ML process by providing the fusion of the trained models.

In this paper, the Res-TCN is chosen as the classifier of the data represented by the time series. TCNs has shown optimal performance for time series based data including human action recognition [29] and action segmentation [30]–[32]. In the proposed SMA, we modified the Res-TCN architecture (Fig. 4). Given the physiological factor from a sensor as a time series, the Res-TCN aims at extracting stress features using four residual blocks (Fig. 4(*a*)). Each residual block contains the following attributes:

1) *1D convolution* extracts features from time-series data, in order to learn how the data changes over time.
2) *Causal profiler* limits the convolution process to use only the current and past data. By excluding future data, the resulting prediction is better aligned with the real-life scenarios where the future data is unavailable.

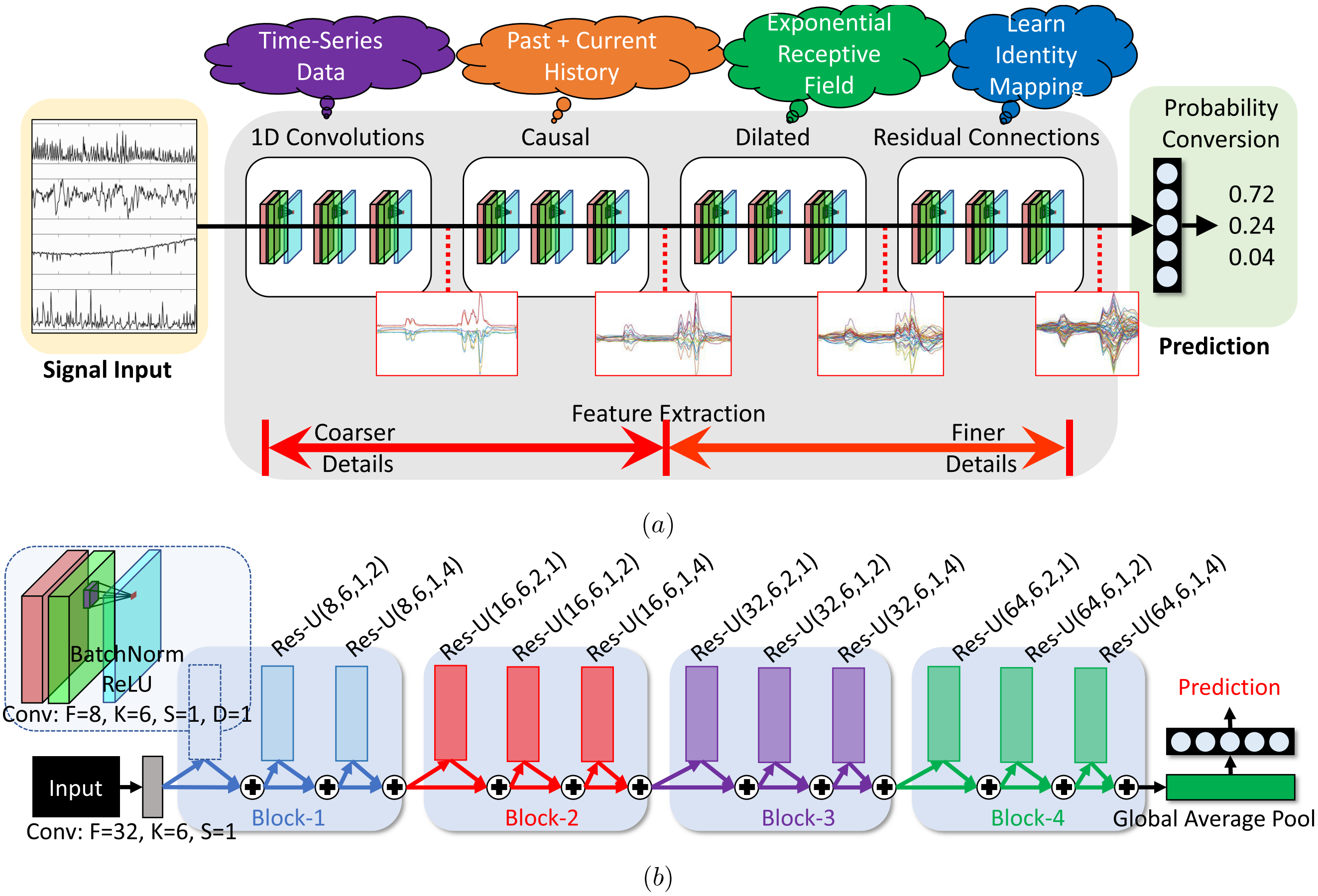

FIGURE 4. The Res-TCN for physiological data processing: (*a*) Architecture and (*b*) Specification.

3) *Dilated convolution* performs the down-sampling instead of pooling layers, allowing for an exponential receptive field. A larger receptive field allows for the detection of longer sequenced patterns.
4) *Residual connection* is a connection between the two layers used to bypass or skip the multiple layers between the two layers. This form of identity mapping allows the gradients to propagate back through the network without going through activation functions, thus prompting faster convergence.

Fig. 4(*b*) provides the following design details and specifications. Each block in the network contains three residual sub-block. Each residual sub-block contains three layers: batch normalization, rectified linear unit, and convolution. The parameters for the convolution layer is defined as Res-U(F, K, S, D), where F is the number of filters, K is the filter size, S is the stride value, and D is the dilation rate:

Block-1 consists of three residual blocks:
Res-U(8, 6, 1, 1), Res-U(8, 6, 1, 2), and Res-U(8, 6, 1, 4). For example, the first residual block is defined as Res-U(7, 6, 1, 1), representing 8 6 × 6 filters with a stride of 1 and a dilation rate of 1. For each proceeding sub-block, the number of filters, filter size, and stride remains the same while the dilation rate is doubled. This results in 8 feature vectors containing coarse details of the signal.

Block-2 consists of three residual sub-blocks:
Res-U(16, 6, 2, 1), Res-U(16, 6, 1, 2), and Res-U(16, 6, 1, 4). Different from Block-1, the number of filters is doubled (from 8 to 16). In addition, the stride for the first sub-block is also doubled. This results in 16 feature vectors containing moderate details of the signal.

Block-3 consists of three residual sub-blocks:
Res-U(32, 6, 2, 1), Res-U(32, 6, 1, 2), and Res-U(32, 6, 1, 4). Similar to block-2, the number of filters is again doubled (from 16 to 32). This results in 32 feature vectors containing finer details of the signal.

Block-4 consists of three residual sub-blocks:
Res-U(64, 6, 2, 1), Res-U(64, 6, 1, 2), and Res-U(64, 6, 1, 4). Similar to block-3, the number of filters is again doubled (from 32 to 64). This results in 64 feature vectors containing yet finer signal details.

Figure 5 illustrates emotion classification using the independent sensory signals from the chest (*a*) and wrist (*b*) sensors. Each sensory signal is passed into its individual Res-TCN, resulting in an emotion prediction.

