Multi-modal Emotion Recognition for Determining Employee Satisfaction

Farhan Uz Zaman, Maisha Tasnia Zaman, Md. Ashraful Alam, Md. Golam Rabiul Alam
Department of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
e-mail: (farhanuzzaman32,maishatasniazaman)@gmail.com; (ashraful.alam, rabiul.alam)@bracu.ac.bd

2021 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE) | 978-1-6654-9552-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/CSDE53843.2021.9718373

***Abstract*- Emotion recognition has been popular in the field of research for quite a while now. In this paper bi-modal emotion detection has been used to find employee satisfaction in workplaces. Interviews were taken of employees from different workplaces and were recorded. The recorded interviews were then used to detect the emotions of the employees from which their satisfaction level was derived. From the interviews, six different entities of emotion were detected, which are: Happiness, Sadness, Neutral, Disgust, Anger and Surprise. Two separate independent neural networks have been utilised. In one the facial expressions were detected and in the other sentiment analysis was done on the speech of the interviews after converting it into text. The extracted emotions were then fed into a Support Vector Machine (SVM) for determining the satisfaction of the employees. The satisfactions were categorized into five different levels which are: Highly dissatisfied, Dissatisfied, Neutral, Satisfied and Highly satisfied.**

***Index Terms*—Deep learning, Emotion recognition, Neural Network, Support Vector Machine, Sentiment Analysis.**

I. Introduction

Employee mental health is one of the most important factors for productivity of any organization. This is intertwined with the satisfaction level of the employees with their respective workplaces. This research was done to determine how satisfied the employees are in their workplaces with the help of deep learning and machine learning techniques. Although emotional recognition has already been applied in several different sectors, the importance of comprehending the emotional states of employees has not been emphasized on. Therefore, this research focuses on capturing the genuine emotional states of employees regarding their work. The satisfaction of a certain employee towards their work has a major impact on their individual performance and as a result the overall success of an organization. Most mental health professionals have studied that the workplace environment plays a vital role in an individual's mental well-being and vice versa [1]. According to the World Health Organization (WHO), mental health is a state of well-being that causes every individual to understand his/her potential. It depends on his/her mental health whether they are able to adjust to regular stresses of everyday life, productivity and thus their contribution in the workplace is determined. WHO also found poor mental health of over 300 million employees around the globe causes the global economy to lose one trillion USD in productivity every year [2].

Convolutional Neural Network (CNN) [3] was used to find facial expressions to detect the emotion from the interviews. The speech of the interviewees in the interview was then converted into text, using a very efficient python script which calls google speech to text converting API, using which sentiment analysis [4] was carried out, which was the second mode for detecting emotions of the employees. Two modes were used instead of one for finding the actual emotions of the interviewees with better accuracy. The results from the individual modes are then provided to a Support Vector Machine [5] to get the satisfaction of the employees.

II. Existing Works

There has been significant progress in developing emotion recognition using EEG with Deep Recurrent Neural Networks, critical frequency bands and channels has been investigated using EEG-based emotion recognition models for three emotions: positive, neutral and negative [6]. ECG signals have also been used effectively in application of emotional recognition. In terms of valence, arousal and dominance, EEG and ECG signals were recorded during effect elicitation by means of audio-visual stimuli [7]. In order to represent the characteristics associated with emotional states, a new effective EEG feature referred to as "Differential Entropy" was introduced and it was confirmed that EEG signals on frequency band gamma related to emotional states more closely compared to other frequency bands [8]. As emotion is a subjective matter, gathering knowledge or the science behind labeled data has been quite challenging for several years. However, with the evolution of deep learning, it has been in the research for a long time. Facial recognition using CNN has been the most efficient model and has given quite good results on image classification problems [9]. Therefore, it has proven to be successful mainly in terms of detecting facial expressions from images. CNN layers are structured using multiple layers. On each layer, groups are formed based on the features that they work on and these groups feed forward to the next layer. At the end, it connects to the fully connected layer [10]. Nonetheless, using only facial expressions to identify emotions is not authentic enough. Speech emotion recognition systems have been used. A recurrent neural network was initially used to classify six different emotions. Their performances were then compared to multivariate linear regression (MLR) and support vector machine techniques [11] . Machine learning algorithms have also been proven to be effective and reliable for opinion mining as well as sentimental classification [12]. The use of emotion recognition has been so far used in different applications. Unfortunately, emotional states of employees and it's impacts have not significantly been highlighted in most research. It has been determined that job crafting predicted intrinsic need satisfaction which as a result predicted the mental well- being of employees in a research [13]. The importance of employee mental health has been highlighted by the World Health Organization multiple times and it's consequences leading to not only very poor performance in the workplace but also an overall global impact [14].

