Journal of Science and Engineering Research
Vol. 2, No. 2, 2023, pp.27-36
eISSN: 2957-7187

Discrepancy Reduction and Transfer Learning in Convolution Neural Networks for Stress Emotion Recognition

David Taniar

Monash University, Clayton, Australia
david.taniar@Monash.edu

DOI: http://dx.doi.org/10.56828/jser.2023.2.2.2
Article history: *Received (September 30, 2023); Review Result (November 5, 2023); Accepted (December 10, 2023)*

Abstract

Assessing the impact of stress on a person's socio-economic well-being necessitates the use of stress management techniques. It was decided to combine the Deep Belief Network with the DBNTL technique to forecast stress levels. In contrast to DBNTL, it does not attempt to reconcile various domain distributions on the same level of abstraction. Closer physical closeness between two people. The result of the research uses an innovative transfer learning method to build cutting-edge neural networks that are capable of detecting small-scale stress and emotional data fields in real-time. TLCNNDR can learn top-layer features from marginal and joint distributions, which are two types of distributions. TLTNNDR discovered more dependable features at the top of the list. A personality may be assigned to every emotion and stressor. TLCNNDR layers are taught to be equally effective, allowing for the assessment of many levels in MDD and JDD. These two imbalances may aid in the reduction of emotional and stress transfer. TLCNNDR is used to classify CNN and DNTL in clinical studies.

1. Introduction

Seventy percent of the population of the United States is always concerned. In addition to an increased risk of cancer, chronic stress has [1] been related to an increased risk of cardiovascular disease, depression, type 2 diabetes, and pharmaceutical addictions. It is shown that stress, both physical and emotional, is detrimental to one's health. It is critical to research and develop sensitive techniques for detecting and assessing stress. It may be possible to detect stress in real time using these technologies [2]. Health practitioners may be able to treat stress-related diseases more effectively in the future as a consequence of people's increased capacity to cope with stress in their daily lives.

Researchers have devised a multitude of methods for detecting stress, which are detailed below. To identify video strain, a deep neural network model is used. Detection of psychological stress to detect stress in speech throughout [3] the study, recurrent neural networks were utilized throughout. Researchers have created a variety of techniques for analyzing physiological data collected by sensors attached to the human body to detect stress and identify emotional states.

These methods include It has been shown that physiological markers of stress [4], such as blood pressure and sweat production, are very efficient in identifying human discomfort. The identification of stress may be accomplished via the use of non-invasive methods based on physiological data analysis. Because of this, these techniques [5] have the potential to substantially improve human well-being. In the past, machine learning was used to identify stress based on physiological data collected from the subject.

It was all over the place. Here, we propose two new types of complex neural networks that are capable of carrying out prior techniques for detecting stress and emotional states [6]. Stress has been the subject of many studies that have used stress detection and physiological indicators.

The researchers did not look at any previous technologies, except coetaneous electrocardiography and electromyography sensors. With the use of physiological data, conventional machine learning methods were applied to assess the stress and emotional state of the participants.

A variety of machine learning techniques [7], including vector support machines, K-nearest neighbours, and random forests, were used to assist in various decision-making processes. Healey and Picard's research on stress was one of the first scientific attempts to demonstrate the existence of stress in humans. The researchers conducted their tests with the use of sensors that measured electrical activity, breathing rate, and skin conductance. A total of twenty-two characteristics were manually created from the data set described above. It was determined if events were stressful or not stressful using the LDA [8] machine learning method, which categorized occurrences into stressful and no stressful groups using two mutually exclusive variables to determine whether they were stressful or not stressful.

An accelerometer (ACC), a person's blood volume (VVP), the electrical dermal activity of the skin, the heart rate, and the temperature of the skin are all examples of sensors. Two out of the 63 characteristics discovered were deemed acceptable for machine [9] learning, bringing the total to two. It was found that using a random method to categorise forestry equipment resulted in an accuracy rate of 72 percent in the classification. The researchers utilized electromyography, speed, and cadence, as well as an ECG [10] and sensors, to assess their participants' feelings. Volunteers were instructed to listen to music to generate a variety of emotional responses. Using the LDA [11] machine learning method, the scientists were able to categorise emotions after manually creating characteristics that were specific to machine learning technology.

