Stress Recognition Using Facial Landmarks and Cnn (Alexnet)

Ramesh Naidu P1, Pruthvi Sagar S2, Praveen K3, Kiran K4 and Khalandar K5
Department of Computer Science and Engineering, Nitte Meenakshi Institute of Technology, Bangalore, Karnataka, India.
Email: ramesh.naidu@nmit.ac.in

Abstract. Stress is a psychological disorder that affects every aspect of life and diminishes the quality of sleep. The strategy presented in this paper for detecting cognitive stress levels using facial landmarks is successful. The major goal of this system was to employ visual technology to detect stress using a machine learning methodology. The novelty of this work lies in the fact that a stress detection system should be as non-invasive as possible for the user. The user tension and these evidences are modelled using machine learning. The computer vision techniques we utilized to extract visual evidences, the machine learning model we used to forecast stress and related parameters, and the active sensing strategy we used to collect the most valuable evidences for efficient stress inference are all discussed. Our findings show that the stress level identified by our method is accurate is consistent with what psychological theories predict. It is vital to have a gadget that can collect the appropriate data. The use of a biological signal or a thermal image to identify stress is currently being investigated. To address this limitation, we devised an algorithm that can detect stress in photos taken with a standard camera. We have created DNN that uses facial positions points as input to take advantage of the fact that when a person is worried their eye, mouth, and head movements differ from what they are used to. The suggested algorithm senses stress more efficiently, according to experimental data.

Keywords: AlexNet, Stress, Strain, Machine Learning, Facial, Visual.

1. Introduction
Stress is a feeling of being overworked that can be brought on by a number of things in our daily lives. Stress management is critical in today's world to keep stress levels low and health risks to a minimum, as stress is one of the leading causes of chronic illnesses.

Recently, as a result of the high levels of stress experienced by modern individuals, a system has been designed to detect whether a user is stressed and to provide feedback aimed at lowering stress. Many strategies based on bio signals have been investigated in the subject of stress identification. However, because bio signal measuring equipment must be connected to the user's body in order to measure bio signals, the user may be rejected. As a result, several research on stress detection using thermal pictures have been undertaken. However, because stress cannot be identified without thermal imaging equipment, it is impossible to perceive stress in everyday life. In the case of stress recognition study utilizing a generic picture, on the other hand, the majority of studies have employed a rather basic characteristic. Stress disorders are a widespread problem among today's working professionals in every business. When employees' lives and work cultures change, they are more likely to experience stress. Despite the fact that many industries and organization provide mental health-related program and make efforts to enhance the workplace environment, the problem persists.
We present a technique for identifying stress using high-dimensional characteristics extracted from facial images captured by a generic camera in this study[1,2]. We also utilize the placement of face landmarks that indicate a significant shift when stressed in order to learn more efficient features. Machine learning is a subset of Artificial Intelligence that allows computers and computing systems to learn and improve without being explicitly programmed by humans. Machine learning is the process of creating computer programmes that can retrieve data and learn on their own. This is especially valuable in healthcare, where there is a large amount of data that can be put into an intelligent system and trained, the resultant prediction model will be unrivalled, free of human mistakes, and minimize diagnostic time.

1.1 Sequential model
A sequential method is suitable for stacks of layers with one input tensor and one output tensor for each layer. When the following conditions apply, a sequential model is not appropriate:
Your model contains a lot of inputs and outputs.
Any of your layers can have many inputs and outputs.
Layer sharing is required.
You want a topology that isn't linear (e.g. a residual connection, a multi-branch model)

1.2 Convolutional Neural Network (AlexNet/CNN)
CNNs are Deep Learning algorithms that use visual input to determine which attributes and objects should be weighted and balanced, which enables them to be isolated and classified. [3,4]. The amount of pre-processing required by a Conv Net is much less than that necessary by other classification methods. While basic approaches need hand-engineering of filters, Conv. Nets can learn these filters/characteristics with enough training.

