Pattern Recognition Letters 145 (2021) 81–87

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Hierarchical deep neural network for mental stress state detection using IoT based biomarkers

Akshi Kumar[a,*], Kapil Sharma[b], Aditi Sharma[a]

[a] Department of CSE, Delhi Technological University, Delhi 110042, India
[b] Department of IT, Delhi Technological University, Delhi 110042, India

ARTICLE INFO

Article history:
Received 9 May 2020
Revised 18 December 2020
Accepted 24 January 2021
Available online 4 February 2021

MSC:
41A05
41A10
65D05
65D17
Deep learning
IoT sensors
Biomarkers
Mental health
Stress

ABSTRACT

Affective state recognition at an early stage can help in mood stabilization, stress and depression management for mental well-being. Pro-active and remote mental healthcare warrants the use of various biomarkers to detect the affective mental state of the individual by evaluating the daily activities. With the easy accessibility of IoT-based sensors for healthcare, observable and quantifiable characteristics of our body, physiological changes in the body can be measured and tracked using various wearable devices. This work puts forward a model for mental stress state detection using sensor-based bio-signals. A multi-level deep neural network with hierarchical learning capabilities of convolution neural network is proposed. Multivariate time-series data consisting of both wrist-based and chest-based sensor bio-signals is trained using a hierarchy of networks to generate high-level features for each bio-signal feature. A model-level fusion strategy is proposed to combine the high-level features into one unified representation and classify the stress states into three categories as baseline, stress and amusement. The model is evaluated on the WESAD benchmark dataset for mental health and compares favourably to state-of-the-art approaches giving a superlative performance accuracy of 87.7%.

1. Introduction

The healthcare industry has radically changed as the Internet of Things (IoT) have recalibrated endless applications within the structure. The current generation, healthcare 4.0 improves clinical treatment such that medical practitioners can monitor personal health information shared through sensors to be more watchful and connected with the patients proactively. The smart IoT based devices available in the market have helped patient management by remotely monitoring health conditions and timely alerting the hospital about any irregularities using biomarkers on daily basis [1]. Smart healthcare, as shown in Fig. 1, works both on clinical and non-clinical data. Clinical trials for any disease require the subject to visit the hospital and always be available for physical examination. With the help of IoT-based sensors, the health condition of the user can be tracked remotely using wearable like a wristwatch, or with the implantation sensor like a pacemaker. In the non-clinical collection of data, the bio-signals of the subject can be traced with the help of their smart devices such as mobile phones, daily/ monthly activities like walking, running, and sitting to track the health of the user.The benefits of using IoT in healthcare includes Higher patient engagement, better patient outcomes, timely intervention and diagnosis, proactive treatments.

To have healthy lives, we rely on three basic and interrelated human capacities: *Affect*: the feelings that we experience as part of our everyday lives. *Behaviour*: The way we act and present ourselves in social interactions. and lastly, *Cognition*: What we think and the way we connect our thinking in social world [2]. These three are the Tripartite Model(ABC) of attitude in social psychology. Affect defines our feelings which we experience in form of mood and emotion, our mood and emotion change based upon the surrounding events. The affective state of an individual is a neuro-physiological state, which is consciously accessing the simple raw primitive feeling that a person have, it may be non-reflective in nature too. Broadly, a person have two affective states: affective emotional state and affective psychological state, the affective state of a person changes in response to a change in emotion, mood or affect. Affective psychological state of an individual identifies the mental state of a person, as a person only in sound affective mental state can perform the daily activities without any complications and can focus on their work. Stress in response to day-to-day activities can result in change of affective psychological/mental state of a person. the changes in the affective state can be expressed as a tone of voice, a frown, pressed lips, tear, an eye gaze along

* Corresponding author.
E-mail address: akshikumar@dce.ac.in (A. Kumar).

https://doi.org/10.1016/j.patrec.2021.01.030
0167-8655/

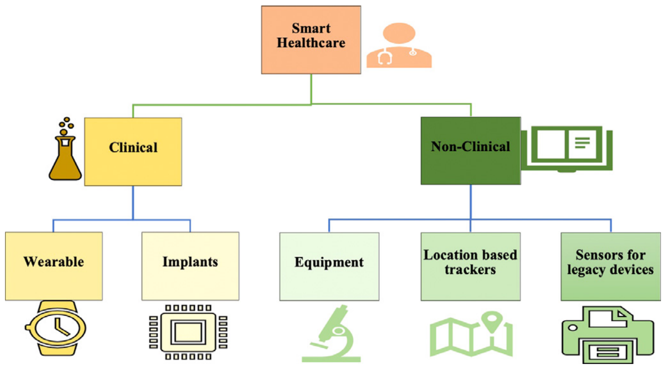

Fig. 1. IoT based sensors in health care 4.0.

with internal changes in the body such as change in temperature, heartbeat rate etc. Affective state recognition at an early stage can help in mood stabilization, stress and depression management for mental well-being.

