International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
doi : https://doi.org/10.32628/IJSRST

Human Stress Monitoring System Using Convolution Neural Network

Prathiksha M[1], Arunachalam CV[1], Gokulakrishnan A[1], Ponraj V[1], Lakshmanan V[2]
[1]UG Scholar, [2]Assistant Professor,
Department of Computer Science and Engineering, Akshaya College of Engineering and Technology, Coimbatore, Tamil Nadu, India

Article Info
Volume 8, Issue 2
Page Number : 527-533

Publication Issue
March-April-2021

Article History
Accepted : 18 April 2021
Published : 24 April 2021

ABSTRACT

Psychological problems are becoming a major threat to people's life. Mental stress is a major issue nowadays, especially among youngsters and working people. The age that was considered once most carefree is now under a large amount of stress due to the surroundings. It is important to detect and manage stress before it turns into a severe health issue. Stress increase nowadays leads to many problems like depression, suicide, heart attack, and stroke. In this paper, the stress of the working people is being monitoring in their work environment. Image processing technique is used to monitor the stress of a person. The stress is identified by the face detection mechanism. In this face detection project, a computer system will be able to find and recognize human faces fast and precisely in images or videos that are being captured through a surveillance/web camera. It helps in conversion of the frames of the video into images so that the face of the student can be easily recognized for their face expressions. The dataset is taken from various people and their face expressions. These expressions are stored as values in XML files. These values are used to train the system using Convolution Neural Network algorithm. It predicts the results based on the values in the dataset. The resulting values will be stored in the text document. It will be helpful to their organisation counsellor for further future reference.

Keywords : Consumer Face Detection, Face Expressions, Stress Monitoring, Neural Network, Convolutional Neural Network.

I. INTRODUCTION

The technology aims in imparting a tremendous knowledge oriented technical innovation these days. Machine Learning is one among the interesting domain that enables the machine to train itself by providing some datasets as input and provides an appropriate output during testing by applying methods. Stress is our body's response to pressure. Many different situations or life events can cause stress. It is often triggered when we experience something new, unexpected or that threatens our

sense of self, or when we feel we have little control over a situation. The physical effects of prolonged stress are numerous, including a greater susceptibility to illness, a lack of energy, problems with sleep, headaches, poor judgment, weight gain, depression, anxiety, and a host of other ills. Stress is one of the mental disorders which can also be identified by facial expressions. It is a feeling of emotional or physical tension.

It can come from any event or thought that makes you feel frustrated, angry, or nervous. Stress is your body's reaction to a challenge or demand. In short bursts, stress can be positive, such as when it helps you avoid danger or meet a deadline. Not all stress is bad. But long-term stress can lead to health problems. Preventing and managing long-term stress can lower your risk for other conditions like heart disease, obesity, high blood pressure, and depression. "Prevention is better than cure"

In this paper, the stress of a person is identified and monitoring through facial expressions using the Convolutional Neural Network. The early detection of stress can help the person to get out of the stress easily. The early prevention of stress is to save a life of a person.

II. PROCESSES IN CNN

a. Dataset collection

Appropriate datasets are required at all stages of object recognition research, starting from training phase to evaluating the performance of recognition algorithms. All the images collected for the dataset were downloaded from the Internet, searched by name on various sources in different languages.

b. Image Preprocessing and Labelling

Images downloaded from the Internet were in various formats along with different resolutions and quality. In order to get better feature extraction, final images intended to be used as dataset for deep neural network classifier were preprocessed in order to gain consistency. Furthermore, procedure of image preprocessing involved cropping of all the images manually, in order to highlight the region of interest.

c. Augmentation Process

The main purpose of applying augmentation is to increase the dataset and introduce slight distortion to the images which helps in reducing overfitting during the training stage. Image data augmentation is a technique that can be used to artificially expand the size of a training dataset by creating modified versions of images in the dataset.

Training deep learning neural network models on more data can result in more skillful models, and the augmentation techniques can create variations of the images that can improve the ability of the fit models to generalize what they have learned to new images.

d. Neural Network Training

The main goal of training the network is for neural network to learn the features that distinguish one class from the others. Therefore, when using more augmented images, the chance for the network to learn the appropriate features has been increased.

e. Testing Trained Model with Valuation Data

Finally, the trained network is used to detect the disease by processing the input images in valuation dataset and results are processed.