The categorization ratio is independent of the topic matter 70% of the time, according to the data. Scientists doing research on stress and [12] emotion detection have looked at the use of physiological markers to identify certain emotions. They carried out their research by analyzing physiological data collected by chest and wrist sensors, among other methods. Using machine learning, this study were able to do [13] binary classification (difference between stressed and unstressed states) as well as three-class categorization (differentiating between a baseline condition, a stressed state, and an amused state). For each goal, this study evaluated several methods such as decision trees, RAF, AdaBoost, LDA, and K-neighborhood [14] search. To accomplish both of the ultimate objectives, machine learning methods were shown to be the most accurate in 75.21 and 76.60 percent of cases, respectively, for the 3-class classification of wrist and chest movements. When it came to the categorizing wrist and chest instances, machine learning techniques performed the best: they achieved an accuracy of 87.12 percent in 87.12 percent of cases and 92.83 percent in 92.83 percent of cases. Almost all conventional machine-learning methods, by and large, need characteristics to be manually programmed, which is a major disadvantage.

The bulk of previous research has been constrained by physiological circumstances and data availability. One example [15] of using the characteristics of ECG sensor data is the usage of cardio frequency and cardiac variability, as well as associated statistics such as the average, variance, and energy of low, medium, and high frequency [16] bands. Breathing rate data were categorized based on their mean and standard deviation, which were calculated for intake, exhalation, and duration of breathing time.

Electromyography (EMG) data and associated statistics such as mean and standard deviation to determine the sensor signal characteristics for electromyography sensors this study used and also the sensor signal dynamics were examined. [17] Calculating them is difficult since each sensor must be manually translated into a unique set of characteristics. Furthermore, there is a paucity of scientific evidence to support the precise physiological signals that underlie these characteristics, and there is no assurance that the signal functioning of all algorithms has been adequately accounted for by taking into account the characteristics of previous methods.

2. Literature Survey

Camille Daudelin-Peltier et al., [15], proposed research to examine the emotions experienced by healthy young males when they are subjected to extreme social stress. Each participant received a TSST-G as well as a control group. They performed an unexpected mega mix during a house party to demonstrate the wide variety of facial emotions. Participants were expected to express all six fundamental emotions to complete the assignment. When a person is under stress, their ability to tolerate disgust is diminished. Individuals may do so to adjust to a new situation, alleviate stress, or avoid social judgment. Second, our findings indicate that the severity of surprises diminishes with time, which suggests that it enhances performance. A more surprising response in stressful social settings, according to our theory, reflects an evolutionary change that improves the probability of collecting critical information about possible dangers.

Domes [16] proposed research recognizing the emotions of people who need the ability to read their facial expressions. Although face emotion recognition research is still in its infancy, it has the potential to affect acute stress and the physiological responses that follow. Participants were assigned to one of two conditions in the virtual world: a stressful virtual reality version of the Trier Social Stress Test (n = 23) or a non-stressful control condition in which they did not experience stress. After that, using three distinct signal detection theory expressions, a computerized face recognition test assessed the actual emotions of joy and anger experienced by both groups. Seven separate studies collected saliva to measure free salivary cortical and alpha-amylase. According to the findings of the research, individuals who were stressed needed more time to recognize and react to emotional signals. While stress does not improve performance on tasks, the synthesis of cortical and alpha-amylase does so. Anxiety increases people's sensitivity to social signals, allowing them to recognize danger or social support more quickly and effectively.