The craving to succeed in this fast-paced life takes away the time to overhaul oneself. An individual encounters constant pressure to excel at everything, job stress, nagging by parents and peer pressure leading to increased risk of developing mental health problems. Undeniably, stress is a common problem in modern life psychology. As per a report by British Health and Safety Executive (HSE), stress accounted for 37% of all work-related ill health cases in 2015. Mental health is not visible to anyone and is hard to identify the real affective psychological state of a person. In contrast, physical health whenever deteriorates has observable signs & symptoms and we seek immediate diagnosis, advice and treatment for it from the healthcare professionals. Moreover, some of the mental health conditions like depression, acute stress syndrome, anxiety and insomnia have common symptoms, so classifying the correct type becomes challenging. For example, stress and anxiety both have symptoms like a dry mouth, sudden sweating, and increased breathing rate. Therefore to classify them as stress or anxiety, the subject needs to be evaluated for an extended period as an anxiety attack lasts only for few minutes and is often triggered by an external stimuli. The physical and psychological impacts can be cyclically linked: emotional distress and poor mental health can trigger or flare a physical health problem and, as a result, cause further distress. Likewise, poor physical health can lead to an increased risk of developing mental health problems. A mild amount of stress can be favourable, as it has been observed that a person gives near-optimal works performance under mild-stress. Eustress or beneficial stress [3] is often related to a positive challenge as compared to distress which has negative implications. However, prolonged and chronic stress can severely impact person's health, affect the whole body and increase the risk of developing certain illnesses. It can have several physical or psychological symptoms, which can make functioning on a daily basis more challenging.

A typical mental status examination (MSE) done by a professional is a standardized procedure used to evaluate the person's mental and emotional functioning. It involves a precise series of observations as well as some specific questions. The complete diagnostic picture is analysed using seven key evaluation parameters described as: Appearance, psychomotor behaviour, and attitude; Characteristics of speech; Affect and mood; Thought-content, Thought-form (Delusions, Illusions, Hallucinations), and concentration; Orientation; Memory ; General intellectual level ; Insight and judgement.

Formally, mental illnesses are health conditions involving changes in emotion, thinking or behaviour (or a combination of these) [4]. The general cognitive function is hindered to an extent that it can trigger inappropriate responses because those responses are based upon inaccurate thoughts. That is, the person finds difficult to stay focused, process information, store it in memory, and accurately respond. Mental illness is conceptualized as a clinically significant behavioural dysfunction or psychological syndrome. There are many different categories of mental/ psychological disorders defined in the ICD-10, 10th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD), a medical classification list by the World Health Organization (WHO) known as mental and behavioural disorders, ICD codes F00 to F99. Mental healthcare needs various biomarkers to detect the affective mental state of the individual by evaluating the daily activities. An individuals behaviour needs to be evaluated in different scenarios like his feeling while watching a movie, while driving, or while doing office work. Only then a psychologist can identify the actual mental condition of the person. Each psychological disorder has its own characteristic symptoms which can be measured through biomarkers and some general warning signs to alert the need of professional help. An intelligent mental illness diagnostic can support clinicians with early detection. Most of the work conducted to detect the affective mental health status of a user involves datasets formulated with the help of medical questionnaires [6,7]. These databases suffer due to lack of standardization of questionnaires, dishonest & unconscientious answers, unanswered questions and differences in understanding and interpretation. With healthcare 4.0, wearable IoT utilities can gather information, assess activity and other biomarkers, and even deliver interventions for various mental health conditions of an individual [5].

This work puts forward a model to detects the mental stress at an early stage by evaluating the different biomarkers indicative of affective mental health. As every individual acts differently on encountering the same situation. Stress is not dependent on a single attribute, and may work differently for different individuals. Therefore, to evaluate the stress level of an individual without prior medical history is a difficult task [6]. Biomarkers are the measures of observable and quantifiable characteristics of our body. Digital Biomarkers along with Traditional (bio-signal based) biomarkers can be used to track the affective state of an individual. Traditional biomarkers measures the physiological change in the body with bio-signals such as change in heartbeat, respiration rate and temperature etc using medical equipments in the lab. whereas, Digital biomarkers can be derived from various sources such as natural interaction with digital games, social media activities and other physiological changes in the body but collected through digital devices such as portables, implants, wearables etc. With the advancement in healthcare and information technology, the traditional biomarkers like heart rate, Blood pressure, electrocardiogram can also now be measured in real-time by having sensors in implants or wearable devices such as fitness watch. Although the biomarkers in psychology cannot be restricted to electro-physiological signals, as behavioural signals like posture, along with speech signals, social interactions and social media activities can also be used as observable traits that can indicate the status of mental health of an individual [10]. Digital biomarkers can measure observable traits from diverse sources such as smartphone, social media, wearable devices to accurately predict the affective mental state [8]. Some of these digital biomarkers are shown in Fig. 2.

