FACIAL EMOTION RECOGNITION AS SPATIAL IMAGE USING GABOR FILTER

Shubham Luharuka, Meghana R, Pallavi K J and Asha S Manek

Department of Computer Science and Engineering,
RV Institute of Technology and Management, India

ABSTRACT

The state of mind of a person can be easily understood from the human face. This paper proposes a methodology to recognize facial expression using Gabor filters, ResNet and two other custom models. The image is taken as the input data from the camera. This input can be used to extract information to infer a person's mood. First, we develop an algorithm for detecting image of an individual from entire set of images using Haar Cascade face detection algorithm. Then, we apply Gabor filter for extracting facial features in the spatial domain. Using the Gabor filter can effectively reduce computation and size, and in some situations even improve recognition. Gabor filters are used to capture the entire frequency spectrum in all directions. Finally, facial expressions are successfully classified by proposed Convolutional Neural Network model using extracted important facial features from the facial image after applying Gabor filter as input. The results of testing images from the CK+ dataset show the reliability and the best recognition rate of the proposed method.

KEYWORDS

CK+ Dataset, CNN Model, Data Augmentation, Facial Emotion Recognition, Gabor Filter, Normalization of Images.

1. INTRODUCTION

Emotions are psychological states based on individual thoughts, behaviours and actions. They are mainly conspicuous methods used for conveying purpose and internal feelings. They can be a verbal or non-verbal representations. Facial Emotion is nonverbal representation and scientists are doing research to discover a new and efficient way to predict facial emotions. The human facial actions can be classified using "Action Units (AU)" with the help of "Facial Action Coding System (FACS)" structure [1]. In today's world, Facial Expression software is used by game developers to know how the user feels while playing their games. Companies are allowed to conduct market research and measures moment-by-moment facial expressions/emotions using Facial emotion recognition software. This automatic extraction of facial emotion measures helps companies to aggregate results and analyse the market products/services. Hoteliers and Restaurant Owners use this software to take feedback from customers. Electroencephalography (EEG), Electrocardiography (ECG), Galvanic Skin Response (GSR) etc., are various methods available to recognize emotions [2]. But these systems require costly hardware and a complex process. It is also true that no one can show false expressions or cheat these processes.

This research work aims to build the software which can easily find expressions. The Computer Vision is used for the same to understand the muscle movements of the face and predict emotions. Emotions can be represented as 2D vectors. X-axis and Y-axis represent Valence and

Arousal respectively. In Figure 1 shows a neutral point at the origin. The value of the X-axis increases if emotion tends to be happiness and Y-axis if it tends to excited state.
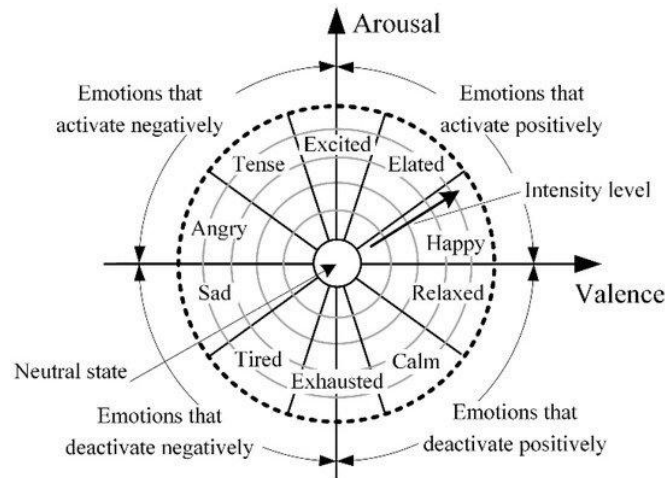

Figure 1. 2D Vector Representation of Emotions

The remainder of the paper is organized as: Section II explains the related work, Section III provides the Methods and Techniques, Section IV describes System Design and Implementation. Section V shows Experiment and Evaluation of proposed work. Performance Analysis is explained in section VI and Section VII concludes the paper.