Zhang et.al proposed [17] a Physiological response to mental and physical difficulties may be considered negative emotional stress. Long-term stress exposure may have serious impacts on individuals like depression, which can in extreme cases lead to suicide; thus, it is essential to assess and manage stress in real-time. This study provides a novel paradigm in this work for detecting stress in real time. The framework recognizes three expressions linked to stress to identify stress: anger, fear and sorrow. In addition, it provide a linked coevolutionary network for the development of a deep network for sensing facial expressions by combining low- and high-level data. If a certain threshold surpasses the number of frames associated with stress, the framework calls on users to rest. Experimental results show that this approach exceeds others in the identification of facial expressions and provides high-performance stress detection.

Attwood et.al proposed [18] a high level of trait anxiety has difficulty interpreting others' emotional signals. However, just a few researches have looked at the relationship between state anxiety and emotional face processing. A 7.5 percent CO2 model was used to calculate state anxiety in both an experimental and large observational study on emotional expression identification, and it was shown to be accurate (anger, sorrow, surprise, disgust, fear, and pleasure). Worry reduced the accuracy of global emotion identification and increased the bias in interpretation (a tendency to identify rage over pleasure). State anxiety seems to act as a moderator in the relationship between trait anxiety and state anxiety, according to some research. As a consequence of increasing the frequency, intensity, or duration of anxious state episodes in patients with anxiety disorders, these findings may be influenced.

Alberdi et al., [19] proposed an algorithm and it may lead to health and financial problems for both people and organizations. As a result of the increasing mental load and technological advances, there has also been a rise in the number of changes and the need to adapt. Recognizing the early symptoms of stress may assist in preventing long-term consequences. A method for detecting stress does not exist at this time that is automated, continuous, or discrete. Because of the multimodal character of stress and the research conducted on it, the proposed method incorporates several approaches. However, the research looked at various metrics across three major modalities: psychological, physiological, and behavioural; as well as environmental factors and factors that influence behaviour.

3. Proposed Methodology

Grid-structured data and image-based classification are two areas in which convolution neural networks may be put to use as applications. They are also often used for image-based classification, which is another use. CNNs [20] were chosen because they are capable of performing complex image recognition, segmentation, detection, and retrieval tasks, among other things. The usage of CNNs improved sparse interaction, parameter sharing, and equivalent representation, which were all found to be important results. Deeper networks with billions of parameters can deal with complicated and unique picture characteristics more effectively and efficiently than previous training techniques because they have more parameters to work with. In contrast, fully linked networks, which need a fixed-size input, do not require this. Each grid of input data is processed by a matrix (a filter) that divides the grids into smaller groups, with each group being multiplied by a kernel matrix to produce an Output Function Map. In a similar vein, the receptive region of each neuron receives some information while discarding the remainder. To correctly train a fully connected neural network with the same functional identification and classification of 1000 pictures per neuron as in this experiment, it would take much longer. However, CNNs would have learned functions that did not affect the size of the kernel. Layer sequences are usually comprised of downsampling maps, traditional fully connected layers, and layer bundling, among other things. Except for the basic components (as previously mentioned), the number of layers and filter sizes used in the literature are variable.