In this paper, we have proposed a hierarchal deep neural network that takes as input wearable stress and affects detection dataset (WESAD) that contains the bio-signals of 15 individuals collected from the wrist-wearable device and chest-worn device for a time-span of 2 h. The hierarchal deep neural network consists of three levels. As WESAD contains bio-signals from 2 devices, so, to fetch the optimal values for every feature at each instance, at the first level, the sub-sub networks(SSN) for each bio-signal

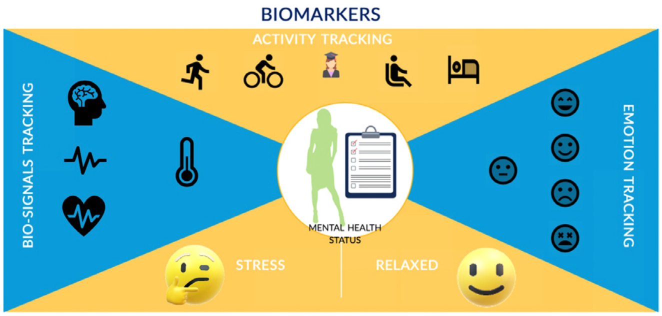

Fig. 2. Mental health predictor biomarkers.

are used. 10 SSNs are a 1-dimensional convolution neural network (1D-CNN) containing two convolution layers with batch normalization and max-pooling and one dense layer. These SSNs generate a high-level representation of the respective wrist and chest-based biomarkers which are input to the respective sub-networks (SNs) at the second level. The SNs at the second level are 1D-CNN containing two convolution layers with batch normalization and max-pooling and one dense layer. This produces a combination of the high-level representation features of each device type biomarker. In next level, separately learned device type biomarkers are combined into one unified representation realizing a model-level fusion strategy. Thus, the shared representation is given to a convolution layer which generates the final feature vector and uses a denser layer to process the feature vector. The output layer generates a regression output with linear activation to finally detect the mental state. The network is compared with 5 machine learning based models (Decision tree and Random forest classifier, LDA, KNN, AdaBoost), and one deep learning model (CNN with late fusion), which is the state-of-the-art result for the WESAD dataset proposed by Schmidt et al. in 2018 and Lin et al. in 2019. The primary contributions of this research are:

- A deep neural network with hierarchical learning capabilities is proposed for mental stress state detection.
- Both wrist-based and chest-based sensor bio-signals are used to generate high-level representation of features with generalization capabilities.
- Model-level fusion strategy is proposed to elucidate the correlation in data as different sub-networks are used to operate over features which are learned separately for each input type and then combined into one unified representation.

1.1. Related work

Mental health is as vital as physical health. Most of the organizations try to arrange motivational sessions or activity sessions to ensure their employees get a mental vacation and feel relaxed. This is important as a stressed or depressed person will always find it challenging to focus and so will always take extra time to pursue the same work in comparison to a mentally healthy and relaxed person. As the early stages of mental illness have invisible symptoms, it is not always possible to notice the change until the symptoms are persistent, increase in frequency and severity and interfere with life activities and roles [7]. Thus, early identification and intervention are necessary to recover and reclaim lives. Various artificial intelligence based techniques have been reported to supplement clinical practice in various mental healthcare studies. To analyze the behaviour of a person under stress, researchers have proposed machine learning-based techniques; however, the availability of a WESAD dataset has always been an issue. In 2018, Schmidt et al. [11] created and applied five different machine learning algorithms namely, Random Tree Classifier (RT), Decision Tree (DT), AdaBoost, K Nearest Neighbour (KNN), and Linear Discriminant Analysis (LDA) and given the state of the art results by providing the accuracy of 80%. In 2019, Lin et al. proposed a deep fusion network on the WESAD to optimize the accuracy of the prediction of stress. They used the late fusion method in deep neural networks and divided the model into four subnetworks, one tuned on the chest sensor dataset, and rest three tuned on the wrist sensor dataset, first one on EDA and Temperature, second on BVP and last on ACC. They attained the highest accuracy of 85% and F1 score as 0.86, which was a significant improvement on the result provided by Lin et al. [12]. The latest work on stress detection using WESAD is proposed by Indikawati and Winiarti [13]. They used three classifiers, namely logistic regression, decision tree, and random forest, and rather than evaluating the result into three categories, they added one more output category as meditation. Also, rather than applying each classifier on the complete dataset, they applied it on individual subjects, resulting in an accuracy of 88% to 99% for the individual subject.

WESAD is considered as the most recent benchmark dataset to analyze the mental health of a person since it contains the maximum number of biomarkers on a single subject to determine the affective state of the subject. Previously, many researchers have created datasets to evaluate the stress level. Picard et al. [14] built a dataset containing physiological data of a single person depicting eight different emotions for 20 days. As different individuals can represent different behaviors for the same emotional feeling; therefore, the dataset collected from a single source cannot predict accurately for all the users. Healey et al. [15] also created a dataset for evaluating stress using ECG, Electrodermal Activity and Respiration, and Electromyogram data. However, this dataset was only used to evaluate the stress of a driver, so it did not apply to all the subjects performing different actions. In 2012, DEAP was published by Koestra et al.; which contained the facial videos and EEG signals of the users to analyze the emotions using peripheral signals [16]. Although it used multiple subjects, the features used were limited, so DEAP was also unable to predict the emotion of all the users accurately. Maxhuni et al. [17] used mobile phones to construct data for analyzing the level of stress in a user. They used biomarkers like physical activity level, social interaction, and social activity along with the location of the user. A transfer learning model was used to predict the stress level and an accuracy of 76% was reported. In 2017, Gjoreski et al. [9] used only bio-signals from a wristwatch to evaluate the level of stress from 5 subjects evaluated over a span of 55 days using machine learning.