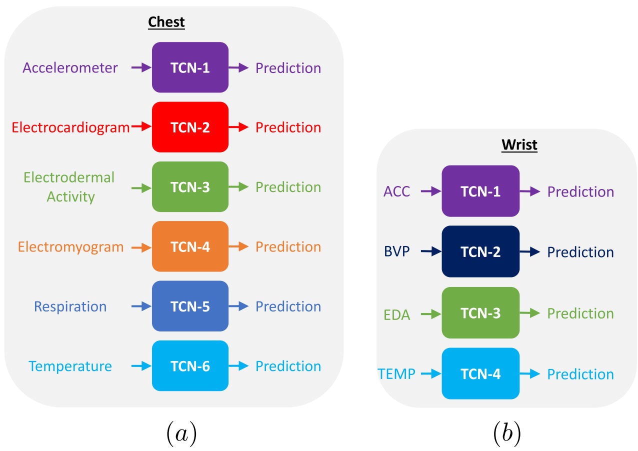

FIGURE 5. Emotion classification using independent: (*a*) chest sensory signals: ACC, ECG, EDA, EMG, RESP, and TEMP. (*b*) wrist sensory signals: ACC, BVP, EDA, and TEMP.

For time-series data, the best hand-crafted features for classification are generally the mean, peaks, and moments at the selected time interval; however, given the Res-TCN, the features extracted is obtained through supervised machine-learning. The feature output of the Res-TCN is 1×64 feature vector. This vector represents the most important aspects of the signal and how it changes through time in a compressed form. Analysis of this feature vector can be done to perform emotion classification.

C. FUSION

In a multi-sensory system, the ultimate objective is usually to consolidate the data gathered from multiple sources [33], [34]. Sensors may provide inaccurate, incomplete, and conflicting data. However, the right approach to data fusion can increase system robustness and fault tolerance, as well as reduce the system's sensitivity to sample-specific, poor quality, or erroneous input. Moreover, when sensors are not deployed in an optimal fashion, unwanted uncertainty occurs on the input of the classifier or detector, thus reducing the classification accuracy. Intelligent classification support, e.g. residual TCNs alleviate these problems.

The following fusion methods are used in the SMA prototyping (Fig. 6):

Sensor Level (SL) fusion that combines physiological data from multiple sensors prior to feature extraction.

Feature Level (FL) fusion that combines the features extracted from each signal.

Decision Level (DL) fusion that combines the decisions of separate residual TCNs.

The left plane (a) on Fig. 6 shows the SL fusion of the wrist sensory data. The signals from the few sensors are concatenated before passing through a TCN. The central plane (b) illustrates the FL fusion for the wrist sensors. The signal from each of the wrist sensors is passed into the respective TCN. Each TCN produces a 1-D logit vector which can be concatenated with other TCN's outputs in order to create a feature vector to be passed on to a multilayer-perceptron. The perceptron performs the final emotion prediction based on the amalgamated feature vector. The right plane (c) of Fig. 6 illustrates the process of DL fusion. Each classifier prediction is fused together to obtain a more confident prediction. The logic function of such fusion can be AND (all predictions must point to the class of interest), OR (at least one decision must point to the class of interest), or other logic.

In this paper, we have chosen a machine reasoning approach to fusion. The model of such fusion is a causal network. The designed causal network is shown in Fig. 7. Note that the designed causal network in this paper is only a baseline. It can be modified and/or improved to include additional information such as demographics, education, experience, etc. This information can be incorporated as parent nodes that can influence the performance of the different signals. The "Valid" node in this network represents the probability of the classifier in predicting a positive (valid) or negative (invalid) emotion. The parent nodes to the "Valid" node represent the different modalities that affect the stress recognition or stress detection performance. The "Match" node determines whether the positive or negative prediction matches the ground truth label accordingly to the"Type" of modality.

Given the causal network and the corresponding Conditional Probability Tables (CPTs), a probabilistic reasoning model called a Bayesian network (BN) is built. Posterior probabilities can be calculated by applying Bayesian inference using the BN, prior probabilities, and the current observation [35]. This is the mechanism we use in this work for ensemble fusion of multiple modalities.

VII. EXPERIMENT SETTINGS AND PROTOTYPING

In this experiment, we used the dataset for WEarable Stress and Affect Detection (WESAD) [3]. Other recent datasets of stress includes the MSP-Podcast [36], Multimodal Dataset of Stressed Emotion (MuSE) [37], and Interactive Emotional Dyadic Motion Capture (IEMOCAP) [38]. WESAD dataset was chosen as it is the only dataset that provides physiological signals. Other datasets provide visual and/or audio data which is not of key interest in this paper. Although the MuSE dataset monitors heart rate, it was ultimately not chosen for evaluation because they do not provide other physiological signals like ECG.

A. DATASET

WESAD dataset contains ECG data from 17 participants [3]. A RespiBAN Professional sensor was used to collect signals at a sampling rate of 700 Hz. Each signal is labeled with four different affective states: neutral, stressed, amused, and meditated. Four different test scenarios were created. During the first 20 minutes, neutral data were collected; the participants were asked to do normal activities such as reading a magazine and sitting/standing at a table. In the amusement scenario, the participants watched 11 funny video clips for a total length of 392 seconds. Next, the participants went through public

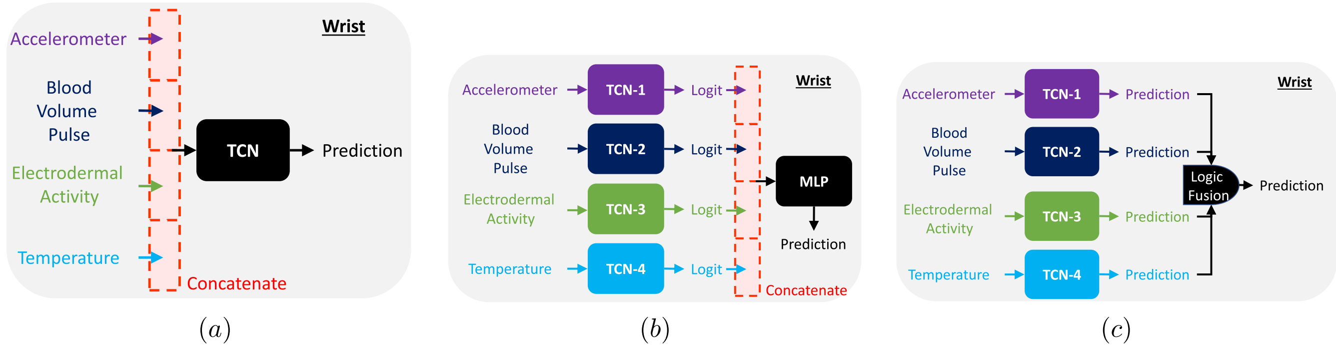

FIGURE 6. The SMA fusion: *(a)* SL fusion on wrist sensors; *(b)* FL fusion on wrist channel; *(c)* DL fusion on wrist channel.

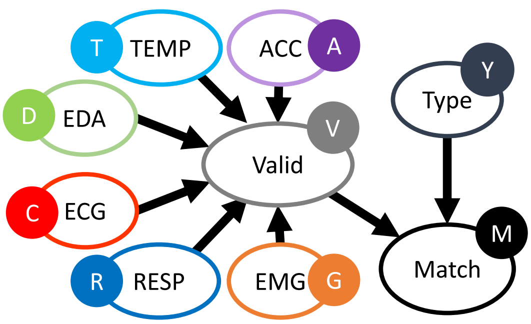

FIGURE 7. Causal network of an ensemble of modalities for stress classification and detection.

speaking and arithmetic tasks for a total of 10 minutes as part of the stress scenario. Finally, they went through a guided meditation session of 7 minutes in duration. Upon completion of each trial, the ground truth labels for the affect states were collected using the Positive and Negative Affect Schedule (PANAS) scheme [39].