2. RELATED WORK

Sarvakar et al., [3] presented FERC system using neural networks coevolutionary face emotion detection method. The system developed also included decision making tree technique for determining the high and low percentage of emotions. The proposed CNN model achieved 54% accuracy with split ratio of 80:20 percentage. Wu et al., [4] explained that "Gabor motion energy filters (GME)" could be used to predict "facial expressions". This paper concentrates on generating feature descriptors such as pose, illumination, scale, occlusion and rotation using Local Directional Covariance Matrices (LDCM) method.

Salama et al., [5] proposed a multimodal human emotion recognition framework from the "electroencephalogram (EEG)" signals, and the video data of human faces by extracting the spatiotemporal features. This paper used final fusion predictions with fusion of data augmentation and ensemble of classifier techniques to achieve accuracy of 96.13% and 96.79% for valence and arousal classes respectively. Islam et al., [6] classify the images using SVM by converting coloured images to Grayscale. The converted grayscale images then segmented the facial parts like nose, eyes and lips. The Gabor filter is applied after segmentation to extract features. These features are used to train the SVM classifier. In this paper, different databases such as JAFFE, CK+, RaFD are used to get 94.89%, 99.57%, 99.50% accuracy respectively.

Ramos et. al., in paper [7] proposed a "Haar-Cascade classifier" for face detection. "Eigenfaces API" and "Gabor filter" is used for feature extraction. The model is trained using Support Vector Machine algorithm and obtained 80.11% accuracy. The model is tested by observing emotions in a lecture with teaching pedagogies. It analyses from the student's facial expressions that adaptive learning methods can be applied in lecture class. Lee et al. [8] recommends two various variants of "EmotionNet Nano", a CNN model trained for 200 epochs. "EmotionNet Nano" model has achieved an accuracy of 97.6% by using CK + dataset on an ARM embedded processor.

In paper [9], Cheng et al. explained a shallow CNN. It is an improved version of "VGG-19". "Alex-Net", improved "VGG-19" and "deep CNN" are used for training and analysing facial expression data on the Extended "Cohn–Kanade" database. The experimental results indicated accuracy of 96% in recognition of facial expression. Aamir, Muhammad, et al proposed a multilevel "deep CNN" to identify "facial expressions" with their intensity [10]. First level named as "Expression-Net" and "Intensity-Net" as second level used to classify face expressions and estimating the intensity of "facial expression" respectively. The model achieved 98.8% and 97.7% accuracy using the extended "Cohn–Kanade" and "Japanese Female" Facial Expression datasets respectively.

Shi, M. et al., introduced "Facial expression recognition (FER)" into fields such as safe driving, medical care, distance education and marketing assistance. But the training model have faced some problems such as low recognition rate and low training speed which affected for real production process [11]. This paper presents FCM to enhance the "feature extraction (FE)" capacity of the model, and proposes a FER algorithm utilising a further developed CNN (F-CNN) to solve the problems faced by earlier CNN model and achieved an accuracy of 83.86%.Mohammed et. al. proposed a model that classifies the emotions of facial expression into different types, natural, smile, surprise, disgust, squint and scream by using the deep CNN model. The "Principle Component Analysis (PCA)" method is used for extracting features and K-NN for classifying the "Extended Cohn-Kanade (CK+)" and the "Japanese Female Facial (JAFFE)" datasets and achieved 98.5% accuracy [12].

In paper [13], Wong et. al. proposed a system that determines the optimal number of clusters by applying EM clustering technique for the collection of 3000 playlist from radio station. Analysing the resulting clusters and all playlist data, the research paper concludes that the music in particular era is popular in that particular time period. Further, the various "data visualization" techniques and "serendipity recommendation" systems are delivered in this paper. In paper [14], Sen et. al. proposed a user intuitive smart music player which acquires the "facial expression" of a person working on system and achieved 83% accuracy on testing data. The software developed using Viola Jones algorithm on JAFFE dataset. K-Nearest Neighbours (KNN) classifier is used for classification of songs into various genres.

3. METHODS AND TECHNIQUES

In this paper, the "Gabor filter" is used to propose the novel model. The dataset used for this proposed method is the CK+ dataset. We are classifying emotion into seven classes that are happy, sad, disgust, surprise, anger, contempt and fear. Gabor Filter is the feature extractor from a facial image like moved muscles in different directions. A CNN classifier is used after obtaining images after applying Gabor Filter.