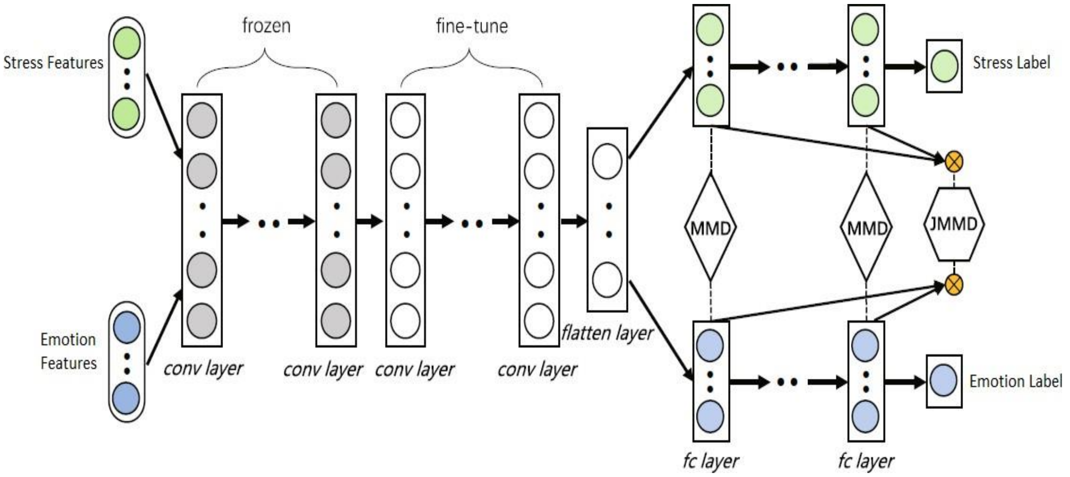

Figure 1: Overall architecture of the TLCNN DR method

It is possible for emotional input to have an impact on the whole pre-trained model rather than just one layer because of this. VGG-16 is a pertained model that consists of five convolution blocks (each with two or three convolution layers) plus a bundling layer that has been pre-trained (Figure 4). 2D convolution and bundling are operations that act on a two-dimensional image. Cascade Convolution is comprised of two convolution layers, one Max Pooling layer, and one additional layer (the first convolution block). The numbers 10, 2021, and 2021, 10,10368 are the inputs to Conv Block 2. Assume that Block 1 has the value 224. The VGG-16 takes as input a 224x224 three-input colour image with a resolution of 224x224. A 7x7 512 linear vector array is created after many layers of coating and pooling are applied. This layer has been flattened into a linear vector of size 25.088. Using the linear method, this study was analyzed and can invert a 1000-vector, which can then be fed into the dense layer with 128 values. The ultimate thick covering may elicit a wide range of feelings in the wearer.

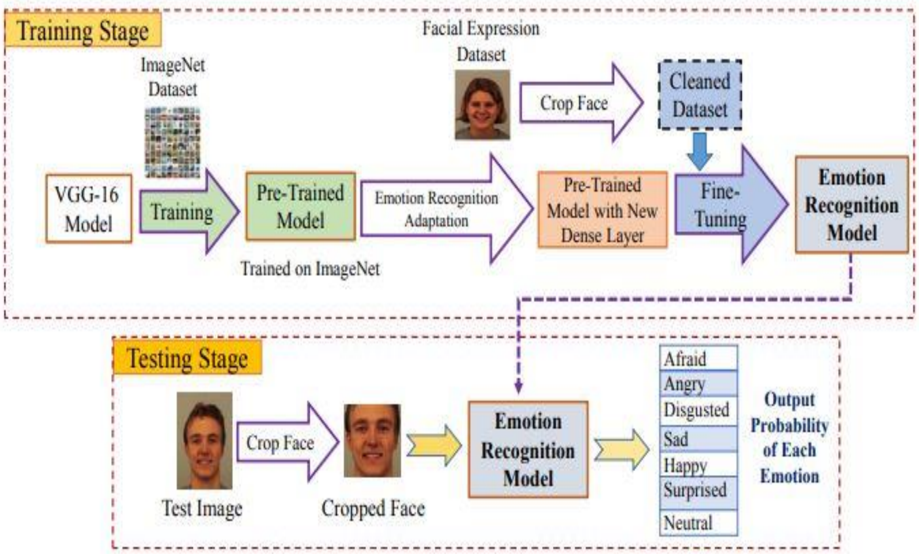

Figure 2: Showing VGG16 architecture for the proposed methodology

4. Results and Discussions

The investigation was divided into two phases: before the catastrophe and after it. Early in the model's development, pre-trained convolution layers were frozen, and fully-connected layers were fine-tuned to improve performance (as described in Section 3). It is recommended to avoid CNN's networks such as VGG16, Inception V3, and Exception since their accuracy rates are low. It is also recommended that InceptionResnetV2, which provided the best results, be kept in place. The second stage was to unfreeze InceptionResnetV2 so that the network's layers could be fine-tuned.