III. Related Work

Multimodal Emotion Recognition Integrating Affective Speech with Facial Expression

Previous works are focused on eliciting results from unimodal systems. Machines used to predict emotion by only facial expressions or only vocal sounds. After a while, multimodal systems that use more than one features to predict emotion has more effective and gives more accurate results. So that, the combination of features such as audio-visual expressions, EEG, body gestures have been used since. More than one intelligent machine and neural networks are used to implement the emotion recognition system.

Multimodal recognition method has proven more effective than unimodal systems by Shiqing et al. Research has demonstrated that deep neural networks can effectively generate discriminative features that approximate the complex non-linear dependencies between features in the original set. These deep generative models have been applied to speech and language processing, as well as emotion recognition tasks. Martin et al. showed that bidirectional Long Short-Term Memory (BLSTM) network is more effective that conventional SVM approach.; In speech processing, Ngiam et al. proposed and evaluated deep networks to learn audio-visual features from spoken letters. In emotion recognition, Brueckner et al. found that the use of a Restricted Boltzmann Machine (RBM) prior to a two-layer neural network with fine-tuning could significantly improve classification accuracy in the Interspeech automatic likability classification challenge.

The work by Stuhlsatz et al. took a different approach for learning acoustic features in speech emotion recognition using Generalized Discriminant Analysis (GerDA) based on Deep Neural Networks (DNNs). Yelin et al. showed three layered Deep Belief Networks (DBNs') give better performance than two layered DBNs' by using audio visual emotion recognition process. Samira et al used Recurrent neural network combined with Convoluted Neural Network (CNN) in an underlying CNN-RNN architecture to predict emotion in the video. Some noble methods and techniques also enriched this particular research. They are more accurate, stable and realistic. In terms of performance, accuracy, reasonability and precision these methods are the dominating solutions. Some of them are more accurate but some are more realistic. Some take much time and require greater computation power to produce the more accurate result but some compromises accuracy over performance. The idea of being successful might differ but these solutions are the best possible till now.

Data driven framework to explore patterns (timings and durations) of emotion evidence, specific to individual emotion classes

In Previous work explore whether a subset of an utterance can be used for emotion inference and how the subset varies by classes of emotion and modalities. They propose a windowing method that identifies window configurations, window duration, and timing, for aggregating segment-level information for utterance-level emotion inference. The experimental results using the IEMOCAP and MSPIMPROV datasets show that the identified temporal window configurations demonstrate consistent patterns across speakers, specific to different classes of emotion and modalities. They compare their proposed windowing method to a baseline method that randomly selects window configurations and a traditional all-mean method that uses the full information within an utterance. This method shows a significantly higher performance in emotion recognition while the method only uses 40–80% of information within each utterance. The identified windows also show consistency across speakers, demonstrating how multimodal cues reveal emotion over time. These patterns also align with psychological findings. But after all achievement, the result is not consistent with this method.

Towards robust emotion recognition in the wild

A. Yao, D. Cai, P. Hu, S. Wang, L. Shan, and Y. Chen used a well-designed Convolutional Neural Network (CNN) architecture regarding the video-based emotion recognition. They proposed the method named as HOLONET has three critical considerations in network design.

(1) To reduce redundant filters and enhance the non-saturated non-linearity in the lower convolutional layers, they used modified Concatenated Rectified Linear Unit (CReLU) instead of ReLU.

(2) To enjoy the accuracy gain from considerably increased network depth and maintain efficiency,

they combine residual structure and CReLU to construct the middle layers.

3) To broaden network width and introduce multi-scale feature extraction property, the topper layers are designed as a variant of the inception-residual structure. This method more realistic than other methods here.

Semi-Supervised Learning (SSL) technique

Zixing Zhang, Fabien Ringeval, Eduardo Coutinho, Erik Marchi and Björn Schüller proposed some improvement in SSL technique to improve the low performance of a classifier that can deliver on challenging recognition tasks reduces the trust ability of the automatically labelled data and gave solutions regarding the noise accumulation problem - instances that are misclassified by the system are still used to train it in future iterations. they exploited the complementarity between audio-visual features to improve the performance of the classifier during the supervised phase.

