Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 3, 2021, Pages. 8364 - 8374
Received 16 February 2021; Accepted 08 March 2021.

Emotional and Physical Stress Detection and Classification Using Thermal Imaging Technique

R. Reshma[1]

[1]Department of Medical Electronics,Sengunthar College of Engineering, Tiruchengode,Tamilnadu, India.
E-mail: reshmapraveen027@gmail.com

ABSTRACT

Introduction: Stress has become a significant issue in the current society anddamaging human life. This is because of work, family, personal pressures as well as surrounding environments. Image-based stress detection is emerging with the help of computer-aided systems.

Background: Here we proposed ahybrid deep learning technique to detect and analyze the effects of stress. The combination of these two deep neural networks that are Alexnet and Vgg-16 are pre-trained, The main advantages are the input of the Vgg-16 networks is feature maps.

Methods: The first network (DNN-1) input is the raw image, and the second network (DNN-2) input is the Frequency features which are created by wavelet transform. The frequency features help to differentiate the frequency-related information from the images. Finally, these two networks are combined for better classification results. Both the network suitable for detecting stress but the combined network provides better performance in terms of accuracy because the features have been extracted from both the network and combined with its mean value so both features are available for the classification. it will boost the detection results.

Results: Also, we experiment with Alexnet and VGG-16 and combined network separately. the combined or hybrid network provides 96.2 % accuracy which is higher than the separate network and existing machine learning techniques like SVM and KNN as well conventional deep learning techniques.

Conclusion: This method can be used in health care systems for identifying the human stress for further treatments.

KEYWORDS

Convolutional Neural Network (CNN), Support Vector Machine (SVM), Classifier, Deep Neural Network (DNN), Rectified Linear Unit (ReLu), Dropout.

Introduction

In the modern world, stress has become a part and parcel of livelihood. It has widespread more and more in this digital era. If stress is not monitored carefully or controlled, it can lead to numerous health issues, undermining people's sentiments, emotions, practices, and prosperity. Having the option to identify stress can help individuals find multiple ways to deal with stress before unwanted consequences are brought. Classical stress detection strategies depend on psychological questionnaires [1] or consultation [2]. Since the questionnaire depends greatly on the opinion given by people, the stress measure is dependent on the perception of the individual. At the point when individuals decide to communicate their mental states with reservations, the outcome scale would be one-sided. To outsmart the constraints of the questionnaire studies, the techniques for automatically finding stress by detecting the proactive tasks through wearable gadgets i.e. cell phones equipped with sensors [3–8] or dependent on physiological signals for eg. heart rate variation, Electrocardiogram (ECG), Galvanic Skin Response (GSR), high blood pressure, Electromyogram (EMG), Electroencephalogram (EEG), and so on from appropriate sensors [9–12] have been designed. While these techniques can detect the stress states of the people, they normally request wearable gadgets and sensors that could hardly acknowledge contact-free detection. As of now, the pervasive deployment of contactless camcorders in the environment, along with the faster advancement of information gathering and analysis methods, offers us another way to identify stress using image sequences acquired from a monitoring camcorder. In comparison to the sensory gadgets, the later provides three benefits. To start with, it is more advantageous, especially in places like schools, clinics, and limited and jails, where no portable gadgets are required or permitted. Second, it has an exceptionally long reserve time and can undoubtedly contact the mass crowd easily. Third, the successive frames help to analyze stress states without any artificial traits and factors.

Literature Review

There are a few ongoing studies and findings that reveal facial cues and articulations can give knowledge about stress

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 3, 2021, Pages. 8364 - 8374
Received 16 February 2021; Accepted 08 March 2021.

identification [13–15], and the basic symptoms related to the changes in the physiological signs (e.g., heart pulse, blood pressure, GSR, and so on) and proactive tasks [16]. A larger part of the current literature work is focused on identifying facial cues for eg. mouth action, head movement, pulse, flicker rate, gaze spatial distribution, and eye movements from various facial areas [13,17,18], or using the Facial Action Coding System (FACS) [19] and taking Action Units (AUs) from the face outlines for stress identification [20,21]. Deep learning has been broadly and effectively applied in numerous fields like computer vision, emotion identification, etc.

Image-based Stress Detection

Seeing that the indications of stress could be all the more effortlessly distinguished by taking a glimpse of the state of the face, especially the lines or wrinkles around the nose, mouth, and eyes, [22,23] researched three facial parts (the eyes, nose and mouth) that are important for stress detection. The Gabor filter and Histogram of Oriented Gradients (HOG) features [23] from each portion of the face in pixels using visual image encoding technique, and gave them into three various SVM classifiers. The individual results are fed into slant binary tree to get the final outcome. Experimental validation were performed on JAFFE dataset, where each patient has a stress expression and a neutral expression image [24]. The result shows that the nose is a portion of the face that eventually indicates stress, and about 86.7% of detection accuracy is achieved. Along a similar line, relevant facial features [22] are extracted from the pixel using Difference of Gaussians (DoG), HOG, and Discrete Wavelet Transform (DWT) histogram methods, and then fused and reconstructed the resulting multi-histogram features into global features. A Convolutional Neural Network (CNN) with three convolutional layers and two max-pooling layers was trained on the color FERET face database. The stress recognition accuracy attained about 95%. As the stress symptoms are normally related with variations in physiological (e.g., heart rate, blood pressure, galvanic skin response, etc.) and proactive tasks [16], such facial features like gaze spatial distribution, eye movement, pupil dilation, and filicker rate and so on. It is used to recognize stress levels. The authors [25] identified stress and anxiety using a set of facial cues, including mouth activity, head motion, heart rate, blink rate, and eye movements. Techniques utilized for extracting these features from various facial regions are elaborately discussed and the performance was evaluated on a dataset consisting of 23 subjects.