Therefore, the weights of the neural network were reduced to retain the model with the greatest accuracy obtained during the training phase. After iterating the method in the Table 1, it achieved an overall accuracy of 79.5 percent on the validation set, as shown in Figure 2. VGG16 came in the lowest place when compared to the other networks, whereas InceptionResnetV2 came in first place (Table 2). The layers of InceptionResnetV2 have been modified.

Table 1: Showing confusion matrix for 600 test data set

	Predicted Class		
		Positive	Negative
Actual Class	Positive(1500 for each class)	True Positive 1390	False Negative 100
	Negative(3000 For other classes)	False Positive 110	True Negative 2900

The proposed method was effective in assessing both front and profile views in benchmark datasets, demonstrating its versatility. There are just front views in the JAFFE dataset; on the other hand, there are four extra angles of view in the KDEF dataset. Because only half of the face is seen in the left and right views of KDEF, only half of the face is displayed in the middle perspective. When just one eye and one ear are visible, it makes it more difficult to find anything in the dark. It brings us great pleasure.

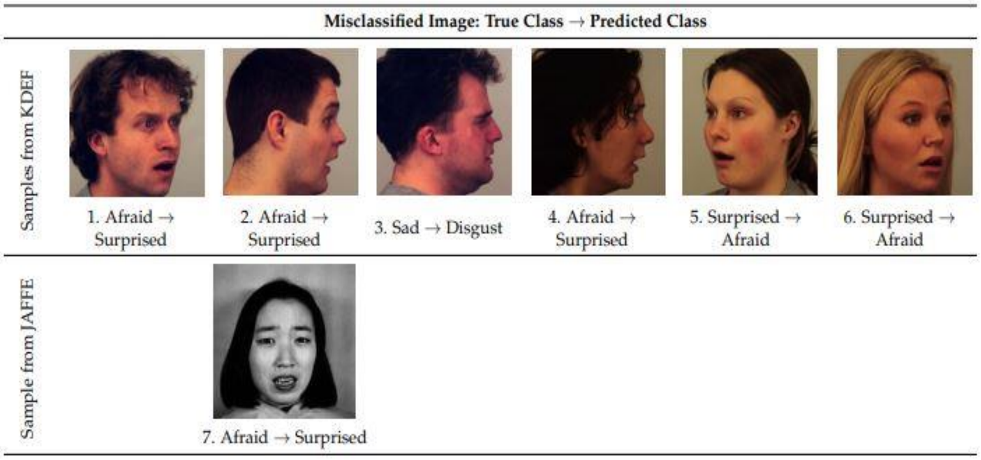

Figure 3: Showing the miss-classified images of the proposed classifier

The results of the tests from two distinct datasets are sufficient to establish competence. More data sets will likely be successful if the proposed strategy is followed. Data sets are often referred to as data, which is incorrect. A combination of pre-processing and method tweaking is needed for low-quality photos or photographs with an uneven balance. study. A similar set of difficulties arises when images and videos are taken in an uncontrolled environment. Future research will pay even greater attention to the sequences that are currently in use.