B. PRE-PROCESSING AND SEGMENTATION

Segmentation and normalization of data streams acquired from sensors are basic operations of pre-processing. In this paper, segmentation of the sensor signals was done using a modified sliding window technique. The technique involves two sliding windows,

1) a sliding window to compute the average value, and
2) a sliding window to retrieve the time slice of interest.

We used a 0.25-second window for averaging and experiment with various duration for time slice and time increment.

Since each of the signals is collected at different frequencies, we have chosen to normalize the sampling rate using the lowest frequency (4Hz for TEMP and EDA signal of the wrist sensor). Based on the 4Hz normalization, the window for averaging values is 0.25 seconds. The conversion process is defined as follows:

$$s(x) = \frac{1}{f_s} \sum_{i=x}^{x+f_s} \texttt{signal}(i) \tag{1}$$

where f_s is the sampling frequency of the sensor, `signal` is the original signal of the sensor, and $s(x)$ is the desired averaged value for index x of the `signal` values within the following interval: $[x; x + f_s]$.

Fig. 9 illustrates the modified sliding window technique used in this paper. The original signal is first normalized based on a $f_s/4$ second window which averages all values within the window. After normalization, another sliding window of 60 seconds is used to extract the time slice of interest. This 60-second window extracts a time slice every 0.25 seconds to be used as the input to the TCN.

Figure 8 shows the result of the re-sampled signals for a single time slice. Because each data point is 0.25 seconds apart, a time slice represents 240 data points for a 60-second window (60 × 4).

C. IMPLEMENTATION AND TRAINING

We implemented all the models using Keras and Tensorflow. The Res-TCN model to classify emotions is trained using the Adam optimizer for 100 epoch, with a base learning rate $lr = 0.001$. The Adam optimizer default parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = e^{-8}$ were set to train the network for the various validation methods. The training was done on a NVIDIA V100 16 GB.

Similar to the original work of the WESAD dataset [3], we adopted a leave-one-subject-out-cross-validation (LOOSCV) for performance evaluation. For the WESAD dataset, which contains 15 subjects, the LOOSCV performs 15-fold cross-validation. Each fold is separated into a 14:1 split, 14 subjects used for training and the remaining subject for validation. This process is repeated for each subject, and the performance is evaluated based on the arithmetic mean of the 15 folds.

Another form of validation we performed was suggested in [40]. It uses 10 rounds of random sampling. In this validation procedure, 90% of the data is randomly sampled to be used as training and the remaining 10% is used for testing. This process is repeated for 10 rounds with the results being averaged.

D. PERFORMANCE MEASURES

In the WESAD dataset, some classes, e.g. normal emotional state, have a large quantity of data; these are the majority classes. Minority classes contain instances that can be

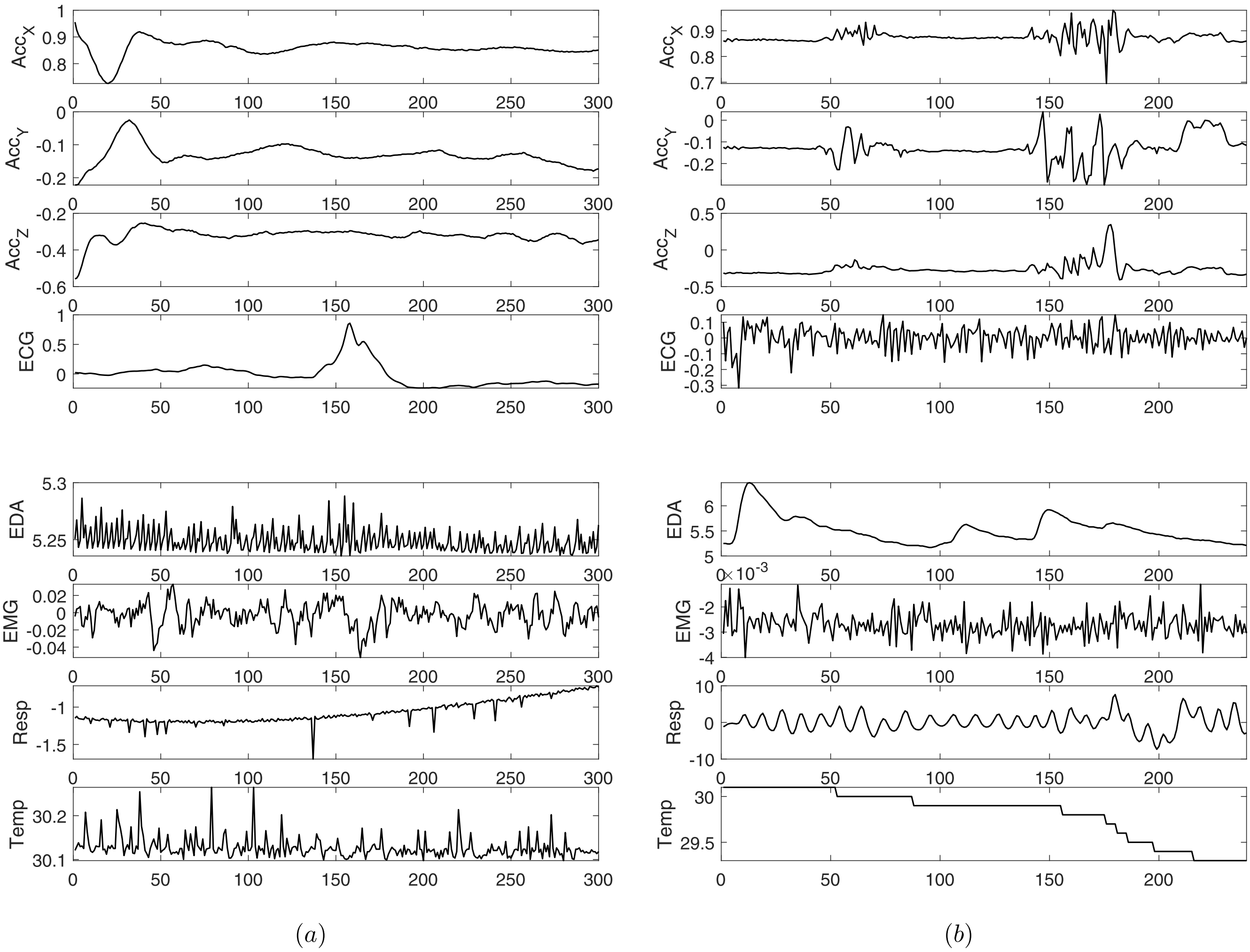

FIGURE 8. (*a*) Example of chest sensor data from ACC, ACG, EDA, EMG, RESP, and TEMP. Data is shown for the first 300 instances sampled at 700hz. (*b*) Example of extracted time slice based on sliding window from ACC, ACG, EDA, EMG, RESP, and TEMP. A 240-frame time slice is shown for the chest signals.