III. Proposed Model

We decided to use two Neural Networks for strongly separated architectures, each of the networks works independently on its own domain.The final decision is made by putting together the outcome of the neural networks, which are commonly called expert networks or agents, into a machine learning algorithm for combining the separate results and finding the satisfaction of the employees. The modular architecture combines the two different outputs from the neural networks, this enhances the overall generalization which could be significant in a high dimensional space [15]. Interviews were conducted and were recorded in the Mp4 format. The recording frames were used for facial expression recognition (FER). And the audio was converted into text for sentiment analysis.

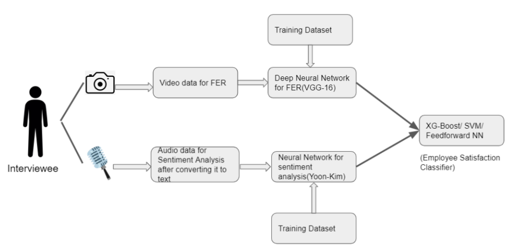

Fig. 1. Proposed Model

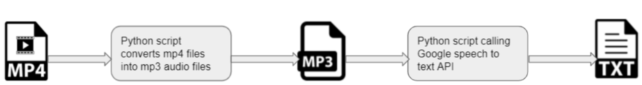

Fig. 2. Deriving text from interviews for Sentiment Analysis

IV. Implementation

We conducted interviews of different employees from various professions, not focused to any particular profession or age range. The interviews consisted of questions where we have asked people regarding their jobs according to the questionnaire which is discussed in details below. The interviews were recorded as mp4 videos. In total we conducted 160 interviews. The 80% of the interviews were then used to train and the rest 20% are used for testing.

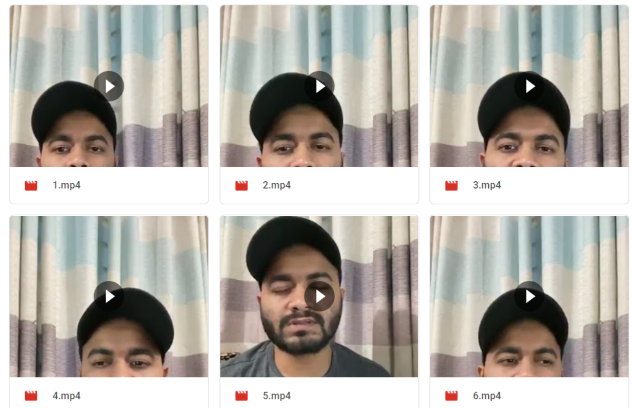

Fig. 3. Example of video dataset of interviews

A. *Questionnaire Development*

A set of questionnaires has been developed, studying the different reasons and areas of a workplace that are capable of affecting the emotional state of an employee [2]. One of the most obvious factors is the environment of the workplace which also include the health and safety policies. It is also essential that the individual finds their work to be enjoyable and meaningful as finding their work monotonous or feeling that they are performing unpleasant tasks results in poor emotional state. An employee's emotional state is largely impacted by how he/she is treated by their supervisor/manager [1]. Moreover, whether the individual is being able to cope with the work pressure or not is an important issue. Another factor that affects an employee's mental state is the system of promotion. Additionally, if an employee is not pleased with his/her co-workers/colleagues, this majorly affects their emotional state. Lastly, the salary of every employee turns out to be either work as a huge inspiration to work or a discouragement depending on whether or not they are satisfied with it [16]. Based on these criteria we prepared the following set of questionnaire:

- How is your working environment?
- How enjoyable and meaningful do you find your work to be?
- Are you satisfied with how your supervisors treat you?
- How stressful do you find your work to be?
- How satisfied are you with your system of promotion and your content of work?
- Do you have complaints regarding your co-workers and/or working in groups?
- What do you feel when you come to the office?

- Are you satisfied with your salary?

The emotional state of the employee is judged according to the answers that are received. The answers are received in the form of mp4 video. The speeches are converted to texts and the videos are converted into a stream of frames. Those inputs are then used for facial recognition using a neural network and sentimental analysis. The final result of the emotional state is derived by judging the outputs of the two neural networks according to the Likert Scale range [17], [18] with the help of a Support Vector Machine, SVM. Likert scales are used to evaluate statements of agreement, frequency, quality, importance, and likelihood, etc. We used it in order to understand the intensity of the emotional state expressed by the employees. In other words it measures the level of mental satisfaction of the employee. The scale ranges from 1 to 5 as follows:

1) Highly dissatisfied
2) Dissatisfied
3) Neutral
4) Satisfied
5) Highly Satisfied

B. Dataset Preparation

We have interviewed 160 different employees from various job sectors starting from employees working in grocery markets as salespeople to employees who work as managers in their company, including doctors, teachers, engineers, bankers, etc. Ensuring diverse backgrounds. We have ensured is that our video recordings are done only after getting the consent of the interviewees. Due to privacy issue the dataset cannot be exposed, since some of our interviewees are insecure whether this might affect their professional lives.

Fig. 4. Glimpse of the Interviewees

V. Sentiment Analysis

Sentimental analysis is done using different types of neural networks. CNN [18], [19], GRU [20] and LSTM [21] have been analysed. All were trained using Twitter Us Airline Sentiment [22]. This dataset is public and can be downloaded by anyone. Results from the neural networks are then compared and the one with the highest accuracy is used in the final model to derive the actual result.

A. Long Short Term Memory

LSTM is a very popular version of RNN in the field of sentiment analysis. Result from this test is provided below.

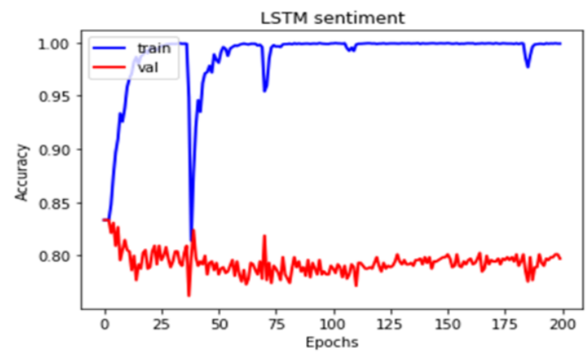

Fig. 5. LSTM training graph of Accuracy against Epochs

B. Gated Recurrent Unit

GRU is a very modern version of RNN in the field of sentiment analysis. Result from this test is shown below.

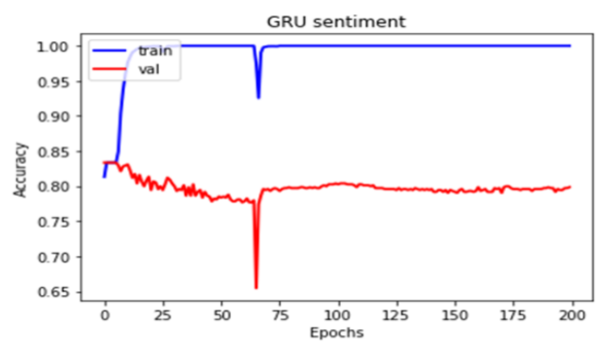

Fig. 6. GRU training graph of Accuracy against Epochs

C. YOON KIM'S Model of CNN

YOON KIM's Model of CNN [23] is a fastest model in the entire experiment having showed great efficiency. Having the highest accuracy. The result derived from this model is given below.