In the next section, we discuss the dataset used along with the model proposed. It is followed by a discussion on results and finally, the conclusion is presented in the last section.

2. Material and methods

Stress is a mental state of a person that can be triggered by external or internal stimuli, and these stimuli vary for different individuals. Consequently, to train a model, we need to understand the behaviour of a person under different situations. And as each person have different response on encountering an external stimuli, training a model on one subject will not identify the affective state of a different person. To understand the behaviour of different healthy individuals under the same situation, WESAD dataset was used. It has physiological changes occurring in 15 subjects while encountering different stimuli under controlled lab environment for the period of two years, collected with the help of wrist wearable and chest wearable digital biomarkers. The prediction of the level of stress in a user depends on several biomarkers like dry mouth, dilation of pupils, decreased digestion, but measuring every aspect from an individual with a single device is not possible, e.g to monitor the pupil dilation a continuous face recording cam-

Table 1
Biomarkers of the proposed model.

Biomarker	Device	Feature/ Significance
Blood Volume Pulse	RespiBAN, Empatica E4	BVP is the amount of blood in blood tissue during a certain time period. BVP also provides pulse rate and blood flow volume, as it is obtained by photoplethysmography.
Electrocardio- gram	RespiBAN	ECG provides the frequency of cardiac cycles. It is sensed using photodetectors, so is not able to be detected by wrist-wearable device.
Electrodermal Activity	RespiBAN, Empatica E4	EDA gives the flow of electricity through the skin. The changes arise in skin when brain sends the signal due to different emotion activation. Skin conduction increases when a person is under stress.
Electromyo- gram	RespiBAN	EMG is used to detect musculoskeletal movements. These signals can detect face and hand gestures.
Body Temperature	RespiBAN, Empatica E4	Skin temperature of the subject is measured using thermistor sensor. Body temperature is negatively correlated with stress.
Respiration	RespiBAN	RESP gives the person inhalation and exhalation rate. The slowed respiration rate shows the level of stress in a user.
Three Axis Acceleration	RespiBAN, Empatica E4	ACC gives an indication of different activities like lying, sitting, standing, walking, running and cycling by recording the human movement in all the three dimensions. Fast hand movement over short time depicts sign of mental stress.

era is must whereas automatic digestion monitoring is not feasible. However, it invades the privacy of the user. Thereby, data collection used is collected through sensors in a way that does not invade the privacy of the users.

2.1. Dataset

To identify the mental stress using behaviour biomarkers, we have used a sensor collected multimodal dataset which features psychological and motion data from both a wrist-worn device Empatica E4, and chest-worn device RespiBAN over 2 hours in a lab under controlled environment named WESAD. It is a publicly available data set for variable stress and affects detection collected by Schmidt et al. in a lab study on 17 subjects (but 2 subjects data being discarded), were 12 being male and 3 been female having a mean age of 27.5. Different biomarkers that were used to monitor the stress level of a person are Blood Volume Pulse (BVP), Electrocardiogram (ECG), Electrodermal Activity (EDA), Electromyogram (EMG), Body Temperature (TEMP), Respiration (RESP) and Three-Axis Acceleration (ACC) motion, the brief application of these biomarkers are shown in Table 1.

The data collected by RespiBAN was at 700 Hz, whereas the data collected by the wrist device was at low resolution, Empatica E4 records BVP at 64 Hz, EDA at 4 Hz, TEMP at 4 Hz, and ACC at 32 Hz. All these biomarkers contribute to identifying the mental state of a person, whether he is amused, stressed, or is in normal state i.e., baseline. Baseline state of each subject was recorded for 20 minutes doing basic activities like reading and walking, to understand the neutral affective state of the subject. To understand the affect of amusement in the subjects, they were shown 11 different short funny videos clips. And for identifying the stress state, they were exposed to Trier Social Stress test consisting of activities like public speaking and mental arithmetic tasks. Each subject has 12 features, and the results were self-reported by the users [8]. The dataset contains a total of 63000000 instances [8].

2.2. Convolution neural network (CNN)

Convolution Neural Network is the most commonly used deep neural network proposed for image processing but now validated for all types of data by converting it into feature vector representation [18,19]. Typically, in a CNN, the input is passed through convolution layers such that the output of the primary layer becomes the input for the subsequent layer. Non-linearity is added post every convolution operation using an activation function such as ReLU, to create a rectified feature map. Each non-linear layer is followed by a pooling layer which performs a down-sampling operation. Pooling operation helps to progressively reduce the size of the input representation and control over-fitting too [20], we can either use max, average or sum pooling. A fully connected layer also known as the dense layer is then attached to this series of convolution, non-linear and pooling layers which output the information from the convolutional networks. Structure of the CNN model is shown in Fig. 3.