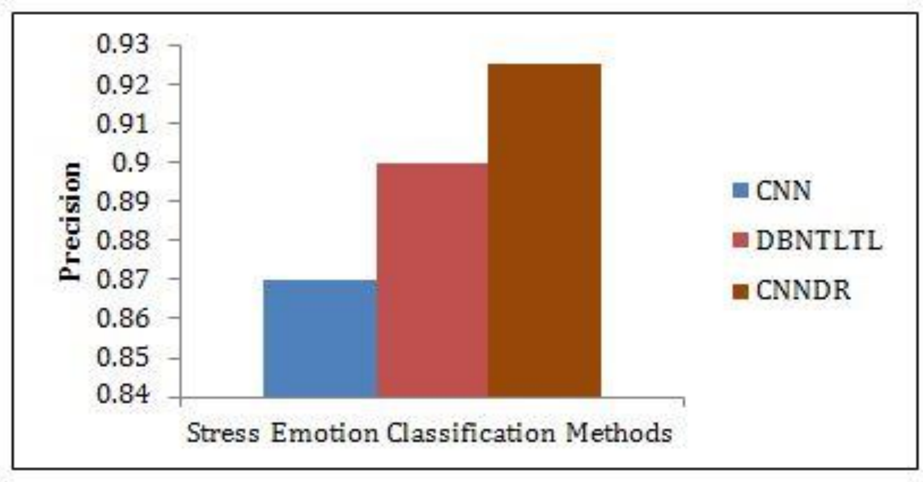

Figure 4: Showing performance of the classifier

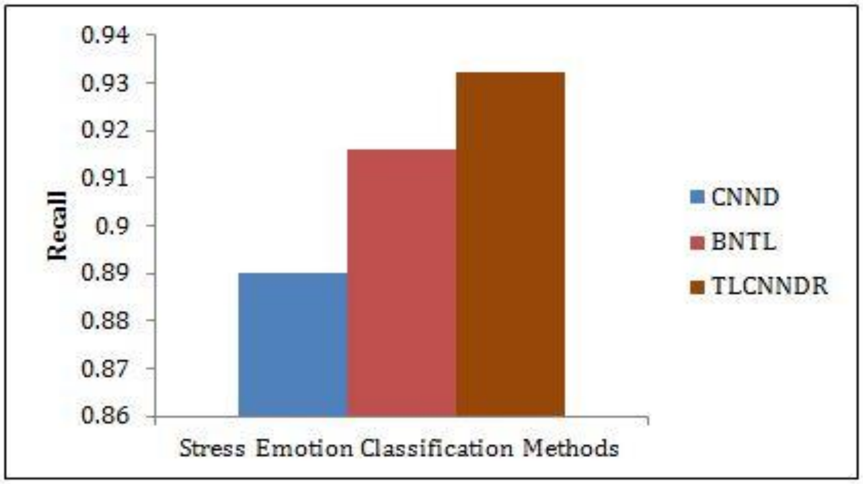

Figure 5: Showing performance of the classifier with recall

For example, if the precision score for a grade C class is 1.0, it indicates that all items identified as part of the class have received the grade (but not the number of incorrectly marked items from the class), and the recall score for the same class is 1.0, it indicates that all items identified as part of the class have been excluded from the class. However, it is not known how many more Class C goods were erroneously classified at this time. Precision and reinforcement are often associated with one another. For example, a hypothetical situation such as brain surgery might be used to demonstrate this point. A brain tumour has been removed by an operator, and a patient has been brought in. Surgeons must remove all cancerous cells from the body to keep it cancer-free. Healthy brain cells, on the other hand, must not be destroyed since doing so would impede the patient's ability to function. More brain tissue may be removed than is required to ensure full eradication of cancer cells by the surgical team. This also acts as a reminder, although one that is less accurate than the last one.