Thermal Image-based Stress Detection

Stress could be effectively identified from thermal imaging because of changes in skin temperature due to stress [26]. As the emerging utilization of both thermal spectrum (TS) and visible spectrum (VS) imaging, the formulation, diagnosis and perceiving facial cues have become a lot more easier. A stress identification method is proposed [27–34] by fusing visual and thermal spectrums of spatio-temporal facial information. The ANUStressDB database consisting of videos of 35 patients watching stressful and non-stressful movie clips was used. It utilized a hybrid Genetic Algorithm (GA) and Support Vector Machine (SVM) to identify notable divisions of facial block regions and choose whether utilizing the block regions can improve the performance of stress identification. The test results showed that visible improvement of stress detection performance after the fusion of facial textures from VS and TS videos compared to the individual videos. Besides the GA selection method gave improved performance in contrast to the facial block divisions. The HDTP (dynamic thermal patterns in histograms) features fused with LBP-TOP (local binary patterns on three orthogonal planes) features for TS and VS videos utilizing a hybrid GA and SVM achieves 86% accuracy. Further extension [35] is towards the representation of a thermal image as a class of super-pixels, and feature extraction from them instead of extracting directly from the pixels [34]. As per [36], super-pixel representation is utilized for face recognition. Moreover, a thermal super-pixel is hence a class of pixels with same color (temperature) which appears as though a more common representation for thermal images as compared to partioning images into non-overlapping blocks. Thus with greatly correlated nearby pixels grouped, the stress recognition can be improved and the execution speed can be increased. The test results on ANUstressDB database showed that the technique performed better [34], achieving a 89% classification accuracy.

Materials & Method

Network Architecture

The convolutional neural network is commonly made out of the convolutional layer, Rectified Linear Unit (ReLU),

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 3, 2021, Pages. 8364 - 8374
Received 16 February 2021; Accepted 08 March 2021.

and pooling layers. Convolutional layers which are the structure block are used to extract the low, medium, high level features from the images these features very helpful to differentiate the images. The ReLU layers help to allow only the effective feature map data to the following stage. It gives non-linearity to the feature maps. The maximum pooling layer is utilized to limit the computational complexity. Figure 1 shows,

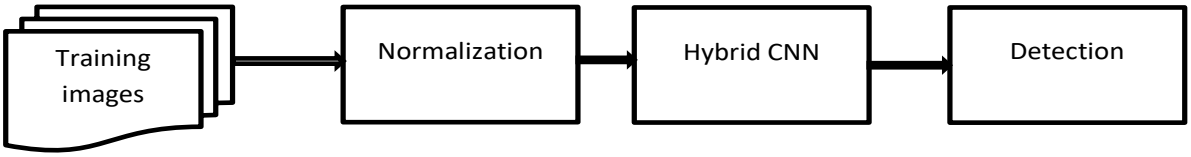

Figure1. Proposed Architecture

Face Detection &Z-Score Image Normalization

Face has to be detected from the input thermal images. Viola jones face detector used here.This viola jones face detector working based on haar like features and cascade classifiers.Image normalization is important processing in the deep learning training process. because the training needs thousand of images so all the images should be normalized for better training, we used the Z-score normalization which means the normalization carried out based on mean and standard deviation values so all the image pixel values come under n a specific range Figure 2 and Figure 3 represents the normal or healthy samples respectively. Figure 4 and Figure 5 represents the original data and normalized data samples.

Figure2. Themal image– Normal; **Figure3.**Themal image -Stress samples

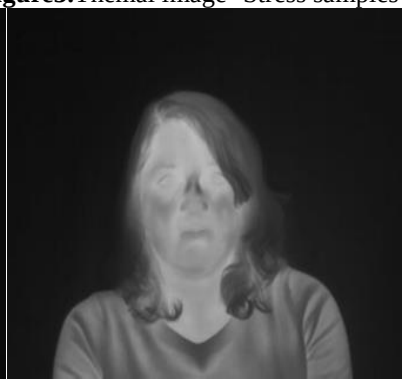

Figure 4. Normalized – Normal; **Figure5.**Normalized - Stress samples

Deep Network Layers

1) Convolutional Layer

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 3, 2021, Pages. 8364 - 8374
Received 16 February 2021; Accepted 08 March 2021.