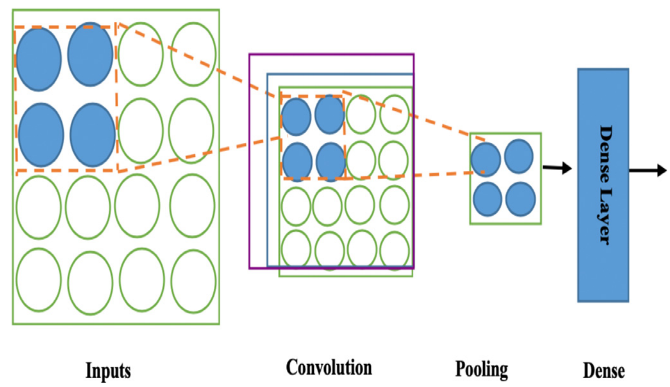

Fig. 3. CNN structure.

In our proposed hierarchical network we have three levels. At the bottom level, we use 1-D CNN for creating 10 different SSNs to generate the optimized feature value for every bio-signal of RespiBAN and Empatica E4. In the second layer, the output of SSNs are passed to two different SN depending upon the feature is obtained by RespiBAN or Empatica E4, for producing a combination of the high-level representation features of each device. Both SNs are also a 1D-CNN containing two convolution layers with batch normalization and max-pooling and one dense layer. And the top level is the classification level, where the result of SNs, separately learned device type biomarkers are combined into one unified representation realizing a model-level fusion strategy. As the dataset has approximately more than half-million values for a single attribute over a single subject, the fusion strategy is used. Typically, fusion strategies can be categorized into early, model-level and late fusion strategies shown in Fig. 4. The early fusion strategy involves concatenation of input features whereas the model-level fusion involves concatenation of high-level feature representations from different SNs and the late multi-lingual fusion involves fusion of predictions from different SNs.

The model-level fusion strategy helps to complete the essence of hierarchical learning proposed in this work. Unlike early fusion, this strategy helps to circumvent the curse of dimensionality and synchronization between different features and at the same time

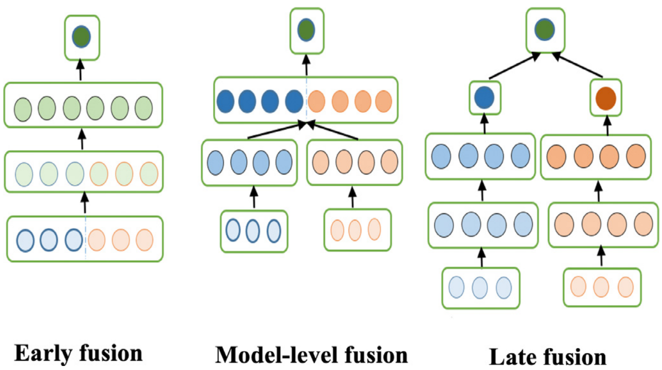

Fig. 4. Fusion strategies.

does not isolate interactions among different features as in late fusion. Finally, the fused representation is given to a convolution layer, which generates the final feature vector and uses a denser layer to process the feature vector. The output layer, i.e. SoftMax layer generates a regression output with linear activation to finally detect the mental state into one of the three categories, baseline, stress, and amusement.

3. Model

3.1. Data pre-processing

Each person have different affect even under the same conditions, they encounter different physiological changes. To have standardization, min-max normalization was applied on WESAD so that all the subjects have test results in the same scale range. Since the data collection is done over a period of 2 hours, in which a subject is presented with different stimuli, it results in different behaviour at different times. Having a time series data, with variation shown at each second using analog signals were converted to discrete forms by getting the bio-signals at each second. The sliding window of 0.25 seconds is used to better predict the mental state, as the effect of stimuli can happen over even a short interval of 2 seconds. On discretizing the continuous time frame data, if the change of stimuli started at one instance(second), but the major change in the biosignals is shown in the next one second interval then to observe the change, the previous instance values are required as well. Therefore to observe the changes in between two instances the sliding window of 0.25 second was used.

The target class is also given the numerical value, 1 for baseline, 2 for stress, and 3 for amusement; Baseline is a condition that a subject exists in the first 20 minutes of the experiment setup, where no external stimuli has been used. For all those scenarios, where target class is undefined is given the value of 0.

3.2. Architecture

After the data has been fetched in the scalable format, data is still astronomical and exists as individual units, so we apply the 1D-CNN over each feature signal by constructing a separate SSN. Once all the SSNs gets trained, the input is passed in as subject manner, i.e., all the features of a subject at one sliding window or at a particular time is passed through the SNs, and the output of the two SNs is then passed into the primary model that classifies the subject at a particular instance into one of the three classes defined during the training phase using Soft-max layer. As each subject produces different bio-signals while being in Stress, amusement or baseline, the convolution network while predicting can map a single instance into different classes having different probabilities, therefore application of logistic regression is very important to determine the final output class of the subject at a given instance. Due to 3 classes, Softmax function is applied on top layer to map the output with highest probability class. The proposed hierarchical network is shown in Fig. 5.