Figure 1. Convolutional Neural Network

CNN has built an essential component in its primary building: the convolutional layer. The parameters of the layer are a collection of learnable filters (or kernels) that cover the input volume from the input layer to the output layer. Convolution is used in order to perform a 2-dimensional activation map across the width and height of the input region to simulate the action of passing the football. The key considerations of elections and essence of Pooling, a type of non-linear down-sampling, is another key idea in CNNs. Pooling may be implemented using a variety of non-linear functions, the most popular
of which being max pooling [5,6]. This term describes the process of converting data from a multidimensional array to a one-dimensional array that will be used in the next layer. In order to synthesise a single long feature vector, we use the flattened output of the convolutional layers. Fully linked layer, often known as the final classification model, is connected to this. Conversely, if you're interested in an original method for expressing [7,8], go no further. In order to send the pixel data as a single, connected line, we link the final layer to a single line that includes all of the pixel data. And once more what is the purpose of the last layer?

Figure 2. CNN's main building

Emotions have been researched by psychologists since the nineteenth century, yet there is no commonly recognized description of what emotions are or how they are formed. Emotions and physiology, on the other hand, have been linked for over a century [9,10]. Many studies use physiological data including electro-dermal, cardiovascular, and muscle activity to assess participants' emotional states, including stress levels. Affective states can also be assessed using other tools such as questionnaires and scales.

Figure 3. CNN's Fully connected Layer

These, on the other hand, cannot be given to users without disrupting their work and thereby altering their emotions [11,12]. Aside from the inherent ambiguity of describing emotions in writing, which may impair any sort of self-reporting, the inherent ambiguity of defining emotions in writing might weaken the reliability of such instruments. As a result, building systems that can identify stress through physiology is interesting, and not just for research purposes.
2. Literature Survey

Individuals experience psychological and bodily strain as a result of their daily work responsibilities. Every individual is subjected to strain over time, and how they react with stress is likely to have both perceived and real implications [13,14]. Coping with stress necessitates the use of several personal and environmental resources. In this subject, several research have been undertaken. Workplace stress was investigated by Alireza Bolharieal. (2012). Jong-Ho Kim et al looked at the social lives of college students and found that physical activity can help reduce stress levels. Enrique Garcia-Ceja et al. (2016) investigated the use of cell phones as a possible tool for detecting stress-related behaviour. Stay Active, a stress detection system invented by Panagiotis Kotsiopoulos et al, combines sleep patterns, physical activity, and social contacts to identify stress.

3. Research Motivation And Problem Statement

3.1 Research Motivation

At first, facial expression analysis was primarily a topic of study for psychologists. Recent advances in image processing and pattern recognition, on the other hand, have sparked a lot of interest in autonomous facial expression identification research. Previously, a great deal of work was put on recognizing face emotion in static photos. Many approaches have been used for this goal, including neural networks, Gabor wavelets, and active appearance models [15,16]. The fact that still photos frequently capture the pinnacle of the expression, i.e. the moment when the signs of emotion are most pronounced, is a significant disadvantage of this method.

3.2 Statement of the Problem

Human emotions and intents are conveyed through facial expressions, and the facial expression system's key component is determining an efficient and effective feature. While intelligent man-machine interfaces, intelligent visual surveillance, teleconferencing, and real-time animation are only a few of the uses for facial expressions, real-time animation from live motion images also utilises them. [6]. For effective conversation, face expressions are useful. The majority of facial expression recognition research and systems are confined to six basic expressions (joy, sad, anger, disgust, fear, surprise. Face detection and recognition is a difficult process in which it is critical to pay attention to major components such as face configuration, orientation, and position where the face is positioned.

4. Proposed System

Face image and facial landmark detection are conducted initially in the proposed method for stress recognition. A Convolution Neural Network (CNN) algorithm was utilized. Face photos and expressions detected earlier are fed into the proposed network, which then outputs stress recognition results. A convolution is a combined integration of two approaches that shows you how one affects the other.