The basic building blocks of CNN are the convolution filters. These filters are capable to learn features specific to the input image and its corresponding outcome. The resulting feature maps are moved on to the upcoming layers. This convolution has the property of translational invariant that helps to identify the multiple features in the input image.

2) Max Pooling

Pooling is used to reduce the filter's sensitivity to noise and other illumination effects. It does the task of subsampling and smooths the image by averaging or taking the maximum over a masked region.

3) Rectified Linear Unit

The activation function controls the firing of neurons in CNN to learns specific features. Finally, the fully connected network encapsulates the hardwired neural network to perform the classification or segmentation tasks. In this layer, some non-linearities are added to make it more adaptable to the real-world case. it is defined as

$$Relu=Max(0,a) \quad (1)$$

Hybrid Network

In this proposed method we combine the two different types of deep neural networks to analyze both the spatial and frequency information of the face images. So there is less chance of missing the important features to classify the normal and stress images. The combined these two deep neural networks are custom-designed as well pre-trained, The main advantages is the input of both the networks are feature maps, the first network (DNN-1) input is the raw image, and the second network (DNN-2) input is the Frequency features which is created by wavelet transform. The frequency features help to differentiate the frequency-related information from the images. Finally, these two networks are combined for better classification results, The proposed architecture is shown in figure 1.

1) Alexnet Architecture – CNN-1

Alexnet consists of eight layers in which there are 5 convolution layers and 3 fully connected layers. After all the convolution layers, the ReLU layer is applied to introduce non-linearity into the system and it finally classifies using a fully connected layer. In each convolution layer, there is a group of kernels of the identical size. And the first convolution layer has a kernel size of 11×11 with 96 kernels that are of similar depth of input i.e. 3. Convolution layers are the backbone of any CNN, so it uses 6% of the entire parameters and requires 95% of the calculation. Before the two final fully connected layers, dropout is added to improve its computation efficiency. The input image size of the architecture is 227× 227.

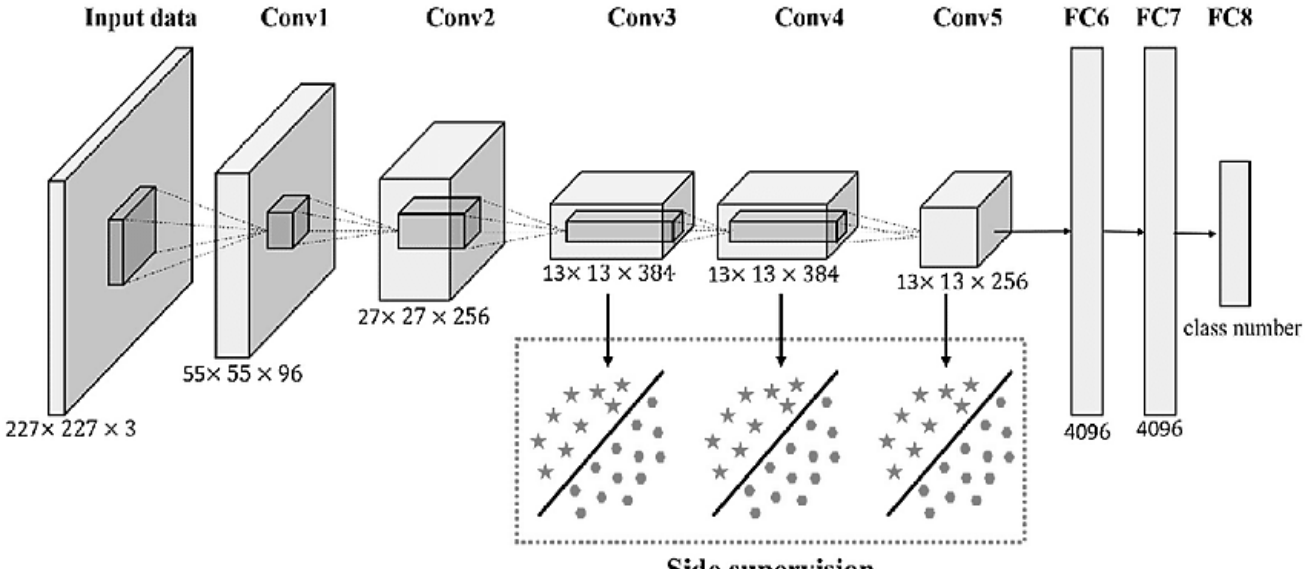

Figure 6. Alexnet Architecture

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 3, 2021, Pages. 8364 - 8374
Received 16 February 2021; Accepted 08 March 2021.

2) Vgg-16 Architecture- CNN-2

TheVgg-16 has sixteen convolutional layers [23]. The number of filters for the first two convolutional layers is 64 and the third and fourth convolutional layer is 128, after this fifth, sixth, seventh convolutional layers number of a filter is 256, then the remaining final 6 layers number of a filter is 512 respectively. It has three fully-connected layers and the final fully connected layer indicated the number of classes required for classification. and all the filter kernel size is 3x3 and Gaussian kernels. Total of four max-pooling layers used to reduce the feature map dimensions. the following figure represents the structure of the VGG-16.it is shown in figure 6.