Table 2
Layers of the proposed hierarchical deep neural network.

High-level Feature Extraction	**CNN-SSN1 to CNN-SSN10 & CNN SN1 and CNN SN2**
	Input
	Convolution
	Pooling
	Convolution
	Pooling
	Dense
cClassification based Model-level fusion	Convolution
	Dense
	SoftMax

Each subject 's data contains signals from two different sensor devices, first RespiBAN, for monitoring chest signals ECG, TEMP, RESP, EDA, EMG, and ACC each recorded with 700 Hz signal, the SSNs created for each of these feature is 1-D CNN with input layer 700*1 except for ACC which is provided with input layer of 700*3, as the ACC signals contain 3 dimension data. The SSNs for features recorded with Empatica E4, have different sized input layers, for ACC 32* 3, BVP has 64*2, TEMP and EDA have input size of 4*1 input layer. 65% of data is used to train the model, i.e., ten subjects are used to train the model, and rest 35% i.e. 5 Subjects data is used to test the data, and 20% of the data (2 subjects) is used to validate the model. 53% of the total instances belong to the baseline class, 30% belongs to stress class, and 17% belong to the amusement class. The structure of different layers used in CNN sub-networks implemented in the proposed hierarchal network is shown in the Table 2.

4. Results & discussion

The proposed deep hierarchal neural network has been implemented on processed discrete data each of one second interval of continuous time series with sliding window of 0.25 seconds using 4-cross fold, each having 65% of data for training, and over 35% of total data, i.e. 5 subjects are used for testing the models. Model is then validated using 20% of the subjects, i.e. 3 subjects. And then model is evaluated in terms of accuracy and F1-Score, where,

$$\text{Precision} = \frac{TP}{TP+FP} \tag{1}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{2}$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

$$\text{F-1 Score} = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{4}$$

Confusion matrix was used to find the values of True Positive(TP), False Positive (FP), True Negative (TN), False Negative(FN) which was used to generate the evaluation matrices of the model. The sample confusion matrix is shown in Fig. 6 for Subject 15.

The average Precision, Recall, Accuracy and F-1 score of each individual subject after 4-cross fold is shown in Table 3. Highest precision of 0.9921 has been attained for Subject 10, and Recall of 0.9973 has been obtained for Subject 1. Best average F-1 score of