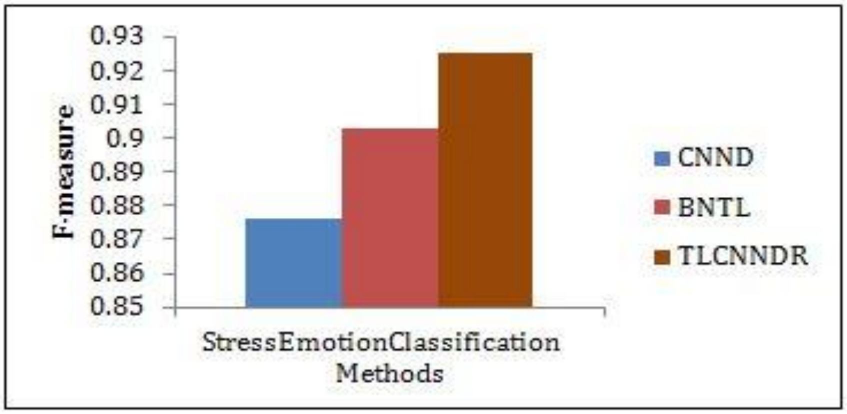

Figure 6: Showing the F-measure of the classifier

Choosing how much brain tissue to remove may cause a surgeon to be concerned about removing all cancer cells from the brain tissue. This option improves the retrieval accuracy by making it more precise. Increasing the probability of removing both healthy and malignant cells may be decreased by improving memory (positive outcome). While accuracy reduces the probability of eliminating cancer cells, it also reduces the possibility of removing all cancer cells from a patient (negative outcome).

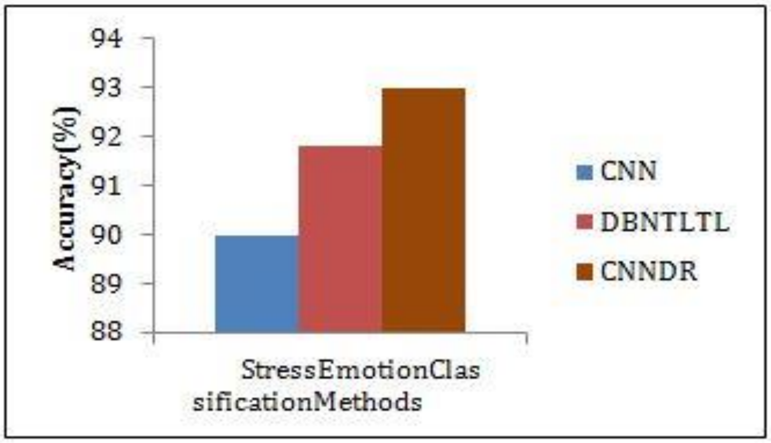

Figure 7: Showing the accuracy of the model

When evaluating classifier outputs against reliable external judgements, classification tasks make use of the words true positives, true negatives, wrong positives, and false negatives (see Type I and Type II mistakes for definitions). For example, the labels "good" and "bad" represent the expectation of the classifier, while the labels "true" and "false" indicate whether or not the classifier's expectation matches the outside judgement of the classification (sometimes known as the observation).

5. Conclusion

The findings of this research were used to build a DCNN by combining TL with a pipeline-altering technique and analysing the results. People's emotional facial expressions may be identified using this method. On the well-known KDEF and JAFFE datasets, eight different pre-trained DCNN models were identified, according to the results. When applied to emotional datasets with changing profile views, the suggested approach achieves remarkable accuracy. A variety of Accuracy of Recognition Experiments was used in this study. The overwhelming majority of profile views are wrongly classified when compared to the pre-trained DCNN model or certain puzzling face pictures. Individual hyperparameters are tweaked to get the desired outcomes. Optimizing results may need the use of pre-trained algorithms and a focus on profile views. Forecasting precision of the study's results may have ramifications for a wider range of real-world industrial applications, such as patient monitoring in a hospital or as part of a security surveillance system. The ability to recognise facial expressions may also be useful. Recognizing particular words or physical gestures should be part of controlling one's emotions.

References

[1] Wardani, R. (2020). Academic hardiness, skills, and psychological well-being on new students. Journal Psikologi, 19(2), 188–200. DOI: 10.14710/jp.19.2.188-200.

[2] Joshua, E. S. N., Chakkravarthy, M., & Bhattacharyya, D. (2020). An extensive review on lung cancer detection using machine learning techniques: A systematic study. Revue d'Intelligence Artificielle, 34(3), 351-359. DOI: 10.18280/ria.340314.

[3] Hemamalini, R., Ashok, V., & Sasikala, V. (2018). A study on stress management and its impact among students. International Journal of Academic Research in Economics and Management Sciences, 7(3). DOI:10.6007/ijarems/v7-i3/4439.