3.1. Dataset Description

In proposed model, the CK+ dataset [15] is used, which consists of 7 types of emotion as "ANGER", "CONTEMPT", "DISGUST", "FEAR", "HAPPY", "SADNESS" and "SURPRISE". The description of the CK+ dataset is as shown in Figure 2. The total number of images used in this experiment is 327. The 327 images used are not enough to train CNN model. Therefore, we have applied Gaussian noise to make a better CNN model and to increase the image dataset.

3.2. Gabor Filter as Feature Extractor

Classifiers depend on the features available to do their task, therefore, feature extraction is a principal process. The proposed model uses a 2D Gabor filter for feature extraction, which is a "Gaussian kernel function" modulated by a complex sinusoidal wave. Mathematically shown in equation (1) and equation (2) below,

$$G_s[i,j] = Ce^{\frac{-(i^2+j^2)}{2\sigma^2}} \sin(2\pi f(i \cos\theta + j \sin\theta)) \dots\dots\dots\dots\dots\dots\dots\dots\dots\text{eq (1)}$$

$$G_c[i,j] = Be^{\frac{-(i^2+j^2)}{2\sigma^2}} \sin(2\pi f(i \cos\theta + j \sin\theta)) \dots\dots\dots\dots\dots\dots\dots\dots\dots\text{eq (2)}$$

3.3. Normalization of Images

The purpose of Normalization is to bring an image to the range that is normal to sense. Normalization in Image processing means each pixel of an image should be in the range of (Min, Max). It can be done with both linear and non-linear types of data. Suppose we have to normalize an n-dimensional Greyscale image. This means the current Image can be written as in equation (3),

$$I:\{X \le R^n\} \to \{currMin, \dots, currMax\} \dots\dots\dots\dots\dots\dots\dots\dots\text{eq (3)}$$

where currMax is the current range of pixel intensity.

Normalized image will be written as in equation (4),

$$I_n:\{X < R^n\} \to \{newMin, \dots, newMax\} \dots\dots\dots\dots\dots\dots\dots\dots\text{eq (4)}$$

The linear normalization on a Greyscale image is done by applying the equation (5) below,

$$I_n = (I - currMin)\frac{newMax - newMin}{currMax - currMin} + newMin\dots\dots\dots\dots\dots\dots\dots\dots\text{eq (5)}$$

And the non-linear normalization follows a sigmoid function then it can be written as shown in equation (6),

$$I_n = (newMax - newMin)\frac{1}{\left(1+e^{\frac{-1+\beta}{\alpha}}\right)} + newMin \dots\dots\dots\dots\dots\dots\dots\dots\text{eq (6)}$$
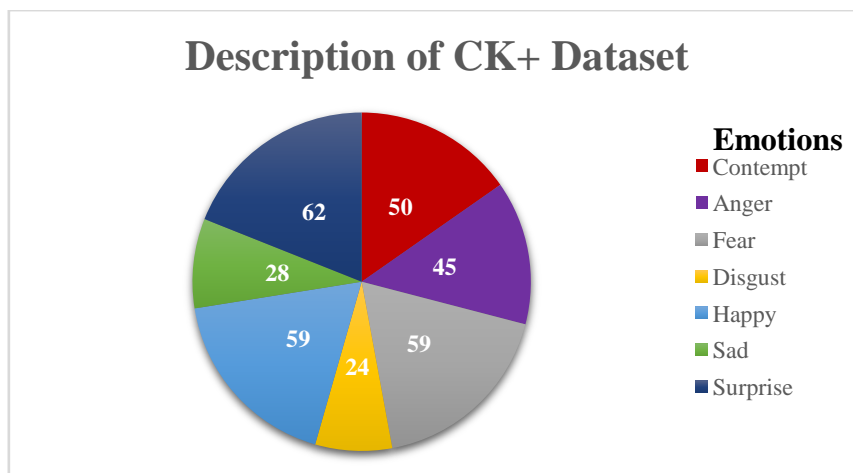
Figure 2. Description of CK+ Dataset

3.4. Data Augmentation

In machine learning, many times we face a shortage of data in the datasets. Our ML model should be able to predict different types of data belonging to the same class. The performance of deep learning "Neural Networks" improves with the amount of data available, but it should not go into the case of overfitting. In simple terms, we can say that Image data augmentation means generating different versions of images by doing some manipulation like shift, flip, zoom, adding noises in the training dataset that belongs to the same class as the original image. In this paper, we have tried to get better accuracy in emotion classification by increasing dataset with adding noise.