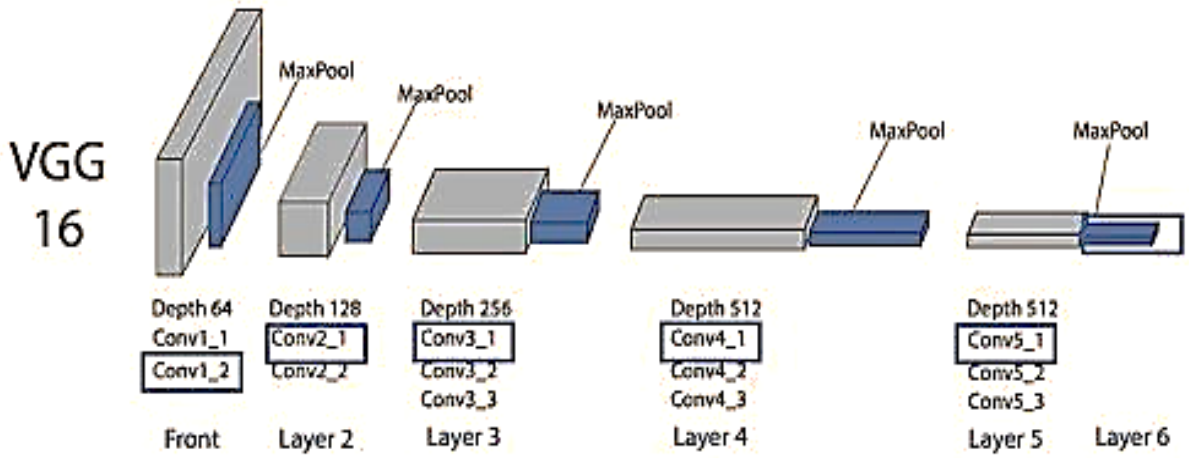

Figure 7. VGG -16 Network Architecture

3)Loss Function L

In order to recognize the class labels, we need to reduce the loss function. The loss function we used here is cross-entropy.
Minimizing the cross-entropy loss function that is defined as

$$L(w)=\sum_{i=1}^{N}\sum_{c=1}^{4}-y_{ic}\ log f_c(x_i)+\in \|w\|_2^2 (2)$$

$f_c(x_i)$ − predicted probability of class c for image x

4)Learning rate (η)

The three most common techniques used for shrinking the learning rate during training are categorized into constant, factored, and exponential decay. As a first step, a constant Ƭ is identified as a constant that can be applied to decrease the learning rate manually with a clear step function. Next, the learning rate can be attuned during training with the following equation:

$$\eta t = \eta 0 \beta^{\frac{t}{\epsilon}} (3)$$
$$\eta 0 = 0.001$$

Where ηt is the learning rate, $\eta 0$ is the initial learning rate, and β is the decay factor with the range of (0,1).

5)Training Algorithm

Input*: Training data samples Tdswith no. ofimage class labels nI, Validation setsVds, Learning Rate Lr, Optimizer*
Output- *Feature map, performance parameters such as accuracy, sensitivity, specificity*
1, Train the network with Adam optimizer to reduce the cross-entropy loss

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 3, 2021, Pages. 8364 - 8374
Received 16 February 2021; Accepted 08 March 2021.

2. Calculate Loss on training data set Tds and validation data sets Vdsby backpropagation based on the correct class labels
3.Initialize the learning Rate Lr
4.For X= 1 to no of images nI
5.Takethe feature map for the last convolutional layer
6.Find softmax of the ouputlayer
7.Return Loss, Accuracy, probabilistic feature map

Network FusionStrategy

The fusion of networkshelps to increase the receptive field in terms of feature vectors so that both learned networks share their receptive fields to get the best receptive field, finally both are combined by the meanmethod.

Result and Discussion

Data Augmentation

Here we have taken a total of 250 sample images it has 120 normal and 150stress images, for training we have taken 250,2400 images respectively by rotation and translation, resized all the images are resized to 256x256. And the training loss calculated by entropy loss and the training method is Adam optimizer, then the 200 epochs given and iteration nearly 100 and we got the training accuracy 100% and training loss 0.001, then we tested the total 250 images and we got 95.1 % accuracy.