[4] Schmidt, P., Reiss, A., Durichen, R., & Laerhoven, K. V. (2018). Introducing WESAD, a multimodal dataset for wearable stress and affect detection. In: Proceedings of the 20th ACM international conference on multimodal interaction, October 16–20, 2018, Boulder, USA. New York: Association for Computing Machinery; https://dl.acm.org.

[5] Kirschbaum, C., Pirke, K. & Hellhammer, D. (1993). The trier social stress test—A tool for investigating psychobiological stress responses in a laboratory setting. Neuropsychobiology. 28, 76–81.

[6] Karthikeyan, P., Murugappan, M., & Yaacob, S. (2013). Detection of human stress using short-term ECG and HRV signals. J Mech Med Biol, DOI:10.1142/S0219519413500383.

[7] Kyriakou, K., Resch, B., Sagl, G., Petutschnig, A., Werner, C., Niederseer, D., Liedlgruber, M., Wilhelm, F., Osborne, T., & Pykett, J. (2019). Detecting moments of stress from measurements of wearable physiological sensors. Sensors. DOI:10.3390/s19173805.

[8] Joshua, E. S. N., Bhattacharyya, D., Chakkravarthy, M., & Byun, Y. –C. (2021). 3D CNN with visual insights for early detection of lung cancer using gradient-weighted class activation. Journal of Healthcare Engineering, Article ID:6695518, 11 pages. DOI:10.1155/2021/6695518.

[9] O'Toole, A. J., Harms, J., Snow, S. L., Hurst, D. R., Pappas, M. R., Ayyad, J. H., & Abdi, H. (2005). A video database of moving faces and people. IEEE Trans. Pattern Anal. Mach. Intell. 27, 812–816.

[10] Fernández-Caballero, A., Martínez-Rodrigo, A., Pastor, J. M., Castillo, J. C., Lozano-Monasor, E., López, , Zangróniz, R., Latorre, J. M., & Fernández-Sotos, A. (2016). Smart environment architecture for emotion detection and regulation. J. Biomed. Inf. 64, 55-73.

[11] Karthikeyan, P., Murugappan, M., & Yaacob, S. (2013). Detection of human stress using short-term ECG and HRV signals. J Mech Med Biol. DOI:10.1142/S0219519413500383.

[12] Cohen, S., Janicki-Deverts, D., & Miller, G. E. (2007). Psychological stress and disease. JAMA, 298, 1685-1687.

[13] Steptoe, A. & Kivimaki, M. (2013). Stress and cardiovascular disease: An update on current knowledge. Annu Rev Public Health, 34, 337–354.

[14] Tzirakis, P., Trigeorgis, G., Nicolaou, M., Schuller, B. W., & Zafeiriou, S. (2017). End-to-end multimodal emotion recognition using deep neural networks. IEEE J Sel Top Signal Process, 11, 1301–1309.

[15] Daudelin-Peltier, C., Forget, H., & Blais, C. (2017). The effect of acute social stress on the recognition of facial expressions of emotions. Sci Rep, 7, 1036. DOI:10.1038/s41598-017-01053-3.

[16] Domes, G. & Zimmer, P. (2019). Acute stress enhances the sensitivity for facial emotions: A signal detection approach. Stress, 22(4), 455-460. DOI:10.1080/10253890.2019.1593366.

[17] Zhang, J., Mei, X., Liu, H., Yuan, S., & Qian, T. (2019). Detecting negative emotional stress based on facial expressions in real-time. 2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP). DOI:10.1109/siprocess.2019.8868735.

[18] Attwood, A. S., Easey, K. E., Dalili, M. N., Skinner, A. L., Woods, A., Crick, L., & Munafò, M. R. (2017). State anxiety and emotional face recognition in healthy volunteers. Royal Society Open Science, 4(5), 160855. DOI:10.1098/rsos.160855.

[19] Alberdi, A., Aztiria, A., & Basarab, A. (2015). Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review. Journal of Biomedical Informatics, 59, 49-75. DOI:10.1016/j.jbi.2015.11.007.