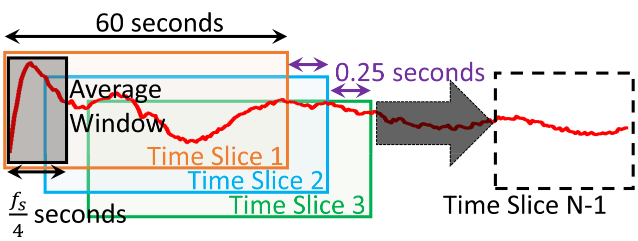

FIGURE 9. Illustration of the sliding window technique for extracting features; it uses a $f_s/4$ window for averaging, 60 seconds for the time slice, and 0.25-second for time increment.

observed less often, e.g. stress states. The rare patterns are difficult to separate from the most frequent ones. These factors normally cause a biased performance of all common classifiers such as support vector machines, decision trees, and neural networks, due to having little quantity of data instances, resulting in an effect that causes minority classes to be ignored in the overall classification process.

The WESAD dataset is an example of an imbalanced set [3]. However, in our application, the minority class is the only class of interest, and the performances of the classifier should be evaluated mostly using this premise.

A major issue in the classification is to determine the most suitable performance metrics to be used. In the case of balanced data, the traditional metric include various derivatives that use the following indicators:

TP – True Positives (correct predictions of emotion),
FN – False Negatives (incorrect predictions of emotion),
TN – True Negatives (correct rejections of emotion), and
FP – False Positives (incorrect predictions of emotion).

These numbers form a 2×2 confusion matrix:

$$\texttt{Confusion matrix} = \begin{pmatrix} TP & FP \\ FN & TN \end{pmatrix}$$

Various measures may be built to summarize its content [41]. The most common are the accuracy, recall, precision, false

TABLE 2. The SMA performance evaluation in the emotion classification mode (Normal vs. Stress vs. Amusement) and stress detection mode (No stress vs. Stress).

Modalities		Emotion Classification Mode Accuracy	F_1-Score	FPR	Stress Detection Mode Accuracy	F_1-Score	FPR
Chest	ACC	71.19 ± 13.18	66.55 ± 15.72	31.37 ± 14.72	90.26 ± 9.04	89.21 ± 10.95	12.92 ± 14.52
	ECG	72.66 ± 13.42	68.36 ± 16.85	30.29 ± 14.92	91.75 ± 9.73	90.48 ± 12.87	10.41 ± 14.91
	EDA	68.58 ± 20.39	64.14 ± 24.49	32.63 ± 19.99	83.91 ± 17.82	82.09 ± 21.73	17.61 ± 16.69
	EMG	67.86 ± 11.66	58.87 ± 14.34	36.27 ± 12.80	79.83 ± 10.79	73.73 ± 16.28	30.82 ± 19.78
	Resp	82.85 ± 9.16	81.44 ± 9.42	18.24 ± 9.18	97.99 ± 1.41	97.99 ± 1.41	2.42 ± 1.88
	Temp	75.02 ± 10.34	67.67 ± 11.62	28.47 ± 10.89	81.51 ± 11.9	76.14 ± 17.05	27.69 ± 21.24
	SL Fusion	88.28 ± 6.06	85.85 ± 8.81	13.09 ± 7.20	96.66 ± 3.98	96.65 ± 3.99	4.16 ± 4.79
	FL Fusion	85.39 ± 9.35	82.95 ± 12.03	15.94 ± 9.96	96.69 ± 4.35	96.61 ± 4.48	4.48 ± 6.15
Wrist	ACC	73.74 ± 16.60	72.41 ± 17.43	27.07 ± 16.64	87.56 ± 12.09	85.78 ± 15.03	16.78 ± 18.31
	BVP	76.02 ± 10.38	72.53 ± 12.60	25.70 ± 9.72	87.81 ± 9.36	87.08 ± 10.64	14.44 ± 12.7
	EDA	67.53 ± 18.18	63.78 ± 18.90	32.63 ± 17.30	81.64 ± 20.41	80.7 ± 22.26	15.89 ± 16.82
	Temp	59.20 ± 7.13	46.33 ± 11.06	45.74 ± 8.68	69.93 ± 0.9	57.56 ± 1.18	49.99 ± 0.04
	SL Fusion	82.70 ± 12.10	80.80 ± 13.42	18.56 ± 12.45	93.45 ± 8.49	92.7 ± 10.11	9.55 ± 13.05
	FL Fusion	83.69 ± 10.85	81.61 ± 12.31	17.56 ± 11.57	94.16 ± 9.09	93.62 ± 10.58	7.24 ± 12.4
Chest + Wrist	SL Fusion	76.73 ± 11.53	71.43 ± 14.11	25.1 ± 10.06	92.51 ± 8.56	91.99 ± 9.84	8.95 ± 11.51
	FL Fusion	86.5 ± 7.94	83.01 ± 10.32	15.27 ± 8.97	97.75 ± 2.55	97.74 ± 2.56	2.59 ± 3.19

positive rate (FPR), receiver operating characteristics, and balanced F_1-score. We used some of them, i.e. the accuracy measure, F_1-score, and FPR:

$$\texttt{Accuracy} = \frac{TP+TN}{TP+FN+TN+FP} \tag{2}$$

$$F_1\texttt{-Score} = 2 \times \frac{\texttt{Precision} \times \texttt{Recall}}{\texttt{Precision} + \texttt{Recall}} \tag{3}$$

$$\texttt{FPR} = \frac{FP}{FP+TN} \tag{4}$$

where $\texttt{Recall}$ (also known as sensitivity) represents the system's ability to detect a specific emotion, $\texttt{Recall} = TP/(TP+FN)$; $\texttt{Precision}$ measure (also called positive predictive value) the system's ability of being correct on a predicted emotion: $\texttt{Precision} = TP/(FP+TP)$.

Accuracy reflects the number of correctly classified patterns among the samples, and, thus, it is a *probability of success* in recognizing the right class of an instance. However, in the case of highly imbalanced datasets, the accuracy measure (2) is misleading. A classifier that is very effective in predicting the majority class, but misses most of the minority classes, may easily have very high accuracy [42].

The F_1-score is a weighted average (harmonic mean) of precision and recall rates, representing the system's balanced ability to detect a specific emotion correctly, where an F_1-score reaches its best value at 1 (perfect precision and recall) and the worst one at 0 [41], [43]. The F_1-score measure provides a way of combining the recall and the precision in order to get a single measure that captures both properties.

VIII. EXPERIMENTAL RESULTS

As described in the previous section, the SMA was trained and prepared for deployment using the real-world signals from the WESAD dataset collected from a wearable sensor network. Depending on the type of modalities and fusion technique, the model can take from 136.06 seconds (for signals such as ECG) to 152.78 seconds (accelerometer data) for training a 100 epoch model on a NVIDIA V100 graphic card. In comparison, an inference will take 0.84 seconds (for signals such as ECG) to 0.89 seconds (accelerometer data) for 2122 samples. That is in general it would take around 0.04 milliseconds for 1 sample. In this section, we report the performance of the proposed SMA. The goals of our experiment are as follows:

1) Estimate the SMA performance in stress classification mode, that is, 'in Normal state' vs. 'Stress state' vs. 'Amusement state'.
2) Estimate the SMA performance in stress detection mode, i.e. detection stress state vs. otherwise states.
3) Compare an achieved performance with the one reported for the known stress classifiers.