As nonlinearities are amplified in this step, the correction function is employed to increase nonlinearity on CNN. Data gathering is comprised of a variety of distinct components, none of which have anything to do with one another. However, despite the fact that it is a non-straight issue, it may be seen as a linear problem, because the information grouping is linear under this function. It is meaningless to talk about the neural network's capacity to identify its own achievements while searching through a data set, since the concept of spatial invariance has nothing to do with it. The functions must then be flattened, because they are pooled in a diagram. When the whole matrix of the drawn-up map is flattened, it becomes a single column of the neural network. When the map has been flattened, the network should then feed it through. All of them are needed in this transfer: the plate, the completely connected plate, and the output sheet. The hidden layer is the same as the completely linked layer. In this example, the ANN layer is connected. The output sheet serves as a resource for organisations. A prediction error is calculated once knowledge is passed over the network.
5. System Design And Architecture

5.1 Architecture Diagram
In the proposed network, Facial photos and face landmarks recognized earlier are fed into the proposed network, which then outputs stress recognition results. The stress recognition results are divided into three categories: Class 1-No stress, Class 2-weak stress, and Class 3-strong stress. To optimize neural network structure Shortcut mapping and bottleneck architecture are used in a network proposal [17,18]. When mapping of shortcuts is done to the neural network structure that has been deeper owing to the various layers, it is feasible to simplify the learning process and decide the direction of learning. Because of the improved depth, the deep neural network may now be easily adjusted to improve accuracy while the number of internal parameters can be lowered, the number of feature maps can also be reduced. resulting in higher performance and reduced calculation time.

![Architecture Diagram](image)

Figure 4. Architecture Diagram

5.2 Flow Chart Diagram
In the proposed network, each entity have respective task to do so each entity accepts the data process the data and feed forwards the data to another entity here our network you can see that camera accepts the real time video data and pre-process the data using our face detection algorithm and it extracts the facial features from the input image and these data is feed forwarded to our proposed algorithm called emotion detection these algorithm by making use of the model which we have trained predicts the emotion and relates it with the 3 class of stress (No Stress ,Weak Stress , Strong Stress).
Figure 5. Flow Chart Diagram

Figure 6. System Representation
6. Methodology and Implementation

In the proposed network, we will be detecting the facial landmarks from real-time video captured data with the help of the haar cascade classifier and detect the facial emotion of the subject. Once done with detecting the facial expression with the help of deep neural network we will proceed with that data to predict the stress level of the subject [19,20]. When gathering photos for training, divide them into categories such as Anger, Happiness, Sadness, Surprise, Neutral, and Fear. Also, the dataset should be categorized by emotion class, and any photographs that are uncertain about the emotion should be removed. Train the dataset with the CNN Algorithm and Classifying Emotions until it achieves maximum accuracy (more you train, higher is the accuracy). After training 60000 photos based on 7 categories and applying a facial crop algorithm, we trained the data using a batch size of 64 and epochs of 200 and achieved an accuracy of 89 percent. We also lowered the training loss by using adaboost activation functions, as seen in the results of loss versus epochs.

Detecting facial landmarks (64) points with the help of haarcascadefrontalface.xml and shape predictor.dat file and these files are available open and provided by open CV which help us to detect the facial landmarks and also detect the front face of the human body. Using OpenCV and a trained dataset model, it detects facial expressions and predicts stress levels based on distinct class emotions [21,22]. Bringing all of the phases together to produce the ultimate product of our project [23]. Iteratively test the project for bugs and ensure that the outcome meets the project's criteria.
7. Result and Test Cases

Our aim will be to forecast the seven classes of emotion based on the algorithm that we have suggested and the model that we have developed by putting it to the test with real-world data. These seven types of emotions are examined in relation to three types of emotions (i.e. Weak Stress, No-stress, Strong Stress). The following are the TestCases outcomes after supplying real-world inputs: If the subject is at the second stage of stress. It indicates that the person is experiencing minor stress and should take precautions before it progresses to the final degree. If not addressed, it begins to manifest symptoms such as weariness, headaches, and mental instability, which are all signs that a person is in the last stages of stress. From an implementation standpoint, we must continuously monitor the subject's stress level by collecting an average report on the subject on a weekly, monthly, or even yearly basis based on the subject's conditions, and these data must be provided to the concerned doctor for analysis.