D. Comparison of Accuracy

After training three different models, it was time to select the one which would be used in the final model. All the models were trained with an epoch number of 200 which can be seen from the training graphs. This was done so that valid comparison could be drawn easily and effectively.

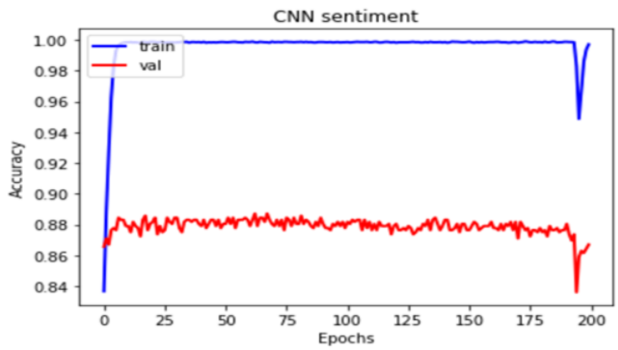

Fig. 7. Yoon Kim's CNN training graph of Accuracy against Epochs

The comparison was made by looking at the accuracy of the models. The one with the highest accuracy has been chosen and put into the final model. It was concluded that Yoon Kim's CNN would be used in the final model since it has highest accuracy as shown in the table, which was a little unexpected as commonly RNNs are considered the reliable forms of neural network when it comes to sentiment analysis.

Models	Accuracy in %
LSTM	73
GRU	77
Yoon Kim's CNN	80

TABLE I
ACCURACY COMPARISON OF SENTIMENT ANALYSIS.

VI. FACIAL EXPRESSION RECOGNITION

For image analysis such as facial expression recognition, the most trusted forms of neural network are the ones that belong to the family of CNN. They perform their task using the convolution operations, after a suitable number of operations have been carried out, the neural networks learn how to differentiate the images it has been trained upon. For determining the emotions of the employees we have trained our CNN architectures using Extended Cohn-Kanade Dataset (CK+) [24], Multimedia Understanding Group (MUG) Dataset [25] and Acted Facial Expressions in the Wild Database (AFEW) [26] using the labels "angry", "disgust", "happy", "sad", "surprise" and "neutral". There are a total of 28,709+x1 images for training and 3589+x2 number of test images. We have used 2 CNN architectures to classify the human expression, and in the end selected the one with the highest accuracy just as in the case of sentiment analysis.

A. VGG16

VGG16's architecture [27] is rather simple given how powerful it is, it follows a few continuous blocks of 2-3 layers of convolution layers followed by a pooling layer, the last 3 layers are however Dense layers. After 100 epoch and 6 hours of training the accuracy stopped improving and the graph became horizontally flat which can be seen from the results from this are provided below in graphical form.

Fig. 8. Different Entities of Emotion

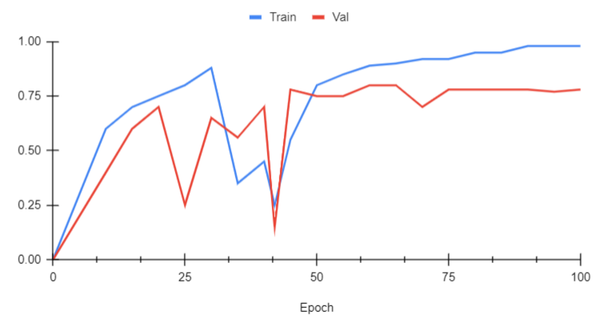

Fig. 9. VGG16 training graph of Accuracy against Epochs

B. MobileNet

MobileNet [28] is highly optimized for low powered systems such as mobile ,embedded machines and other machines powered by SoCs, it has a lower amount of parameters and a smaller complexity, it achieves these feats without loss in significant accuracy using Depthwise Separable Convolution. the acheived results from this is given below.