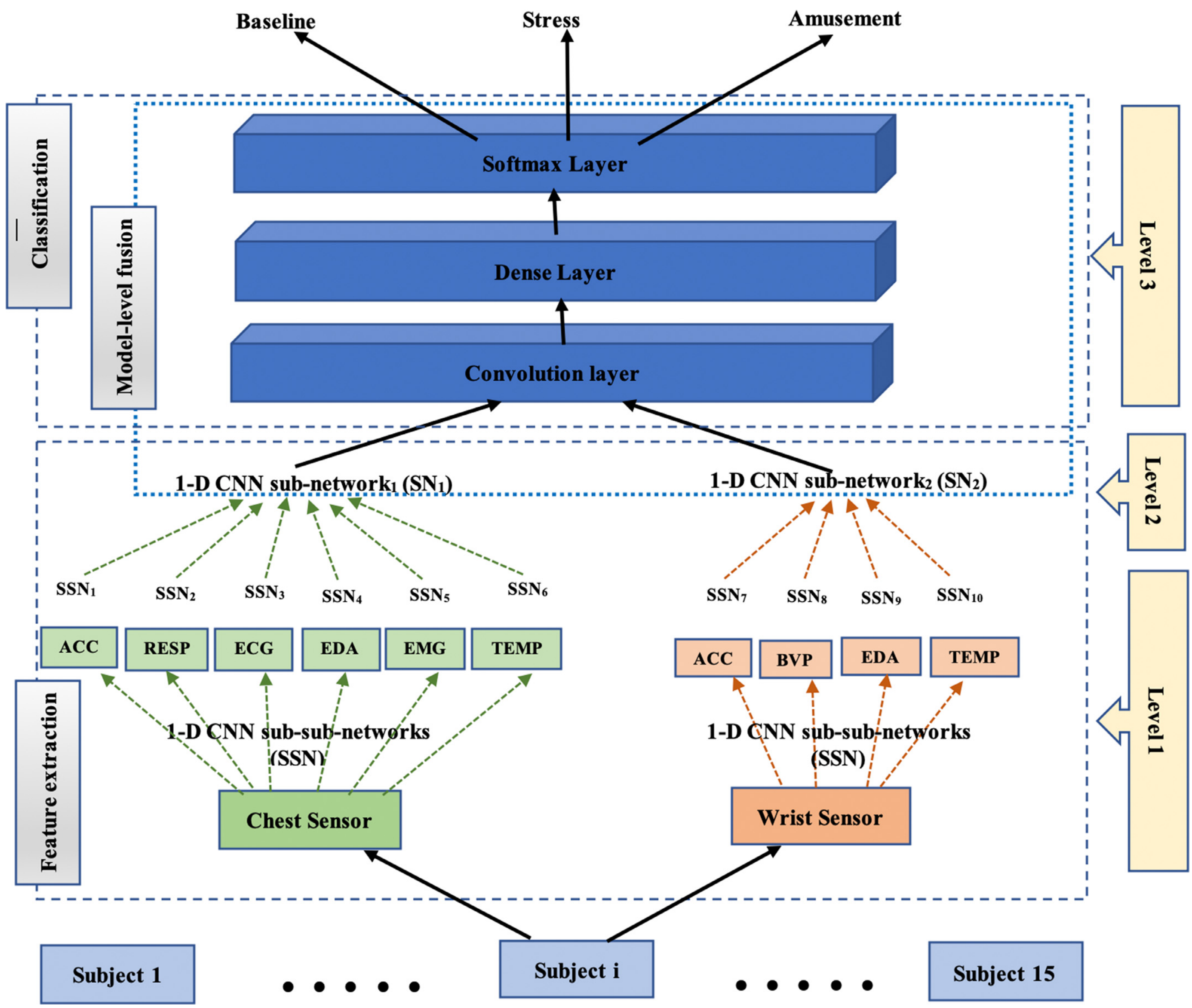

Fig. 5. Architecture of proposed deep hierearichal model.

	Baseline	Stress	Amusement
Baseline	0.9023	0.0412	0.0211
Stress	0.0415	0.9311	0.0112
Amusement	0.726	0.0036	0.8954

Fig. 6. Confusion matrix- subject 15.

Table 3
Results of proposed model.

F-1 Score	Precision	Recall	Accuracy	Subjects
0.9467	0.901	**0.9973**	93.39	**S1**
0.998	0.9474	0.9861	**96.98**	**S2**
0.612	0.826	0.9193	88.70	**S3**
0.9292	0.9079	0.9693	95.07	**S4**
0.96855	0.5518	0.9051	74.9	**S5**
0.8548	0.7709	0.9593	87.92	**S6**
0.8601	0.8353	0.8877	86.79	**S7**
0.6836	0.8761	0.5605	72.19	**S8**
0.798	0.8145	0.8756	87.24	**S9**
0.9348	**0.9921**	0.8831	92.73	**S10**
0.834	0.873	0.913	93.46	**S11**
0.8577	0.8222	0.8974	87.24	**S12**
0.93	0.92	0.961	88.72	**S13**
0.649	0.789	0.681	76.84	**S14**
0.917	0.95	0.925	93.56	**S15**

0.998 and accuracy of 96.98% for Subject 2 was produced by Subject 2. The accuracy curve of the model is shown in Fig. 7 . It is observed that the accuracy of the model varies from 72% to 96%. The average accuracy achieved by the proposed model is better than the state-of-the-art results. The F-score of the model is 0.8325, and the average accuracy is 87.7% which is better than the state-of-the-art results provided by Kumar et al. [8], Gjoreski et al. [9] although F-score generated by CNN with late fusion is more than the proposed model. The comparison of the deep hierarchal model with results of other models is shown in Table 4.

As validated from the results obtained, the accuracy of the model depends upon the subject's data, as different individuals have different level of stress for same scenario and the expression of stress is also different for each individual. Therefore to understand the affective state of a person, it is required to have per-

Table 4
Accuracy of different classifiers over WSEAD.

Classifier	Accuracy	F-score
Decision Tree [8]	0.64	0.58
Random Forest [8]	0.75	0.64
KNN [8]	0.56	0.48
LDA [8]	0.75	0.71
AdaBoost [8]	0.79	0.69
CNN [9]	0.85	0.86
Proposed Hierarchical Model	**0.877**	**0.83**

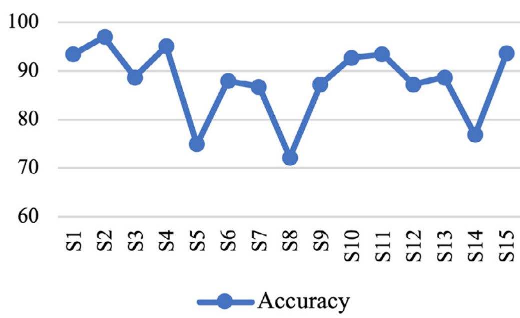

Fig. 7. Accuracy of subjects.

sonal characteristics of the subjects like medical history, traumas faced along with the continuous monitoring of the individual using IoT based wearable devices. The proposed model currently only takes the subject's bio-signals of 2 hours under controlled lab environment, this can further be extended to continuously analyse the data received from wearable IoT sensors.