8 Figure 9. Outcome 1: There is no tension, i.e. the subject is not under any stress.

8 Figure 10. Outcome 2: Weak Stress, i.e., if the subject is under light stress, special precautions must be taken.
Figure 11. Outcome 3: As a result, there is a lot of stress. i.e. the subject is stressed and the doctor is concerned. need to be watched with extreme caution.

8. Conclusion
The goal of this study was to use technology to track the subject's stress level by new modern methodology in order to overcome the old school methods to detect the stress. Our project would help in continuous monitor of subject so that exact prediction can be found. These projects would eliminate the Concern of doctor manually since these output data would be monitored by doctor himself at remote location and also eliminates the old school paper methods and body fluid testing methods where subject should be present manually and also those methods are not helpful in continuous monitoring of the subject. Our project is prototype which can be implemented in hardware which have camera access to read real time data and also easy to implement. A map is made up of images of people and landmarks. As a result of the study, we were able to confirm that employing facial landmarks increased stress recognition performance. We also discovered that better detecting stress-related information boosted performance. Our algorithm makes use of CNN architecture and reads the time data from subject face and based on the algorithm we proposed it would help us in detecting the facial expression based on the 7 class of emotion which in turn is manipulated with respect to 3 class of stress detects the stress level of the subject.

References
[1] T. Jeon, H. Bae, Y. Lee, S. Jang and S. Lee, "Stress Recognition using Face Images and Facial Landmarks," 2020 International Conference on Electronics, Information, and Communication (ICEIC), 2020, pp. 1-3, doi: 10.1109/ICEIC49074.2020.9051145.
[2] G. Giannakakis, D. Manousos, V. Chaniotakis and M. Tsiknakis, "Evaluation of head pose features for stress detection and classification," 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), 2018, pp. 406-409, doi: 10.1109/BHI.2018.8333454.
[3] G. Giannakakis, M. R. Koujan, A. Roussos and K. Marias, "Automatic stress detection evaluating models of facial action units," 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020), 2020, pp. 728-733, doi: 10.1109/FG47880.2020.00129.
[4] T. Baltmšaitis, M. Mahmoud and P. Robinson, "Cross-dataset learning and person-specific normalisation for automatic action unit detection in", Automatic Face and Gesture Recognition(FG) 2015 11th IEEE International Conference and Workshops on, vol. 6, pp. 1-6, 2015.
[5] P. Lucey, J. F. Cohn, K. M. Prkachin, P. E. Solomon and I. Matthews, "Painful data: The unbc-mcmaster shoulder pain expression archive database", Automatic Face Gesnrre Recognition and Workshops (FG2011) 2011 IEEE International Conference on. IEEE, pp. 57-64, 2011.
[6] K. Anis, H. Zakia, D. Mohamed and C. Jeffrey, "Detecting depression seventy by interpretable representations of motion dynamics", 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG2018), pp. 739-745, 2018.
[7] D. V. Kishore, Shivashankar, A. C. Ramachandra, and C. Pandurangappa, “Optimization of motorcycle pitch with non linear control,” in 2016 IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology, RTEICT 2016 - Proceedings, 2017, pp. 1656–1660, doi: 10.1109/RTEICT.2016.7808114.

[8] Kishore D.V, Shivashankar, and S. Mehta, “MANET topology for disaster management using wireless sensor network,” in International Conference on Communication and Signal Processing, ICCSP 2016, 2016, pp. 0736–0740, doi: 10.1109/ICCSP.2016.7754242.

[9] A. C. Ramachandra, M. N. Thippeswamy, C. Pandurangappa, and P. Ramesh Naidu, “Modelling and performance evaluation of anti-lock braking system,” Journal of Engineering Science and Technology, vol. 14, no. 5, pp. 3028–3045, 2019.

[10] M. Penna, J. J. Jijesh, and Shivashankar, “Design and implementation of automatic medicine dispensing machine,” in RTEICT 2017 - 2nd IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology, Proceedings, 2017, vol. 2018-Janua, pp. 1962–1966, doi: 10.1109/RTEICT.2017.8256941.