4. SYSTEM DESIGN AND IMPLEMENTATION

4.1. Image Filtration using Gabor Filter method

In the proposed work, we have tried to gain more accuracy after Gabor Filter is applied. In our proposed model we have applied Gabor Filter only one time. After applying Gabor Filter, used ResNet 50, VGG16 and developed two custom models. In these two custom models, one model is denser than the other. After loading the dataset, a "face detection algorithm", the "Haar-Cascade" classifier is used to get the image and after cropping the face from the image, Gabor Filter is applied. Here, Gabor filter is used as it gives high output response at the edge and at points where texture changes. Gabor filter has 5 parameters. The table 2 for this parameter and its definition is given below.

Table 2. Parameters used for Gabor Filter.

| Parameters	Definition
K= (x, y)	Gabor kernel's size.
λ	Sinusoidal factor's wavelength in the equation. Higher value means wider ellipse.
θ	The orientation of the Gabor function from the normal to the parallel stripes. It controls the rotation of the ellipse.
ψ	Phase shift of sinusoid.
σ	The Gaussian part's standard deviation.
γ	Spatial aspect ratio of the ellipse

The process is visualized as shown in Figure 3:
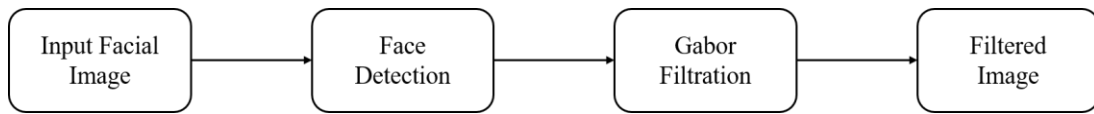
Figure 3. Block Diagram of the Proposed Model

In proposed model, a Gabor filter is applied with the values,

K (x, y) =20, λ=5, θ=3*π/4, ψ=0, σ=1.6, γ=1.5

For data augmentation, we added noise of type gaussian with mean=0.1 and variance=0.1. Sample of image after applying the filter and noise is as shown in Figure 4.

4.2. Proposed CNN Model

The model summary of the proposed "CNN network" is as shown in Table 3. After Data pre-processing, the image is passed through all layers of the proposed CNN model. CNN model will return a vector with seven modes as output. "ANGER", "CONTEMPT", "DISGUST", "FEAR", "HAPPY", "SADNESS" and "SURPRISE" are the seven modes.

As in the proposed method, the image is passed through two adjacent "Convolutional Neural Network" first by applying a convolution of a 32 x 32 and 64 x 64 filter on the image. Next, MaxPooling layer is applied. It reduced the dimension to 63x63x64. Next, again a convolutional network is applied to the 128 x 128 filter size and then a MaxPooling layer to reduce dimension to 31 x 31 x 128.

Figure 4. Sample image after applying Gabor filter and noise.

In the next step, flatten layer is used and it converts all data to a vector of the size 123008. Then, again a dense layer is used to convert the vector of size 123008 to the vector of length 1024 and then again changed to 512 and finally, it is reduced to seven, which show the 7 different categories of emotional states.

Table 3. Type of Layer, their respective parameters and output shapes.

Type of Layer	Output Shape	Parameters
Conv (activation=ReLU)	126,126,32	2432
Conv (activation=ReLU)	126,126,64	51264
MaxPooling	63,63,64	0
Conv (activation=ReLU)	63,63,128	73856

Dropout	63,62,128	0
MaxPooling	31,31,1128	0
Flatten	123008	0
Dense (activation=ReLU)	1024	125961
Dropout	1024	0
Dense (activation=ReLU)	512	524800
Dropout	512	0
Dense (activation=ReLU)	7	3591
Total params: 126,617,159 Trainable params: 126,617,159 Non-trainable params: 0		