We also address questions related to the biases in the training sets, optimization of the WSN, and machine reasoning mechanism.

A. EMOTION CLASSIFICATION MODE OF THE SMA

We will distinguish two outcomes of the stress analysis: 1) recognize a stress state among other emotional states (emotion classification mode), and 2) detect the stress state (stress detection mode). Both modes are useful for the first responder who uses a WSN, and a personal SMA in training, examination, and operation.

Table 2 reports the results of the SMA performance in the following measures: Accuracy (2), F1-Score (3), and False-Positive Rate (4). The first column represents three groups of modalities: sensors deployed on the 1)'Chest', 2) 'Wrist', and 3) a total sensor network 'Chest + Wrist'. Each of these groups is represented by one of the two fusion strategies: 'SL Fusion' and 'FL Fusion'. The rest of the columns represent two performance measures. The results of experiments presented in Table 2 are valuable from various perspectives such as the deployment of the WSN, fusion strategy, and optimization.

1) DEPLOYMENT OF WEARABLE SENSOR NETWORK

A WSN must satisfy the deployment requirements in various actions and situations encountered by the first responder.

Example 4: It follows from Table 2 that:

– For the wrist sensor deployment is preferable, the maximum overall 'Accuracy' $= 83.69 \pm 10.85\%$ is achieved using the FL fusion.

– The worst accuracy is $83.69 - 10.85 \approx 72\%$ which is unacceptable for decision-making support.

– Taking into account the imbalance of data, the accuracy can further degrade to: 'F_1-Score' $= 81.61 \pm 12.31\%$.

2) FUSION STRATEGY

Choosing a fusion method is of critical importance because many factors must be taken into account. Those include the balance of the fused components (e.g. signals, features, decisions) with respect to their contribution (e.g. comparable weight, information content, reliability of sources, and credibility of information), as well as sensitivity to potential attacks. Despite the rich theoretical and practical experience, there are no common rules, – each case is unique from a fusion perspective. This is the reason that we examined the two fusion methods for the modality groups in our experiment as reported in Table 2.

Example 5: Given the WSN,

– Consider the 'Chest' sensor group. Fusion at sensor level 'SL fusion' is preferable by the measures of efficiency (except the attack risk), e.g. overall accuracy 'Accuracy' $= 88.28 \pm 6.06\%$.

– Consider the 'Chest + Wrist' group. The accuracy can reach $76.73 - 11.53 \approx 65\%$ which is unacceptable for reliable decision-making.

Example 6: Consider the 'Chest + Wrest' sensor group.

– Accuracy of the feature-level fusion 'FL fusion' is $86.50 - 7.94 \approx 78\%$.

– The other measures show much lower performance, and thus, the study for this sensor group proceeds to the stress detection mode only.

Example 7: Consider the 'Chest' sensor group. The RESP, TEMP, and ECG are the most informative factors.

B. BIASES

The SMA development and design face multiple biases such as:

– Training set bias, e.g. the same subject is used for both training and validation;

– Performance bias, e.g. there is a difference between the overall accuracy measure and F_1 score measure.

– Subject-related bias; this is the reason that our approach focuses on the individual-centric SMA. If the subject's identity is known, a subject-biased network is much better in identifying his/her emotions compared to a more generalized model.

There are other kinds of biases that are out of the scope of this paper, in particular, the AI bias, i.e. reasoning, judgment, or decision-making provided by intelligent systems; and the bias of synthetic dataset, i.e. there are biases in training using authentic, synthetic, and semi-synthetic data (benchmarks).

TABLE 3. Emotion classification performance: Normal vs. Stress vs. Amusement.

Method	Modalities	Accuracy	F1-Score
Adaboost [3]	Chest Physio	80.34 ± 0.43	72.51 ± 0.17
RF [3]	Wrist Physio	76.12 ± 0.42	66.33 ± 0.36
Adaboost [3]	All	79.57 ± 0.93	68.85 ± 0.89
MMSF [44]	All Chest	83	81
MMSF [44]	All	85	86
Ours	All Chest	88.28 ± 6.06	82.95 ± 12.03
Ours	All	86.50 ± 7.94	83.01 ± 10.32

All of these biases impact the performance of the SMA. For example, the training set bias is introduced through the random sampling process. Since this process does not distinguish between the subjects, this process is very likely to sample similar data of the same subject that will be used for both training and validation. This process makes the classifier to focus on the unique patterns of individuals rather than a generalized model for various emotions. Table 3 shows the state-of-art performance which involves the use of the validation procedure suggested in [40]. It can be seen that our method has similar performance to the self-supervised CNN method proposed in [40]. It should be noted that this form of validation produces a much higher performance compared to Experiment 1, due to the existence of bias in the training dataset.

C. STRESS DETECTION MODE OF THE SMA

In contrast to the scenarios in which stress must be recognized among other emotional states such as Neutral, Stressed, Amused, and Meditated, the related problem may be resolved for another scenario. For the first responders, the task is formulated as a stress *detection* problem. The decision rule (function) in such case is as follows:

$$\begin{matrix}\texttt{Decision}\\ \texttt{Rule}\end{matrix} \equiv \begin{cases} 1, & \texttt{decide } H_1 \text{ (Stress detected)} \\ 0, & \texttt{decide } H_0 \text{ (otherwise)} \end{cases}$$

This partitions the data space into two regions (instead of many), where H_0 and H_1 are null and alternative hypotheses, respectively. Several important observations can be carried out of the comparative analysis of the results.

Observation 1: All performance measurements are better for the case of detection. For example, in Table 2 (left plane), the accuracy 'Acc.' is reported as 76.73% for 'SL fusion' and 86.50% for 'FL fusion'. In the case of stress detection (Table 2 (right plane)), the accuracy 'Acc.' for the same fusion modes is 92.51% and 97.75%, respectively.

Observation 2: The experimental results indicate the following priority:

$$\overbrace{\overset{1}{\text{RESP}};\quad \overset{2}{\text{ECG}};\quad \overset{3}{\text{ACC}};\quad \overset{4}{\text{EDA}};\quad \overset{5}{\text{TEMP}};\quad \overset{6}{\text{EMG}};}^{\textit{Ranging factors in stress detection mode}}$$

$$\underbrace{\overset{1}{\text{RESP}};\quad \overset{2}{\text{TEMP}};\quad \overset{3}{\text{ECG}};\quad \overset{4}{\text{ACC}};\quad \overset{5}{\text{EDA}};\quad \overset{6}{\text{EMG}};}_{\textit{Ranging factors in stress classification mode}}$$

```
Stress Assistant
09.12.2020
2.45.21pm
Firefighter 01
Warning level: High
Status: High risk of stress –
Suffered severe blast injury
in an explosion
ACC: Extreme Movement
ECG: Abnormal Heart Rate
EDA: Slight Increase
EMG: Abnormal Conductivity
RESP: Heavy Breathing
TEMP: Normal
```

```
Stress Assistant
09.12.2020
2.45.21pm
Firefighter 02
Warning level: Low
Status: Low risk of stress
– Climbed a multi-storey
building during training
ACC: Intense Movement
ECG: Fast Heart Rate
EDA: Slight Increase
EMG: Normal
RESP: Intense Breathing
TEMP: Normal
```

```
Stress Assistant
09.12.2020
2.45.21pm
Firefighter 03
Warning level: Very High
Status: High risk of stress –
Suffered light burn injury –
Damaged Sensors
ACC: Intense Movement
ECG: Extreme Heart Rate
EDA: Disconnected/Damaged
EMG: Disconnected/Damaged
RESP: Increased Breathing
TEMP: Normal
```

FIGURE 10. Examples of the SMA protocols for firefighter combat: high risk status (left), low risk status (central), and very high risk (right).