Then, they iteratively re-evaluated the automatically labelled instances to correct possibly mislabelled data and this enhances the overall confidence of the system's predictions. This technique gives a best possible performance using SSL technique where labelled data is scarce and/or expensive to obtain but still, there are various inherent limitations that limit its performance in practical applications. This technique has been tested on a specific database with a limited type and number of data. The algorithm which has been used is not capable of processing physiological data alongside other types of data.

Personalizing EEG-Based Affective Models with Transfer Learning, Center for Brain-like Computing and Machine Intelligence

Wei-Long Zheng and Bao-Liang Lu proposed EEG-based effective models without labelled target data using transfer learning techniques (TCA-based Subject Transfer) which is very accurate in terms of positive emotion recognition than other techniques used before. Their method achieved 85.01% accuracy. They used to transfer learning and their method includes three pillars, TCA-based Subject Transfer, KPCA-based Subject Transfer and Transudative Parameter Transfer. For data pre-processing they used raw EEG signals processed with a bandpass filter between 1 Hz and 75 Hz and for feature extraction, they employed differential entropy (DE) features. For evaluation, they adopted a leave-one subject-out cross-validation method. Their experimental results demonstrated that the transductive parameter transfer approach significantly outperforms the other approaches in terms of the accuracies, and a 19.58% increase in recognition accuracy has been achieved.

IV. SYSTEM MODEL

Mental stress is a major issue nowadays, especially among youngsters. The age that was considered once most carefree is now under a large amount of stress. Stress increase now a days leads to many problems like depression, suicide, heart attack and stroke. In this paper, we are also calculating the mental stress of students one week before the exam and during the usage of the internet.

Our objective is to analyse stress of working people at different points in his life. The effect that exam pressure or recruitments stress has on the students often goes unnoticed. We will perform analysis on how these factors affect the mind of a student and will also correlate this stress with the time spent on the internet.

The dataset was taken from Jaypee Institute of Information Technology and it consisted of 206 student's data. Four classification algorithms Linear Regression, Naïve Bayes, Random Forest and SVM is applied and sensitivity, specificity and accuracy are used as performance parameter. The accuracy and performance of data are further enhanced by applying 10-Fold cross-validation.

The J48 algorithm, Sequential Minimal Optimization algorithm, Bayesian Network algorithm for predicting stress on the data collected from 16 peoples under four different stressful conditions. And also used HRV

features and EEG signal to predict the stress level. Various features like HRV, heart rate, ECG are used to predict the stress level. Then, authors used decision tree algorithm is applied on a dataset collected from two test completed that these tests to be unsatisfactory.

V. PROPOED SYSTEM IMPLEMENTATION

The proposed system aims at creating a digital way of identifying the stressed people in the work place to avoid the adverse effect of stress. Human Stress Monitoring System is a digital way that would enable us to find the facial expression and identifies the stressed persons saving lots of time. The Convolutional Neural Network (CNN) detects and monitoring the facial expression of a person through camera module. It helps to reduce the unnecessary counselling process for the identification of stressed people, In the initial phase, the developer creates the dataset to train the system. Once the dataset is created, the system will be trained based on the dataset. After completing the training, the system will start to detect and monitor the facial expression of a person and converting them into values to find the stressed level of that person. Finally, the results will be stored in the text document for further reference.

a. Convolutional Neural Networks Working

There are four layered concepts we should understand in Convolutional Neural Networks:

1. Convolution,
2. ReLu,
3. Pooling and
4. Full Connectedness (Fully Connected Layer).

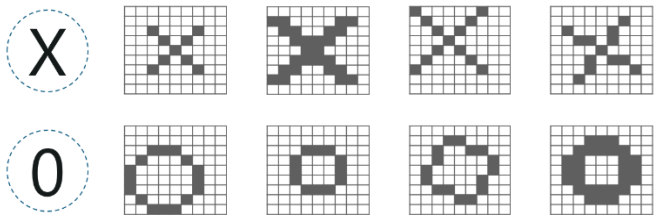

Figure 1 Convolution Neural Network Example

Here, there are multiple renditions of X and O's. This makes it tricky for the computer to recognize. But the goal is that if the input signal looks like previous images it has seen before, the "image" reference signal will be mixed into, or convolved with, the input signal. The resulting output signal is then passed on to the next layer.