The SoftMax and ReLU activation functions used are as shown in equation 7 and equation 8 given below:

SoftMax Function: $\sum_i X_i = 1 \qquad x_i = [0,1]$ ……………………………eq (7)
Rectified Linear Unit: $f(x) = x^+ = \max(0, x)$ ……………………………eq (8)

Finally, the overall method is to first crop the image of the face by using a Haar-Cascade classifier and then apply Gabor Filter by adding noise to images for data augmentation. Then the processed images are fed as input to the CNN model as shown in block diagram Figure 5.
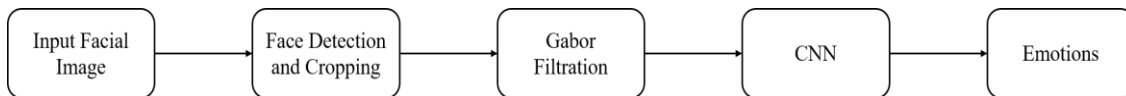

Figure 5. Block Diagram of Proposed Algorithm

5. EXPERIMENT AND EVALUATION

For the experimental result, CK+ dataset is used. Almost 50 images from each class are used for training purpose. For data pre-processing, first resizing of image is done. A Gabor Filter followed by normalization function is applied. Addition of gaussian noise in images completed the data augmentation process. We tested the proposed model on the normal PC with Intel Core i5 and 8 GB of RAM. The model used and accuracy gain while training is as tabulated in Table 4.

Table 4. Models used and accuracy gained while training.

Model Used	Gabor Filter	Augmentation	Training Accuracy (%)	Validation Accuracy (%)
ResNet	False	False	62.12	52.14
VGG16	True	False	70.26	49.23
Proposed Model	True	False	98.52	78.60
	True	True	99.51	96.78

True and False indicates whether that feature is applied or not. 10 per-cent of total dataset is used for validation. Around 99.5% of accuracy is achieved when we train the proposed CNN model with the CK+ Dataset after applying Gabor Filter and noise. 91% of test accuracy is achieved when this model is tested with 10% of the dataset. The proposed model is taking 50 sec (without GPU) to complete one epoch. The proposed model achieved 98.5% of training accuracy without

any data augmentation. Table 5 show the accuracy achieved by proposed model in count of epochs and time taken to train the model itself.

Table 5. Accuracy achieved by the model in epochs and time taken by model.

Accuracy (%)	Epochs	Time (sec)
15	1	51
50	6	311
75	8	412
99.5	21	1056

Total time taken to train the model itself is around 1056 sec and 99.5% of accuracy reported at end of 21st epoch as shown in Figure 6 and 7. In the above graphs as shown in Figure 6 and Figure 7, blue colour of plotting line indicates training parameter and orange colour plotted line indicates validation parameter.

6. PERFORMANCE ANALYSIS

6.1. Comparison of proposed system with work [12]

In research work [12], "Principle Component Analysis (PCA)" technique is used for extracting features and "K-nearest neighbour (KNN)" for classification. Experiments are carried out on the "Extended Cohn-Kanade (CK+)" and the "Japanese Female Facial Expression (JAFFE)" datasets achieved accuracy of 98.5%. With the CNN and Gabor filter, the proposed approach outperformed with 99.5% accuracy as shown in Table 6.
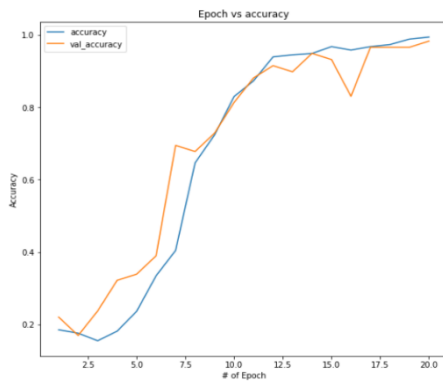

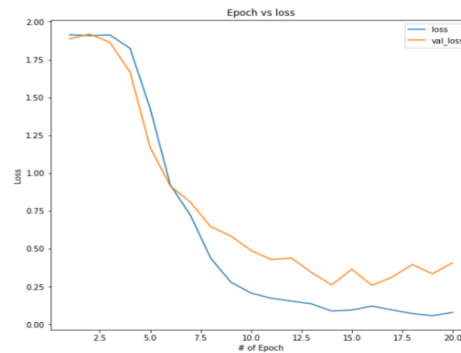Figure 6. A graph illustrating no. of epoch and accuracy attained.