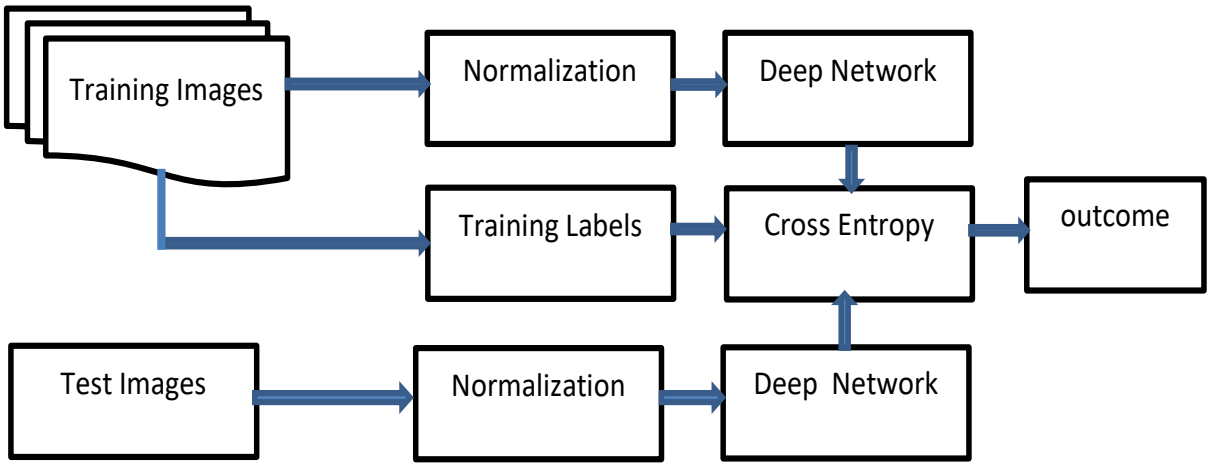

Figure 8. The Experimental setup

The experimental results are shown in figure 8. It represents the setup for both training and testing data. The performance calculation formulated below
Normal: Normal sample
Abnormal: Stress sample
True positive (Tp) =No of identified Normal sample correctly
False positive (Fp) = No of identified Stresssample correctly
True negative (Tn) = No of identified Normalsample as Stress sample
False negative (Fn) = No of identified Stress sample as Normalsample

$$\textbf{Accuracy} = (Tp + Tn)/(Tp + Tn + Fp + Fn) \quad (4)$$
$$\textbf{Sensitivity} = Tp/(Tp + Fn) \quad (5)$$
$$\textbf{Specificity} = Tn/(Tn + Fp) \quad (6)$$

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 3, 2021, Pages. 8364 - 8374
Received 16 February 2021; Accepted 08 March 2021.

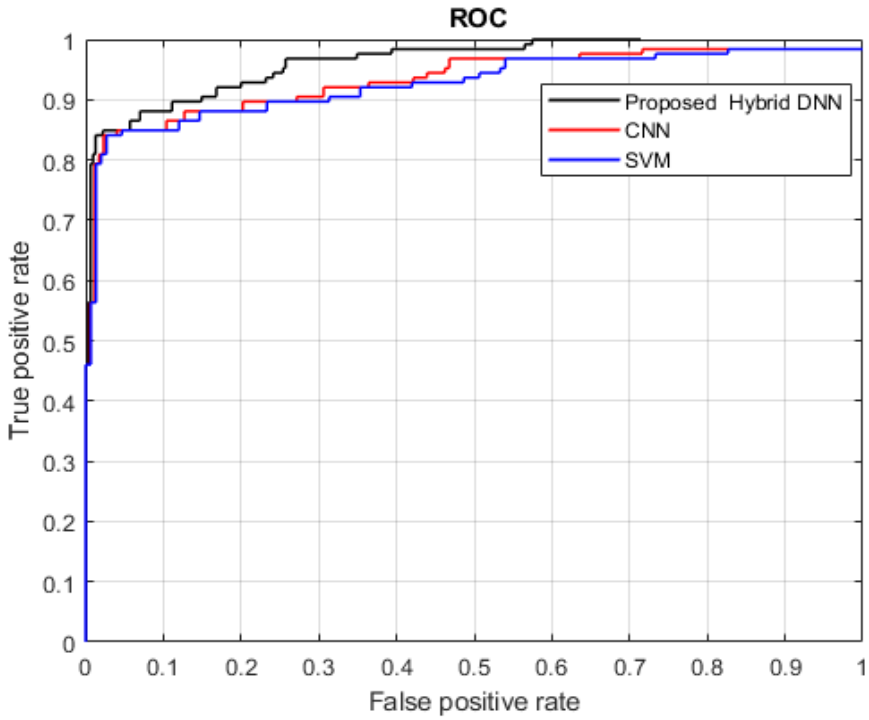

Figure9. ROC ofproposed vs CNN vs SVM

Figure.9 shows the ROC curve for the proposed system and the AUC for this systemis 96.8 which is higher than conventional CNN and SVM,