-1	-1	-1	-1	-1	-1	-1	-1	-1
-1	1	-1	-1	-1	-1	-1	1	-1
-1	-1	1	-1	-1	-1	1	-1	-1
-1	-1	-1	1	-1	1	-1	-1	-1
-1	-1	-1	-1	1	-1	-1	-1	-1
-1	-1	-1	1	-1	1	-1	-1	-1
-1	-1	1	-1	-1	-1	1	-1	-1
-1	1	-1	-1	-1	-1	-1	1	-1
-1	-1	-1	-1	-1	-1	-1	-1	-1

Figure 2 Computer Understanding of Pixels

So, the computer understands every pixel. In this case, the white pixels are said to be -1 while the black ones are 1. This is just the way we've implemented to differentiate the pixels in a basic binary classification.

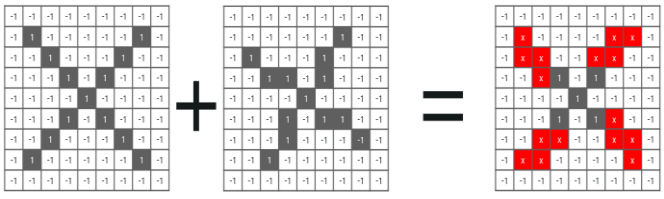

Figure 3 Adding the pixels

Now if we would just normally search and compare the values between a normal image and another 'x' rendition, we would get a lot of missing pixels.

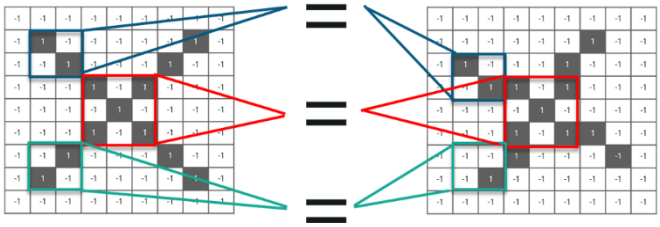

Figure 4. Mapping the pixels

We take small patches of the pixels called filters and try to match them in the corresponding nearby locations to see if we get a match. By doing this, the Convolutional Neural Network gets a lot better at seeing similarity than directly trying to match the entire image.

b. Convolution of an Image

Convolution has the nice property of being translational invariant. Intuitively, this means that each convolution filter represents a feature of interest (e.g. pixels in letters) and the Convolutional Neural Network algorithm learns which features comprise the resulting reference (i.e. alphabet).

There are 4 steps for convolution:

- Line up the feature and the image
- Multiply each image pixel by corresponding feature pixel
- Add the values and find the sum
- Divide the sum by the total number of pixels in the feature

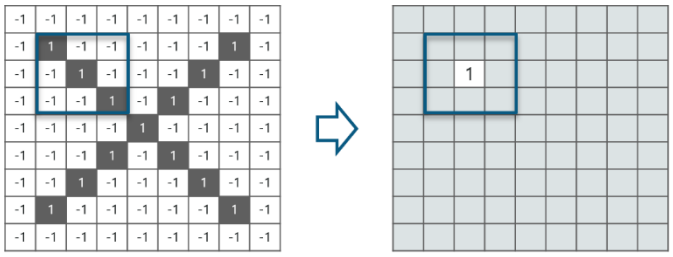

Figure 5: Filtered image example

Now, we can move this filter around and do the same at any pixel in the image.

ReLU Layer

Rectified Linear Unit (ReLU) transform function only activates a node if the input is above a certain quantity, while the input is below zero, the output is zero, but when the input rises above a certain threshold, it has a linear relationship with the dependent variable.

Face detection and Prediction Phase:

The final phase of our system is prediction. It predicts the output based on the dataset values by detecting the face of a person. The results will be emotions of a person who is in front of the camera. The resulted outputs will be stored as a text document.

VI. CONCLUSION

Thus, the Human Stress Monitoring System has gained a greater attraction in most of the fields. The system is designed to overcome the issues in existing similar systems. Core part of the system was achieved using extraction of the facial expressions. The face detection mechanism played the major role in this system. After extraction of facial expression, these values will be stored as a text document for further reference. It will be helpful to counselors to take the necessary preventive measures for the stressed persons. In the proposed system, data sheet in the XML file format is used to identify the facial expressions of a person by fixed values instead of that, adaptive learning method can be used to train the model. We can suggest the remedies to the persons who are stressed, to prevent them from taking any unstable decision. The facial expressions of a person can be monitored and the values are then exported into Excel sheet as report for further future reference. An additional module of sending a text message to his/her organization counselor can be added to the system to make it even more effective. The text message system can be automated to send messages at regular intervals of time.