Figure 7. A graph illustrating the no. of epoch and loss.

Table 6. Comparison of Proposed Model with Reference [12]

Model	Proposed Model	Comparison with Work [12]
Dataset	Extended Cohn-Kanade (CK+)	Extended Cohn-Kanade (CK+)
Feature-Extraction Method	Gabor Filter	Principal Component Analysis (PCA)
Best- Classifier	Convolution Neural Network	K-Nearest Neighbour (KNN)
Accuracy (%)	**99.5**	**98.5**

6. SIGNIFICANCE

Facial Expression Recognition can be used to diagnose mental diseases and detect human social interactions. It can also be used for music therapy which helps people suffering from anxiety, depression and stress. Emotion recognition can monitor citizens in crowds for suspicious behaviours. But the most popular application is in marketing where the companies can know how the customers respond to advertisements. A similar application is in the education industry, where feedback of lecturers and educational content can be recorded by analysing the facial expressions and detecting the emotions from student's faces. Driving in extreme emotional state could be harmful and cause accidents, so it has proven useful in accident prevention by alerting the driver in such cases. A person's facial emotion can give away their intentions. So, it plays a key role in investigation of deception (fraud).

7. CONCLUSION

This paper presents an experiment to predict emotion from spatial image of face. Improved accuracy of the system is obtained just by applying Gabor Filter one time in the proposed model. The worth amount of time is taken to attain a high accuracy i.e., 99.5%. The Gabor filter is a feature extractor and extracted features are provided as input to Convolutional Neural Network.

Future extension of this work will show a hybrid model to measure these findings on video sequence by using Gabor Filter and optical Flow as feature extractor.

ACKNOWLEDGEMENTS

REFERENCES

[1] Neha Jain, Shishir Kumara, Amit Kumara, Pourya Shamsolmoalib, Masoumeh Zareapoor, (2018), "Hybrid Deep Neural Networks for Face Emotion Recognition", Pattern Recognition Letters.

[2] Andrius Dzedzickis, Artūras Kaklauskas, Vytautas Bucinskas, (2020) "Human Emotion Recognition: Review of Sensors and Methods", in MDPI.

[3] Sarvakar, K., Senkamalavalli, R., Raghavendra, S., Kumar, J.S., Manjunath, R. and Jaiswal, S.,(2021), "Facial emotion recognition using convolutional neural networks". Materials Today: Proceedings. Elsevier publication.

[4] T. Wu, M. S. Bartlett, and J. R. Movellan, (2010) "Facial expression recognition using Gabor motion energy filters," IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, pp. 42-47.

[5] Salama, E.S., El-Khoribi, R.A., Shoman, M.E. and Shalaby, M.A.W., (2021), "A 3D-convolutional neural network framework with ensemble learning techniques for multi-modal emotion recognition". Egyptian Informatics Journal, 22(2), pp.167-176.

[6] Bayezid Islam, Firoz Mahmud, Arfat Hossain, (2018), "Facial Expression Region Segmentation Based Approach to Emotion Recognition Using 2D Gabor Filter and Multiclass Support Vector Machine", in ICCIT.

[7] Ramos A.L.A., Dadiz B.G., Santos A.B.G, (2020) "Classifying Emotion based on Facial Expression Analysis using Gabor Filter: A Basis for Adaptive Effective Teaching Strategy", In: Alfred R., Lim Y., Haviluddin H., On C. (eds) Computational Science and Technology. Lecture Notes in Electrical Engineering, vol 603. Springer, Singapore.

[8] Lee, James, Linda Wang, and Alexander Wong. (2020), "EmotionNet Nano: An efficient deep convolutional neural network design for real-time facial expression recognition." Frontiers in Artificial Intelligence 3.

[9] Cheng, S. and Zhou, G., (2020) "Facial expression recognition method based on improved VGG convolutional neural network. International Journal of Pattern Recognition and Artificial Intelligence", 34(07), p.2056003.