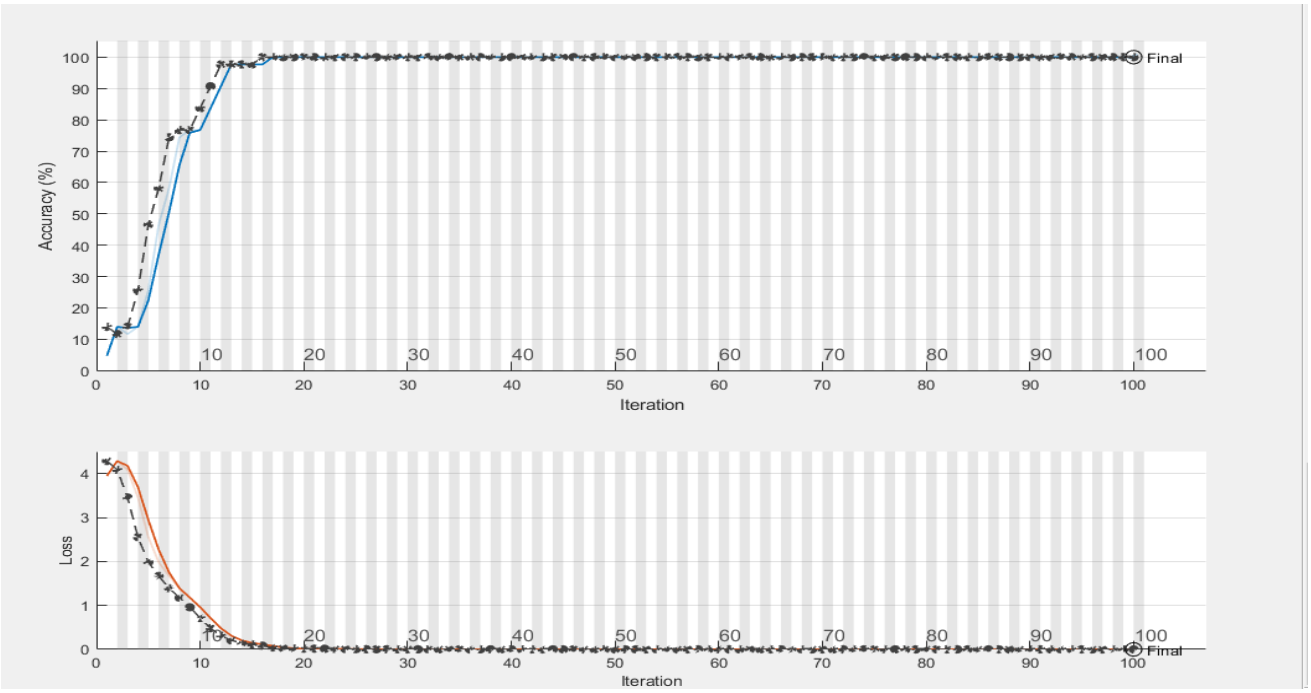

Figure10. Training loss andTraining Accuracy

The training accuracy versus training error loss is represented in figure 10, at the lowest epoch it reached the

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 3, 2021, Pages. 8364 - 8374
Received 16 February 2021; Accepted 08 March 2021.

maximum accuracy which indicates the good training.

Table1. Performance of the Proposed Systemwithout dropout layer

Method	Accuracy	Sensitivity	Specificity
Alexnet	91	93	93
Vgg-16	93	94	94
Proposed (Combined)	**95.1**	**95.6**	**96.5**

Table2. Performance of the Proposed Systemcompared alexnet

Method	Accuracy	Sensitivity	Specificity
Alexnet [26]	91	92	94.5
Vgg-16	93	94	95
Proposed (Combined)	**96.2**	**96.5**	**97.1**

Table3. Performance of the proposed systemwith Dropout-0.5

Method	Accuracy	Sensitivity	Specificity
Alexnt	92	92.3	92
Vgg-16	94.5	94.7	95.7
Proposed (Combined)	**96.2**	**96.5**	**97.1**

Table 1 represents the performance of the proposed system compared with transfer learning Alexnte and vgg-16 without using the dropout layer. Sothe performance little degraded without using the drop-out layer compared to table 3. The alex network comparison is tabulated in table 2.