[11] A. C. Ramachandra, M. N. Thippeswamy, C. Pandurangappa, and P. Ramesh Naidu, “Synthesis and modeling of antilock braking system using sliding mode controller,” Journal of Advanced Research in Dynamical and Control System, vol. 10, no. 12, pp. 208–221, 2018.

[12] C. A. Varun, Shivashankar, M. Sahana, R. S. Varun, and T. Rajesh, “Implementation of swarm intelligence in obstacle avoidance,” in RTEICT 2017 - 2nd IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology, Proceedings, 2017, vol. 2018-Janua, pp. 525–528, doi: 10.1109/RTEICT.2017.8256652.

[13] S. B. Sridhara, K. B. Naveen, M. Ramesha, and G. N. Pai, “Internet of things: Internet revolution, impact, technology road map and features,” Advances in Mathematics: Scientific Journal, vol. 9, no. 7, pp. 4405–4414, 2020, doi: 10.37418/amsj.9.7.11.

[14] P. Ramesh Naidu, N. Guruprasad, “Design and implementation of cryptcloud system for securing files in cloud,” Advances in Mathematics: Scientific Journal, vol. 9, no. 7, pp. 4485–4493, 2020, doi: 10.37418/amsj.9.7.17.

[15] M. Ramesha, Sridhara S.B, Naveena Pai G, “FPGA Implementation of Low Power High Speed BTED Algorithm for 8 Bit Error Correction in Cryptography System,” International Journal of Emerging Trends in Engineering Research, vol. 8, no. 7, pp. 3893–3897, 2020, doi: 10.30534/ijeter/2020/158872020.

[16] P. Ramesh Naidu, N. Guruprasad, “A High-Availability and Integrity Layer for Cloud Storage,” Cloud Computing Security: From Single to Multi-Clouds,” Journal of Physics: Conference Series 1921(1), https://doi.org/10.1088/1742-6596/1921/1/012072.

[17] S. B. Sridhara, M. Ramesha, “Recent advances in graph theory and its applications,” Advances in Mathematics: Scientific Journal, vol. 10, no. 3, pp. 1407–1412, 2021, doi: 10.37418/amsj.10.3.29.

[18] Ramesha, M., Sridhara, S.B., Pai, G.N., Patil, S.K., “Design of Antilock Braking System Based on Wheel Slip Estimation,” Journal of Physics: Conference Series 1706(1),2020, https://doi.org/10.1088/1742-6596/1706/1/012216.

[19] Pai, G.N., Sridhara, S.B., Shashidhara, K.S., Gangadhara, “Signal Analysis and Filtering using one Dimensional Hilbert Transform,” Journal of Physics: Conference Series 1706(1),2020, https://doi.org/10.1088/1742-6596/1706/1/012107.

[20] M. Ramesha, K. Jeevan and B. M. Sathisha, “Implementation of IoT Based Wireless Electronic Stethoscope,” 2020 Third International Conference on Multimedia Processing, Communication & Information Technology (MPCIT), 2020, pp. 103-106, doi: 10.1109/MPCIT51588.2020.9350476.

[21] Puneeth Kumar M. V, Naveena Pai G, Vinay Kumar B. C, "Dynamic Analysis and Control Strategies of an Anti-lock Braking System," 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), 2020, pp. 1677-1682, doi: 10.1109/ICECA49313.2020.9297642.

[22] G. Naveena Pai, M. Swathi Pai, M. Shruthi and B. Naveen K, "Internet of Things: A Survey on Devices, Ecosystem, Components and Communication Protocols," 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), 2020, pp. 611-616, doi: 10.1109/ICECA49313.2020.9297458.

[23] M. Ramesha, Jeeven K.M, “Study and analysis of bted error correction codes for cryptography applications,” International Journal of Advanced Trends in Computer Science and Engineering, vol. 9, no. 5, pp. 8938–8942, 2020, doi.org/10.30534/jitcse2020/293952020.