[10] Aamir, M., Ali, T., Shaf, A., Irfan, M. and Saleem, M.Q., (2020), "ML-DCNNet: Multi-level Deep Convolutional Neural Network for Facial Expression Recognition and Intensity Estimation", Arabian Journal for Science and Engineering, 45(12), pp.10605-10620.

[11] Shi, M., Xu, L. and Chen, X., (2020), "A novel facial expression intelligent recognition method using improved convolutional neural network". IEEE Access, 8, pp.57606-57614.

[12] Mohammed, S.B. and Abdulazeez, A.M., (2021) "Deep Convolution Neural Network for Facial Expression Recognition", PalArch's Journal of Archaeology of Egypt/Egyptology, 18(4), pp.3578-3586.

[13] Wong, C.C., Alias, E.S. and Kishigami, J., (2013) "Playlist environmental analysis for the serendipity-based data mining", International Conference on Informatics, Electronics and Vision (ICIEV) (pp. 1-6).

[14] Sen, A., Popat, D., Shah, H., Kuwor, P. and Johri, E., (2018) "Music playlist generation using facial expression analysis and task extraction", In Intelligent Communication and Computational Technologies (pp. 129- 139), Springer, Singapore.

[15] https://www.kaggle.com/ankur133047/modified-ck-facial-expression-dataset

AUTHORS

Shubham Luharuka is a 3rd-year student at RV Institute of Technology and Management, Bangalore, India where he is majoring in the Bachelor of Engineering program in the Department of Computer Science and Engineering. He has published one paper "Technical Scrutiny of Blockchain Technology Protocols and its Applications" in the IRO Journal on Sustainable Wireless Systems. He learned about Machine Learning, Artificial Intelligence and Quantum Machine Learning. He completed many Online Certification courses from top organizations like "Introduction to Algorithm and its Analysis" and "Cryptography and Network Security" from National Programme on Technology Enhanced Learning (NPTEL), Machine Learning from Stanford University and more. He got an opportunity to work as a Data Science Intern in Exposys Data Labs and Innomatics Research Lab. He participated in many hackathons and other challenges. After being inspired by the research field he wants to pursue a career in the Research and Development Cell of government agencies. In his free time, he studies research blogs and the problem faced by the world that can be solved by his field of interest.

Meghana R is a student pursuing Bachelor of Engineering in Computer Science and Engineering at RV Institute of Technology and Management, Bangalore, India. She is currently in her 3rd year and has participated in various hackathons and workshops on IOT and Robotics. She has completed internships in the fields of Visual Computing, Machine Learning and Web Development along with a few online certification courses from reputed universities. She is also a member of CSI and is constantly working towards improving her skills.

Pallavi K J is currently studying Computer science engineering at RV Institute of Technology and Management Bangalore, India. She is now in her third year continuing to work toward her degree. She holds internship certificates in Visual Computing and Web development. She has also participated in robotics and IOT Workshops. She is a CSI member and looks forward to make the best use of all the opportunities it provides.

Dr. Asha S Manek is currently working in the Department of Computer Science and Engineering, RV Institute of Technology and Management Bangalore, India. She received her BE degree from Nagpur University in 1993 and ME in Information Technology from University Visvesvaraya College of Engineering, Bangalore University, Bangalore in 2008. She holds a Ph.D. Degree in Computer Science and Engineering from JNTU, Hyderabad, India in 2019. In her 24 years of teaching experience at both UG and PG level, she has published more than 25 research papers in various National and International Conferences and Journals including in Scopus indexed, SCI
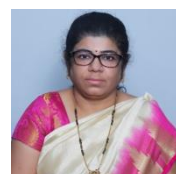
publications, IEEE Access, Springer, WoS and Inderscience. Her research interests include Data Mining, Machine Learning, Web Mining, Information Retrieval, Soft Computing, Text Mining, Social Media Analysis, IoT, Cloud Computing and Wireless Senor Networks. She worked as reviewer for various journals and conferences. She was honoured as Judge, Resource Person, Session chair, TPC member, Advisory committee member at National and International Seminars, Workshops, Conferences, Project Competitions and Toycathon. She is also IEEE, CSI member and lifetime member of ISTE.