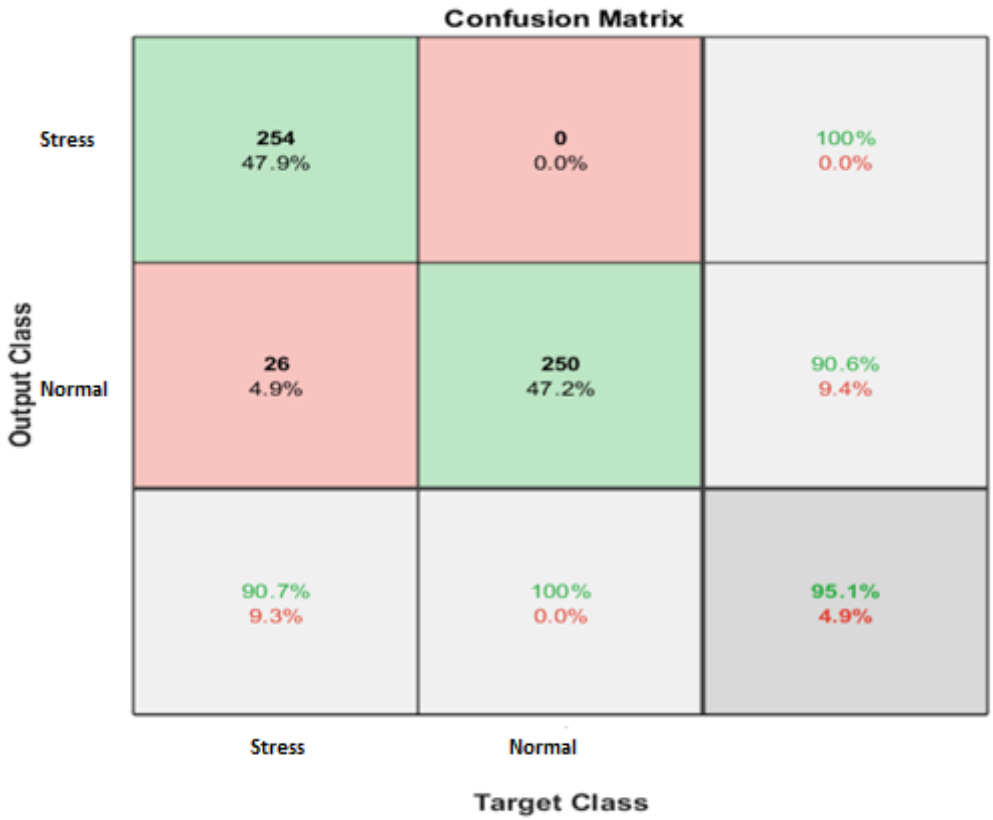

Figure 11. Confusion matrix

Figure 11 represents the confusion matrix, it is used to how the normal and stress data samples perform each other.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 3, 2021, Pages. 8364 - 8374
Received 16 February 2021; Accepted 08 March 2021.

Conclusion and Future Scope

In this work, the hybrid deep learning network is used to identify the stress in the human-based on face images. The two different transfer learning network that is alexnet and VGG-16 has taken and fused for better detection accuracy. Both the network suitable for detecting stress but the combined network provides better performance in terms of accuracy because the features have been extracted from both the network and combined with its mean value so both features are available for the classification. it will boost the detection results. Also, we experiment withalexnet and VGG-16 and combined network separately. the combined or hybrid network provides 96.2 % accuracy which is higher than the separate network and existing machine learning techniques like SVM and KNN.In the future, we can use better feature-based images instead of raw images for better performance.

References

[1] Cohen, S., Kamarck, T., &Mermelstein, R. (1983). A global measure of perceived stress. *Journal of health and social behaviour*, 385-396.

[2] Dupéré, V., Dion, E., Harkness, K., McCabe, J., Thouin, É., & Parent, S. (2017). Adaptation and validation of the Life Events and Difficulties Schedule for use with high school dropouts. *Journal of Research on Adolescence*, *27*(3), 683-689.

[3] Lee, B.G., & Chung, W.Y. (2016). Wearable glove-type driver stress detection using a motion sensor. *IEEE Transactions on Intelligent Transportation Systems*, *18*(7), 1835-1844.

[4] Ciabattoni, L., Ferracuti, F., Longhi, S., Pepa, L., Romeo, L., &Verdini, F. (2017). Real-time mental stress detection based on smartwatch. *In IEEE International Conference on Consumer Electronics (ICCE)*, 110-111.

[5] Han, H., Byun, K., & Kang, H.G. (2018). A deep learning-based stress detection algorithm with speech signal. *In proceedings of the workshop on audio-visual scene understanding for immersive multimedia*, 11-15.

[6] Yogesh, C.K., Hariharan, M., Yuvaraj, R., Ngadiran, R., Yaacob, S., &Polat, K. (2017). Bispectral features and mean shift clustering for stress and emotion recognition from natural speech. *Computers & Electrical Engineering*, *62*, 676-691.

[7] Prasetio, B.H., Tamura, H., &Tanno, K. (2018). Ensemble support vector machine and neural network method for speech stress recognition. *In International Workshop on Big Data and Information Security (IWBIS)*, 57-62.

[8] Harari, G.M., Gosling, S.D., Wang, R., Chen, F., Chen, Z., & Campbell, A.T. (2017). Patterns of behavior change in students over an academic term: A preliminary study of activity and sociability behaviors using smartphone sensing methods. *Computers in Human Behavior*, *67*, 129-138.

[9] Chow, L., Bambos, N., Gilman, A., &Chander, A. (2014). Personalized monitors for real-time detection of physiological states. *International Journal of E-Health and Medical Communications (IJEHMC)*, *5*(4), 1-19.

[10] Cinaz, B., Arnrich, B., La Marca, R., &Tröster, G. (2013). Monitoring of mental workload levels during an everyday life office-work scenario. *Personal and ubiquitous computing*, *17*(2), 229-239.

[11] Sevil, M., Hajizadeh, I., Samadi, S., Feng, J., Lazaro, C., Frantz, N., &Cinar, A. (2017). Social and competition stress detection with wristband physiological signals. *In IEEE 14th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, 39-42.

[12] Mozos, O.M., Sandulescu, V., Andrews, S., Ellis, D., Bellotto, N., Dobrescu, R., &Ferrandez, J.M. (2017). Stress detection using wearable physiological and sociometric sensors. *International journal of neural systems*, *27*(2), 1650041.