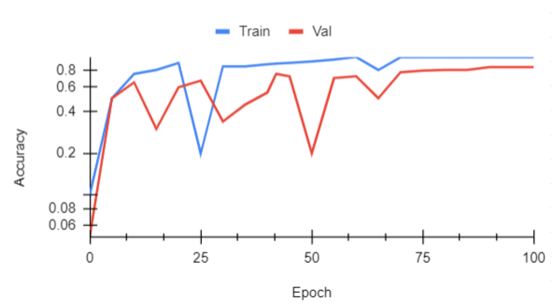

Fig. 10. MobileNet training graph of Accuracy against Epochs

C. Comparison of Accuracy

MobileNet despite being way smaller than VGG16, in terms of srchitecture, gave a relatively similar result. From the table below we decided to choose VGG16 which was marginally better in terms of accuracy.

The results from the independent agents were then fed to the decision making module, SVM. The output from this module is the Likert Scale that comes in the form of integers from 1 to 5 where 1 stands for least satisfied and 5 signifies very satisfied

Models	Accuracy in %
VGG16	76
MobileNet	75

TABLE II
ACCURACY COMPARISON OF FER.

as mentioned above. A simple feedforward neural network was tried at the beginning. However, it was overfitting, since we had only 160 data which was insufficient to train a neural network, thus the results were very poor. On the other hand the result from the SVM were very satisfactory. The confusion matrix below shows result of the SVM.

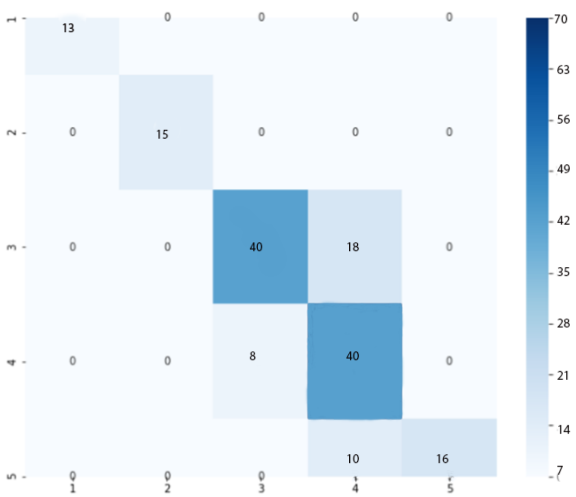

Fig. 11. Confusion Matrix of SVM

VII. CONCLUSION

Emotion detection for employee satisfaction is a major application in the real world since mental health of employees at workplaces is a key factor. This paper has thrown light upon how modern technology can be used for the study of this aspect. The aim of this research is to improve the working environment as well as other aspects which are affecting the mental health of employees all over the world. Machine Learning and Neural Networks which have provided fast answers to various questions can now be used to answer "How satisfied employees are at their respective workplaces." This will enhance understanding of human psychology and can be used for better productivity. This research is the very first of its kind. It has many lackings such as not focusing on any particular age group or profession which could have given more accurate results. We plan to collect more data so that the decision making module could also be a deep learning architecture which probably would bring higher accuracy.

REFERENCES

[1] G. R. Slemp and D. A. Vella-Brodrick, "Optimising employee mental health: The relationship between intrinsic need satisfaction, job crafting, and employee well-being," *Journal of Happiness Studies*, vol. 15, no. 4, pp. 957–977, 2014.

[2] G. Ainomugisha, "Why employee mental health should be every manager's concern," https://inside.6q.io/employee-mental-health-managers-concern/.

[3] H. Yar, T. Jan, A. Hussain, and S. Din, "Real-time facial emotion recognition and gender classification for human robot interaction using cnn."