The RESP and EMG are the most and least informative factor in both modes, respectively. The role of factor TEMP is decreased in stress detection mode (it number 5 instead of number 2).

Table 4 compares the results.

TABLE 4. Stress detection performance: No Stress vs. Stress.

Method	Modalities	Accuracy	F1-Score
[3]	Chest	93.12	91.47
[3]	Wrist	88.33± 0.25	86.10 ± 0.29
[3]	All	92.28	90.74
Ours	Chest	96.69 ± 4.35	96.61 ± 4.48
Ours	All	97.75 ± 2.55	97.74 ± 2.56

Observation 3: The accuracy of 97.75% is achieved in the detection mode of the SMA, against 92% reported in work [3].

D. EXAMPLES OF PROTOCOLS

The stress state of the first responders is continuously monitoring by the SMA and reporting using *protocol* format. The protocol is available to the subject via, in particular, a wearable display, and is transmitted to an operations center. Examples of the protocols are given in Fig. 10:

Firefighter # 1 (Left Plane): A high risk of stress is reported by the SMA. It is caused by suffering a severe blast injury due to an explosion (as detected by the ACC) with high probability.

Firefighter # 2 (Central Plane): A low risk of stress is reported by the SMA. It is caused by climbing a multi-story building during training (as detected by the ACC).

Firefighter # 3 (Right Plane): A very high risk of stress is reported by the SMA. It is caused by suffering a low-degree burn injury; the elevated environment temperature has damaged both the EDA and EMG sensors.

E. DIAGNOSTIC MODE

One of the distinguishing features of the proposed SMA architecture is the possibility to use the causal network in two modes: *feed-forward reasoning* for the fusion of the Res-TCN outcomes, and *backward reasoning* for inference and diagnosis.

From a system view, the causal network (Fig. 7) is an embodiment of decision-level fusion. While each "sensor" node represents the output of a trained machine-learning model, "Valid" node 'fuses' the results from all sensors. "Type" node provides the evidence or condition to be used for performance calculation, i.e. the accuracy of decision-level fusion represented by "Match" node.

Consider the SMA performance given that the only RESP channel is used. The highest accuracy achieved using this channel is 84.84% after probabilistic inference. The original accuracy of 82.85% from the Res-TCN is boosted via the use of a BN for fusion.

Fig. 11 represents an example of an inference test applied to the original causal network shown in Fig. 7. The probabilities shown in Fig. 11 are the marginalized conditional probabilities given the evidence (Type = RESP). The CPTs for each node in the BN are calculated based on the output distribution of the stress detection ML model.

The six "sensor" nodes (D, A, T, G, C, R) are the individual nodes describing the distribution of the predicted emotion classes. For example, node A (accelerometer) contains the probability distribution as follows: Amuse = 15.27, Stress = 26.3, and Base = 58.63. These probabilities represent the total positive predictions (True Positives + False Positives) of the Res-TCN. This indicates that given 100 samples, 15 are classified as Amusement, 26 as Stress, and 59 as Baseline. As seen, the estimated distributions provide only positive predictions, it does not indicate the difference between true positives and false positives.

The aforementioned nodes' predictions are fused together to provide a better estimate of the emotion class. The child node "Valid" represents the combined decision of the six

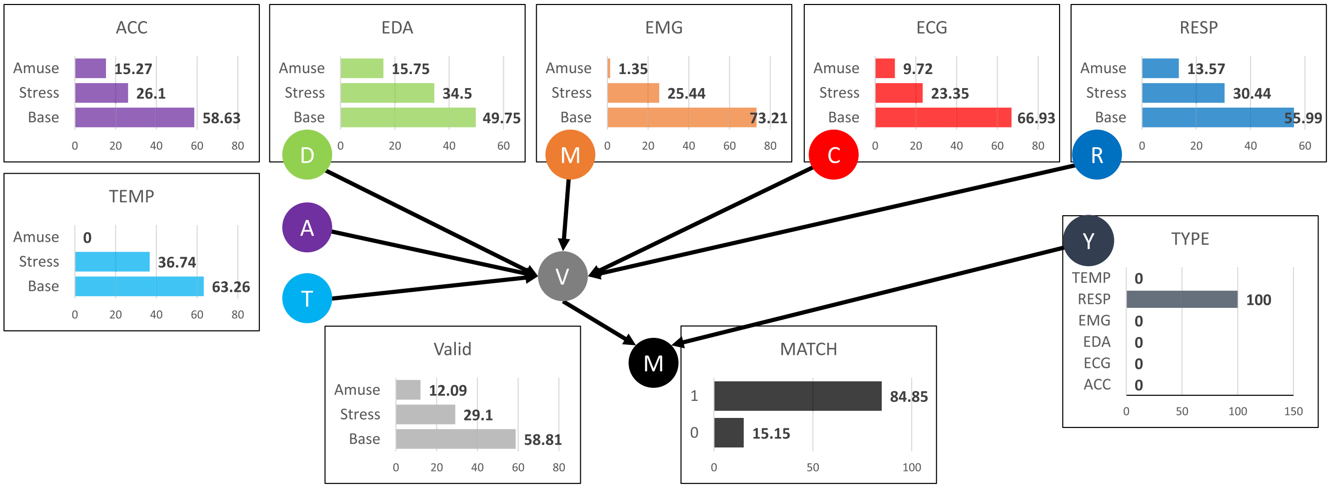

FIGURE 11. Example of inference test. Using the RESP sensor only for decision produces an accuracy of 84.84% which is higher than the original 82.85% when using data from multiple sensors.

parent nodes. For example, if nodes D, A, and T "predict" the Amusement, nodes M and C "predict" Stress, and node R "predicts" baseline, node "Valid" will yield Amusement = 3/6, Stress = 2/6, and Baseline = 1/6.

The sensor nodes as well as the "Valid" node describe the predicted distribution, while the "Type" node describes the ground-truth distribution. The predicted distribution is the probability distribution estimated by the Res-TCN, while the ground-truth distribution is the one provided by the dataset. Since the predicted distribution does not differentiate between true and false positives, the "Match" node is necessary in order to extract the correct predictions (true positives). This process is performed by measuring the similarity between the ground-truth and predicted distributions. A positive match happens once both distributions predict the same class (true positives or true negatives), whereas a negative match occurs when both distributions predict different classes (false positives or false negatives).

Provided the evidence (Type = RESP), the BN estimates the similarity between the ground-truth distribution of the RESP sensor and the predicted distribution (after fusion). In this scenario, we get a matching similarity of 84.84% between the ground-truth and predicted distribution. This value represents the accuracy of decision-level fusion and is greater than the original 82.85%. This suggests that the distribution from the other sensors provide useful information resulting in a 2% increase in performance.