VII. REFERENCES

[1]. Shiqing Zhang, Xiaohu Wang, Gang Zhang, Xiaoming Zhao; Multimodal Emotion Recognition Integrating Affective Speech with Facial Expression; Institute of Image Processing and Pattern Recognition Taizhou University Taizhou 318000 CHINA, Hunan Institute of Technology Hengyang 421002 CHINA, Bay Area Compliance Labs. Corp. Shenzhen 518000 CHINA, 2014.

[2]. Yelin Kim and Emily Mower Provos, Data driven framework to explore patterns (timings and durations) of emotion evidence, specific to individual emotion classes; University of Michigan Electrical Engineering and Computer Science, Ann Arbor, Michigan, USA;2016.

[3]. A. Yao, D. Cai, P. Hu, S. Wang, L. Shan and Y. Chen; HoloNet: Towards robust emotion recognition in the wild, 2016.

[4]. Zixing Zhang, Fabien Ringeval, Fabien Ringeval, Eduardo Coutinho, Erik Marchi and Björn Schüller, Semi-Supervised Learning (SSL) technique.

[5]. Wei-Long Zheng1 and Bao-Liang Lu, Personalizing EEG-Based Affective Models with Transfer Learning, Center for Brain-like Computing and Machine Intelligence, Department of Computer Science and Engineering, Key Laboratory of Shanghai Education Commission for Intelligent Interaction and Cognitive Engineering, Brain Science and Technology Research Center, Shanghai Jiao Tong University, Shanghai, China. 2016.

[6]. Kulkarni, N., Phalle, S., Desale, M., Gokhale, N. and Kasture, K., 2020. A Review on EEG Based Stress Monitoring System Using Deep Learning Approach. Mukt Shabd Journal, 9(6), pp.1317-1325.

[7]. Li, R. and Liu, Z., 2020. Stress detection using deep neural networks. BMC Medical Informatics and Decision Making, 20(11), pp.1-10.

[8]. Hammad, M., Zhang, S. and Wang, K., 2019. A novel two-dimensional ECG feature extraction and classification algorithm based on convolution neural network for human authentication. Future Generation Computer Systems, 101, pp.180-196.

[9]. Hwang, B., Ryu, J.W., Park, C. and Zhang, B.T., 2017, July. A novel method to monitor human stress states using ultra-short-term ECG spectral feature. In 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 2381-2384). IEEE.

[10]. Nath, R.K., Thapliyal, H., Caban-Holt, A. and Mohanty, S.P., 2020. Machine learning based solutions for real-time stress monitoring. IEEE Consumer Electronics Magazine, 9(5), pp.34-41.

[11]. Dahmane, M., Alam, J., St-Charles, P.L., Lalonde, M., Heffner, K. and Foucher, S., 2020. A Multimodal Non-Intrusive Stress Monitoring from the Pleasure-Arousal Emotional Dimensions. IEEE Transactions on Affective Computing.

[12]. Lee, J.H., Gamper, H., Tashev, I., Dong, S., Ma, S., Remaley, J., Holbery, J.D. and Yoon, S.H., 2020, April. Stress monitoring using multimodal bio-sensing headset. In Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1-7).

[13]. Shan, Y., Li, S. and Chen, T., 2020. Respiratory signal and human stress: Non-contact detection of stress with a low-cost depth sensing camera. International Journal of Machine Learning and Cybernetics, 11(8), pp.1825-1837.

[14]. Lim, H.J. and Sohn, H., 2020. Online Stress Monitoring Technique Based on Lamb-wave Measurements and a Convolutional Neural Network Under Static and Dynamic Loadings. Experimental Mechanics, 60(2), pp.171-179.

Cite this article as :

Prathiksha M, Arunachalam CV, Gokulakrishnan A, Ponraj V, Lakshmanan V, "Human Stress Monitoring System Using Convolution Neural Network ", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 8 Issue 2, pp. 527-533, March-April 2021.
Journal URL : https://ijsrst.com/IJSRST2182104