5. Conclusion

Bio-signals are the biomarkers depicting these physiological changes as stress response symptoms during chronically activated situations. Only trained medical practitioners can measure such indicators, which can be tedious and time-consuming, thus delaying early identification and timely intervention. Persistent and chronic stress can lead to long-term health damage. It is imperative to design and develop an intelligent mental illness diagnostic that can support clinicians pro-actively. With the availability of smart sensors, the health condition of a person both physically and mentally can be tracked easily through IoT based wearable devices. This research proffered a deep hierarchal neural network model which, on receiving different sensor-based signals, categorizes the individual mental state to be either in stress or not or in amusement. A multivariate time series data, wearable stress and affects detection dataset, consisting of both wrist-based and chest-based sensor bio-signals, was trained using 3 level hierarchy of convolutional networks. The model used 13 CNN networks at different levels of the hierarchy that provided more efficient results than the state-of-the-art.

6. Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Declaration of Competing Interest

The authors certify that there is no conflict of interest in the subject matter discussed in this manuscript. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

References

[1] R. Dautov, S. Distefano, R. Buyya, Hierarchical data fusion for smart healthcare, J. Big Data 6 (1) (2019) 19.
[2] A. Kumar, Using cognition to resolve duplicacy issues in socially connected healthcare for smart cities, Comput. Commun. 152 (2020) 272–281.
[3] J. Bienertova-Vasku, P. Lenart, M. Scheringer, Eustress and distress: neither good nor bad, but rather the same? in: BioEssays, 2020, p. 1900238.
[4] S. Kumar, Biochemistry of the brain, Examining Biological Foundations of Human Behavior, 2020.
[5] A. Zenonos, A. Khan, G. Kalogridis, S. Vatsikas, T. Lewis, M. Sooriyabandara, Healthyoffice: mood recognition at work using smartphones and wearable sensors, in: In 2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops), pp. 1-6, IEEE, 2016.
[6] L. Tong, Q. Zhang, A. Sadka, L. Li, H. Zhou, Inverse boosting pruning trees for depression detection on twitter, 2019. arXiv:1906.00398.
[7] A. Esposito, A.M. Esposito, C. Vogel, Needs and challenges in human computer interaction for processing social emotional information, Pattern Recognit. Lett. 66 (2015) 41–51.
[8] A. Kumar, A. Sharma, A. Arora, Anxious depression prediction in real-time social data, in: In International Conference on Advances in Engineering Science Management & Technology (ICAESMT)-2019, Uttaranchal University, Dehradun, India, 2019.
[9] M. Gjoreski, M. Lutrek, M. Gams, H. Gjoreski, Monitoring stress with a wrist device using context, J. Biomed. Inf. 73 (2017) 159–170.
[10] S. Barra, M. Fraschini, A. Casanova, A. Castiglione, G. Fenu, PhysiounicaDB: a dataset of EEG and ECG simultaneously acquired, Pattern Recognit. Lett. 126 (2019) 119–122.
[11] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, K.V. Laerhoven, Introducing WESAD, a multimodal dataset for wearable stress and affect detection, in: Proceedings of the 20th ACM International Conference on Multimodal Interaction, 2018, pp. 400–408.
[12] J. Lin, S. Pan, C.S. Lee, S. Oviatt, An explainable deep fusion network for affect recognition using physiological signals, in: In Proceedings of the 28th ACM International Conference on Information and Knowledge Management, 2019, pp. 2069–2072.
[13] F.I. Indikawati, S. Winiarti, Stress detection from multimodal wearable sensor data, in: IOP Conference Series: Materials Science and Engineering, vol. 771, no. 1, p. 012028, IOP Publishing, 2020.
[14] R.W. Picard, E. Vyzas, J. Healey, Toward machine emotional intelligence: analysis of affective physiological state, IEEE Trans. Pattern Anal. Mach.Intell. 23 (10) (2001) 1175–1191.
[15] J. Healey, A. Jennifer, R.W. Picard, Detecting stress during real-world driving tasks using physiological sensors, IEEE Trans. Intell. Transp.Syst. 6 (2) (2005) 156–166.
[16] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, I. Patras, DEAP: a database for emotion analysis; using physiological signals, IEEE Tran. Affect. Comput. 3 (1) (2011) 18–31.
[17] A. Maxhuni, P. Hernandez-Leal, L.E. Sucar, V. Osmani, E.F. Morales, O. Mayora, Stress modelling and prediction in presence of scarce data, J. Biomed. Inf. 63 (2016) 344–356.
[18] A. Kumar, A. Jaiswal, Deep learning based sentiment classification on user-generated big data, Recent Adv. Comput. Sci.Commun. (Formerly: Recent Patents on Computer Science) 13 (5) (2020) 1047–1056.
[19] S.-C.B. Lo, H.-P. Chan, J.-S. Lin, H. Li, M.T. Freedman, S.K. Mun, Artificial convolution neural network for medical image pattern recognition, Neural Netw. 8 (7–8) (1995) 1201–1214.
[20] A. Kumar, K. Srinivasan, W.-H. Cheng, A.Y. Zomaya, Hybrid context enriched deep learning model for fine-grained sentiment analysis in textual and visual semiotic modality social data, Inf. Process. Manage. 57 (1) (2020) 102141.