IX. CONCLUSION, RECOMMENDATIONS, AND FUTURE WORK

The following key conclusions provide insights for developers of stress detectors for first responders:

1) The stress detector shall be personalized to the wearer to reflect the uniqueness of the individual and his/her reaction to stress. On the other hand, this detector is an integral part of the decision support process.
2) Residual TCN is the preferred ML approach used for learning the individual's response to stress, captured by the time series data from the WSN.
3) The proposed architecture and prototype bridge the research and development gaps in the current technological means to support the work of first responders. In general, this work contributes to the *fundamentals* of an intelligent SMA design.

In the intelligent SMA and DSS design (Fig. 1), the V-C-M survival content is present in the perception-action cycle:

– The SMA is an *integral part* of the intelligent decision support process, a perception-action cycle;
– The DSS supports the first responder in *all V-C-M survival dimensions*;
– The SMA and the WSN *capture* the V-C-M survival spectrum, in order to achieve the required performance, and
– The SMA is an *individual-centric* system that continuously learns from the vitals of the first responder in order to make a reliable stress recognition (detection), monitoring, and corresponding decision support.

There are a few recommendations for future work that are derived from our experience in developing the intelligent SMA:

– The problem of stress detection should be considered in a broader context of the first responder support. Our work is only the first step which covers an introduction to a systematic approach. The NIST roadmap for firefighters [2], a roadmap for Homeland Security first responder coping [9], a guide for combat and operational stress control [12], and a guide to military medicine [13] define the research and development horizon for autonomous decision support to human operators, regarding decisions in both actions and in preventing injuries.

– Our experimental study is limited by the WESAD dataset. Large benchmark databases are currently unavailable, thus prompting for development of machine generated or *synthetic data*. For example, a generation of synthetic ECG was proposed in [45], and the EMG has been simulated in [46].

REFERENCES

[1] D. Grossman and B. K. Siddle, "Psychological effects of combat," in *Encyclopedia Violence, Peace Conflict*, 2nd ed. L. R. Kurtz, Ed. Amsterdam, The Netherlands: Elsevier, 2008.

[2] C. Grant, A. P. Hamins, N. P. Bryner, A. W. Jones, and G. H. Koepke, *Research Roadmap for Smart Fire Fighting*. Gaithersburg, MD, USA: National Institute of Standards and Technology (NIST), Special Publication 1191, 2015.

[3] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing WESAD, a multimodal dataset for wearable stress and affect detection," in *Proc. 20th ACM Int. Conf. Multimodal Interact.*, Oct. 2018, pp. 400–408.

[4] S. K. D'mello and J. Kory, "A review and meta-analysis of multimodal affect detection systems," *ACM Comput. Surveys*, vol. 47, no. 3, pp. 1–36, Apr. 2015.

[5] E. I. Gaura, J. Brusey, J. Kemp, and C. D. Thake, "Increasing safety of bomb disposal missions: A body sensor network approach," *IEEE Trans. Syst., Man, C, Appl. Rev.*, vol. 39, no. 6, pp. 621–636, Nov. 2009.

[6] S. Rodrigues, J. S. Paiva, D. Dias, G. Pimentel, M. Kaiseler, and J. P. S. Cunha, "Wearable biomonitoring platform for the assessment of stress and its impact on cognitive performance of firefighters: An experimental study," *Clin. Pract. Epidemiol. Mental Health*, vol. 14, no. 1, pp. 250–262, Oct. 2018.

[7] F. Perroni, L. Guidetti, L. Cignitti, and C. Baldari, "Psychophysiological responses of firefighters to emergencies: A review," *Open Sports Sci. J.*, vol. 7, no. 1, pp. 8–15, Jan. 2014.

[8] D. Dias and J. P. S. Cunha, "Wearable health devices—Vital sign monitoring, systems and technologies," *Sensors*, vol. 18, no. 8, p. 2414, 2018.

[9] Homeland Security. (2020) *COVID-19 Information for the First Responder Community*. Accessed: Sep. 2020. [Online]. Available: https://www.dhs.gov/science-and-technology/covid-19-info-first-responders

[10] G. Giannakakis, M. Pediaditis, D. Manousos, E. Kazantzaki, F. Chiarugi, P. G. Simos, K. Marias, and M. Tsiknakis, "Stress and anxiety detection using facial cues from videos," *Biomed. Signal Process. Control*, vol. 31, pp. 89–101, Jan. 2017.

[11] H. Zhang, L. Feng, N. Li, Z. Jin, and L. Cao, "Video-based stress detection through deep learning," *Sensors*, vol. 20, no. 19, p. 5552, Sep. 2020.

[12] (2016). *Combat and Operational Stress Control*. US Marine Corps. MCTP 3-30E. Accessed: Oct. 2020. [Online]. Available: https://www.marines.mil/News/Publications/MCPEL/Electronic-Library-Display/Article/899535/mctp-3-30e-formerly-mcrp-6-11c/

[13] C. Crump and L. M. Schlachta-Fairchild, "Achieving a trusted, reliable, AI-ready infrastructure for military medicine and civilian care," in *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications II*, vol. 11413. Bellingham, WA, USA: SPIE, 2020, Art. no. 114130C.

[14] D. Dias, N. Ferreira, and J. P. S. Cunha, "VitalLogger: An adaptable wearable physiology and body-area ambiance data logger for mobile applications," in *Proc. IEEE 14th Int. Conf. Wearable Implant. Body Sensor Netw. (BSN)*, May 2017, pp. 71–74.

[15] M. Umlauft, C. Raffelsberger, A. Kercek, A. Almer, T. Schnabel, P. Luley, and S. Ladstaetter, "A communication and multi-sensor solution to support dynamic generation of a situational picture," in *Proc. 3rd Int. Conf. Inf. Commun. Technol. Disaster Manage. (ICT-DM)*, Dec. 2016, pp. 1–7.

[16] J. Blasco, T. M. Chen, J. Tapiador, and P. Peris-Lopez, "A survey of wearable biometric recognition systems," *ACM Comput. Surveys*, vol. 49, no. 3, pp. 1–35, Dec. 2016.

[17] S. Haykin, *Cognitive Dynamic Systems: Perception-Action Cycle, Radar and Radio*. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[18] M. V. Zuck and R. J. Frey, *The Gale Encyclopedia of Medicine*, 5th ed. J. L. Longe and A. Gale, Eds. Farmington Hills, MI, USA: Cengage, 2015.

[19] R. S. Lazarus and S. Folkman, *Stress, Appraisal, and Coping*. New York, NY, USA: Springer, 1984.

[20] A. J. Bandodkar, W. Jia, and J. Wang, "Tattoo-based wearable electrochemical devices: A review," *Electroanalysis*, vol. 27, no. 3, pp. 562–572, Mar. 2015.

[21] M. Yuan, R. Das, R. Ghannam, Y. Wang, J. Reboud, R. Fromme, F. Moradi, and H. Heidari, "Electronic contact lens: A platform for wireless health monitoring applications," *Adv. Intell. Syst.*, vol. 2, no. 4, Apr. 2020, Art. no. 1900190.