[13] Giannakakis, G., Pediaditis, M., Manousos, D., Kazantzaki, E., Chiarugi, F., Simos, P.G., &Tsiknakis, M. (2017). Stress and anxiety detection using facial cues from videos. *Biomedical Signal Processing and*

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 3, 2021, Pages. 8364 - 8374
Received 16 February 2021; Accepted 08 March 2021.

Control, *31*, 89-101.

[14] Sharma, N., & Gedeon, T. (2014). Modeling observer stress for typical real environments. *Expert Systems with Applications*, *41*(5), 2231-2238.

[15] Dinges, D.F., Rider, R.L., Dorrian, J., McGlinchey, E.L., Rogers, N.L., Cizman, Z., & Metaxas, D.N. (2005). Optical computer recognition of facial expressions associated with stress induced by performance demands. *Aviation, space, and environmental medicine*, *76*(6), B172-B182.

[16] Sharma, N., & Gedeon, T. (2012). Objective measures, sensors and computational techniques for stress recognition and classification: A survey. *Computer methods and programs in biomedicine*, *108*(3), 1287-1301.

[17] Pampouchidou, A., Pediaditis, M., Chiarugi, F., Marias, K., Simos, P., Yang, F., &Tsiknakis, M. (2016). Automated characterization of mouth activity for stress and anxiety assessment. *In IEEE International Conference on Imaging Systems and Techniques (IST)*, 356-361.

[18] Gao, H., Yüce, A., &Thiran, J. P. (2014). Detecting emotional stress from facial expressions for driving safety. *In IEEE International Conference on Image Processing (ICIP)*, 5961-5965.

[19] Ekman, P., & Friesen, W.V. (1978). *Facial action coding system: Investigator's guide*. Consulting Psychologists Press.

[20] Viegas, C., Lau, S.H., Maxion, R., & Hauptmann, A. (2018). Towards independent stress detection: A dependent model using facial action units. *In International Conference on Content-based Multimedia Indexing (CBMI)*, 1-6.

[21] Gavrilescu, M., &Vizireanu, N. (2019). Predicting depression, anxiety, and stress levels from videos using the facial action coding system. *Sensors*, *19*(17), 3693.

[22] Prasetio, B.H., Tamura, H., &Tanno, K. (2018). The facial stress recognition based on multi-histogram features and convolutional neural network. *In IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 881-887.

[23] Prasetio, B.H., Tamura, H., &Tanno, K. (2018). Support vector slant binary tree architecture for facial stress recognition based on Gabor and HOG Feature. *In International Workshop on Big Data and Information Security (IWBIS)*, 63-68.

[24] Lyons, M., Akamatsu, S., Kamachi, M., &Gyoba, J. (1998). Coding facial expressions with gabor wavelets. *In Proceedings Third IEEE international conference on automatic face and gesture recognition*, 200-205.

[25] Pediaditis, M., Giannakakis, G., Chiarugi, F., Manousos, D., Pampouchidou, A., Christinaki, E., &Tsiknakis, M. (2015). Extraction of facial features as indicators of stress and anxiety. *In 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 3711-3714

[26] Yuen, P., Hong, K., Chen, T., Tsitiridis, A., Kam, F., Jackman, J., & Lightman, S. (2009). Emotional & physical stress detection and classification using thermal imaging technique.*In Proceedings of the Third International Conference on Crime Detection and Prevention (ICDP)*, 1–6.

[27] Zhao, G., &Pietikainen, M. (2007). Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE transactions on pattern analysis and machine intelligence*, *29*(6), 915-928.

[28] Hernandez, B., Olague, G., Hammoud, R., Trujillo, L., & Romero, E. (2007). Visual learning of texture descriptors for facial expression recognition in thermal imagery. *Computer Vision and Image Understanding*, *106*(2-3), 258-269.

[29] Fasel, B., &Luettin, J. (2003). Automatic facial expression analysis: a survey. *Pattern recognition*, *36*(1), 259-275.

[30] Trujillo, L., Olague, G., Hammoud, R., & Hernandez, B. (2005). Automatic feature localization in thermal images for facial expression recognition. *In IEEE Computer Society Conference on Computer Vision and*

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 3, 2021, Pages. 8364 - 8374
Received 16 February 2021; Accepted 08 March 2021.

Pattern Recognition (CVPR'05)-Workshops, 14-14.

[31] Manglik, P., Misra, U., &Maringanti, H. (2004). Facial expression recognition. *In Proceedings of the International Conference on Systems, Man and Cybernetics*, 2220–2224.

[32] Neggaz, N., Besnassi, M., &Benyettou, A. (2010). Application of improved AAM and probabilistic neural network to facial expression recognition. *Journal of Applied Sciences(Faisalabad)*, *10*(15), 1572-1579.

[33] Sandbach, G., Zafeiriou, S., Pantic, M., &Rueckert, D. (2012). Recognition of 3D facial expression dynamics. *Image and Vision Computing*, *30*(10), 762-773.

[34] Sharma, N., Dhall, A., Gedeon, T., &Goecke, R. (2014). Thermal spatio-temporal data for stress recognition. *EURASIP Journal on Image and Video Processing*, *2014*(1), 1-12.