[4] C. Dos Santos and M. Gatti, "Deep convolutional neural networks for sentiment analysis of short texts," in *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, 2014, pp. 69–78.

[5] N. Cristianini, J. Shawe-Taylor *et al.*, *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000.

[6] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks," *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 3, pp. 162–175, 2015.

[7] S. Katsigiannis and N. Ramzan, "Dreamer: A database for emotion recognition through eeg and ecg signals from wireless low-cost off-the-shelf devices," *IEEE journal of biomedical and health informatics*, vol. 22, no. 1, pp. 98–107, 2017.

[8] R.-N. Duan, J.-Y. Zhu, and B.-L. Lu, "Differential entropy feature for eeg-based emotion classification," in *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 2013, pp. 81–84.

[9] P. R. Dachapally, "Facial emotion detection using convolutional neural networks and representational autoencoder units," *arXiv preprint arXiv:1706.01509*, 2017.

[10] B. Zhang, C. Quan, and F. Ren, "Study on cnn in the recognition of emotion in audio and images," in *2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)*. IEEE, 2016, pp. 1–5.

[11] L. Kerkeni, Y. Serrestou, M. Mbarki, K. Raoof, M. A. Mahjoub, and C. Cleder, "Automatic speech emotion recognition using machine learning," in *Social media and machine learning*. IntechOpen, 2019.

[12] M. Ahmad, S. Aftab, S. S. Muhammad, and S. Ahmad, "Machine learning techniques for sentiment analysis: A review," *Int. J. Multidiscip. Sci. Eng*, vol. 8, no. 3, p. 27, 2017.

[13] S. der Kinderen and S. N. Khapova, "Positive psychological well-being at work: The role of eudaimonia," *The Palgrave Handbook of Workplace Well-Being*, pp. 141–168, 2021.

[14] G. Harnois and P. Gabriel, "Mental health and work: Impact, issues and good practices," 2000.

[15] A. Schmidt, "A modular neural network architecture with additional generalization abilities for high dimensional input vectors," *Manchester Metropolitan University, Department of Computing*, 1996.

[16] K. Matzler and B. Renzl, "The relationship between interpersonal trust, employee satisfaction, and employee loyalty," *Total quality management and business excellence*, vol. 17, no. 10, pp. 1261–1271, 2006.

[17] J. Parkinson, "Measuring positive mental health: Developing a new scale," *NHS Health Scotland, Glasgow*, 2006.

[18] R. Johnson and T. Zhang, "Effective use of word order for text categorization with convolutional neural networks," *arXiv preprint arXiv:1412.1058*, 2014.

[19] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A convolutional neural network for modelling sentences," *arXiv preprint arXiv:1404.2188*, 2014.

[20] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555*, 2014.

[21] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[22] "Twitter us airline sentiment," https://www.kaggle.com/crowdflower/twitter-airline-sentiment, Jul 2016.

[23] Y. Kim, "Convolutional neural networks for sentence classification. corr abs/1408.5882 (2014)," *arXiv preprint arXiv:1408.5882*, 2014.

[24] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," in *2010 ieee computer society conference on computer vision and pattern recognition-workshops*. IEEE, 2010, pp. 94–101.

[25] N. Aifanti, C. Papachristou, and A. Delopoulos, "The mug facial expression database," in *11th International Workshop on Image Analysis for Multimedia Interactive Services WIAMIS 10*. IEEE, 2010, pp. 1–4.
[26] A. Dhall, R. Goecke, S. Lucey, and T. Gedeon, "Acted facial expressions in the wild database," *Australian National University, Canberra, Australia, Technical Report TR-CS-11*, vol. 2, p. 1, 2011.
[27] H. Qassim, A. Verma, and D. Feinzimer, "Compressed residual-vgg16 cnn model for big data places image recognition," in *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, 2018, pp. 169–175.
[28] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.