[22] G. Tartare, X. Zeng, and L. Koehl, "Development of a wearable system for monitoring the firefighter's physiological state," in *Proc. IEEE Ind. Cyber-Physical Syst. (ICPS)*, May 2018, pp. 561–566.

[23] F. Seoane, I. Mohino-Herranz, J. Ferreira, L. Alvarez, R. Buendia, D. Ayllón, C. Llerena, and R. Gil-Pita, "Wearable biomedical measurement systems for assessment of mental stress of combatants in real time," *Sensors*, vol. 14, no. 4, pp. 7120–7141, Apr. 2014.

[24] S. Yoon, J. K. Sim, and Y.-H. Cho, "A flexible and wearable human stress monitoring patch," *Sci. Rep.*, vol. 6, no. 1, pp. 1–11, Mar. 2016, Paper 23468.

[25] L. M. Roa, J. Reina-Tosina, A. Callejón-Leblic, D. Naranjo, and M. A. Estudillo-Valderrama, "Intrabody communication," in *Handbook of Biomedical Telemetry*. K S. Nikita, Ed. Hoboken, NJ, USA: Wiley, 2014, pp. 252–300.

[26] A. Reichman, J.-I. Takada, D. Bajić, K. Y. Yazdandoost, W. Joseph, L. Martens, C. Roblin, R. D'Errico, C. Oliveira, and L. M. Correia, "Body communications," in *Pervasive Mobile and Ambient Wireless Communications, Signals and Communication Technology*. London, U.K.: Springer-Verlag, 2012, pp. 609–660.

[27] Z. Guan, G. E. Santagati, and T. Melodia, "Distributed algorithms for joint channel access and rate control in ultrasonic intra-body networks," *IEEE/ACM Trans. Netw.*, vol. 24, no. 5, pp. 3109–3122, Oct. 2016.

[28] S. S. Therrien, "A Bayesian model to incorporate human factors in commanders' decision making," M.S. thesis, Naval Postgraduate School, Monterey, CA, USA, 2002.

[29] N. Nair, C. Thomas, and D. B. Jayagopi, "Human activity recognition using temporal convolutional network," in *Proc. 5th Int. Workshop Sensor-Based Activity Recognit. Interact.*, Sep. 2018, pp. 1–8.

[30] Y. A. Farha and J. Gall, "MS-TCN: Multi-stage temporal convolutional network for action segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3575–3584.

[31] C. Lea, M. D. Flynn, R. Vidal, A. Reiter, and G. D. Hager, "Temporal convolutional networks for action segmentation and detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 156–165.

[32] S.-J. Li, Y. AbuFarha, Y. Liu, M.-M. Cheng, and J. Gall, "MS-TCN++: Multi-stage temporal convolutional network for action segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Sep. 4, 2020, doi: 10.1109/TPAMI.2020.3021756.

[33] P. Bota, C. Wang, A. Fred, and H. Silva, "Emotion assessment using feature fusion and decision fusion classification based on physiological data: Are we there yet?" *Sensors*, vol. 20, no. 17, p. 4723, Aug. 2020.

[34] L. M. Dinca and G. P. Hancke, "The fall of one, the rise of many: A survey on multi-biometric fusion methods," *IEEE Access*, vol. 5, pp. 6247–6289, 2017.

[35] J. Perl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA, USA: Morgan Kaufmann, 1988.

[36] R. Lotfian and C. Busso, "Building naturalistic emotionally balanced speech corpus by retrieving emotional speech from existing podcast recordings," *IEEE Trans. Affect. Comput.*, vol. 10, no. 4, pp. 471–483, Oct. 2019.

[37] M. Jaiswal, C.-P. Bara, Y. Luo, M. Burzo, R. Mihalcea, and E. M. Provost, "Muse: A multimodal dataset of stressed emotion," in *Proc. 12th Lang. Resour. Eval. Conf.*, 2020, pp. 1499–1510.

[38] C. Busso, M. Bulut, C.-C. Lee, A. Kazemzadeh, E. Mower, S. Kim, J. N. Chang, S. Lee, and S. S. Narayanan, "IEMOCAP: Interactive emotional dyadic motion capture database," *Lang. Resour. Eval.*, vol. 42, no. 4, pp. 335–359, Dec. 2008.

[39] D. Watson, L. A. Clark, and A. Tellegen, "Development and validation of brief measures of positive and negative affect: The PANAS scales," *J. Personality Social Psychol.*, vol. 54, no. 6, p. 1063, 1988.

[40] P. Sarkar and A. Etemad, "Self-supervised ECG representation learning for emotion recognition," 2020, *arXiv:2002.03898*. [Online]. Available: http://arxiv.org/abs/2002.03898

[41] A. Luque, A. Carrasco, A. Martín, and A. de las Heras, "The impact of class imbalance in classification performance metrics based on the binary confusion matrix," *Pattern Recognit.*, vol. 91, pp. 216–231, Jul. 2019.

[42] A. Maratea, A. Petrosino, and M. Manzo, "Adjusted F-measure and kernel scaling for imbalanced data learning," *Inf. Sci.*, vol. 257, pp. 331–341, Feb. 2014.

[43] D. Hand and P. Christen, "A note on using the F-measure for evaluating record linkage algorithms," *Statist. Comput.*, vol. 28, no. 3, pp. 539–547, May 2018.

[44] J. Lin, S. Pan, C. S. Lee, and S. Oviatt, "An explainable deep fusion network for affect recognition using physiological signals," in *Proc. 28th ACM Int. Conf. Inf. Knowl. Manage.*, Nov. 2019, pp. 2069–2072.

[45] P. E. McSharry, G. D. Clifford, L. Tarassenko, and L. A. Smith, "A dynamical model for generating synthetic electrocardiogram signals," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 3, pp. 289–294, Mar. 2003.

[46] M. E. Esquivel-Frausto, J. A. Guerrero, and J. E. Macías-Díaz, "Activity pattern detection in electroneurographic and electromyogram signals through a heteroscedastic change-point method," *Math. Biosci.*, vol. 224, no. 2, pp. 109–117, Apr. 2010.

KENNETH LAI (Student Member, IEEE) received the B.Sc. and M.Sc. degrees from the University of Calgary, Calgary, AB, Canada, in 2012 and 2015, respectively, where he is currently pursuing the Ph.D. degree with the Department of Electrical and Computing Engineering. His areas of interests include biometrics and its application to security and health care systems.

SVETLANA N. YANUSHKEVICH (Senior Member, IEEE) received the Dr.Tech.Sc. degree from the Warsaw University of Technology, in 1999. She is currently a Professor with the Department of Electrical and Computer Engineering, University of Calgary, where she is also directing the Biometric Technologies Laboratory, and conducting research in the area of biometric-based authentication technologies. She is currently a member of the ISATC. She was the Chair of the Biometrics Task Force for the IEEE ISATC, from 2017–2020.

VLAD P. SHMERKO received the Dr.Tech.Sc. degree from the Latvian Academy of Sciences, Riga, Latvia, in 1990. He is currently a Researcher with the Biometric Technologies Laboratory and an Adjunct Professor with the Department of Electrical and Computer Engineering, University of Calgary, Canada. He is a Fellow and a Chartered Engineer with the Institution of Engineering and Technology, U.K.