

Employee Work Motivation Detection using Image and Audio Processing

H.P.H. Wathsara
Undergraduate
Department of Information
Technology
Sri Lanka Institute of Information
Technology

M.L.P.D. Dhananjana
Undergraduate
Department of Information
Technology
Sri Lanka Institute of Information
Technology

D.A. Nanayakkara
Undergraduate
Department of Information
Technology
Sri Lanka Institute of Information
Technology

B.B.K.R. Dharmarathne
Undergraduate
Department of Information
Technology
Sri Lanka Institute of Information
Technology

L.H Sanjeevi Chandrasiri
Senior Lecturer
Department of Information
Technology
Sri Lanka Institute of Information
Technology

J.K. Joseph
Assistant Lecturer
Department of Information
Technology
Sri Lanka Institute of Information
Technology

ABSTRACT

Customer Support Centers have reached their best level now, being the most preferred way of the modern communication world. These centers play a vital role in the telecommunication industry, helping to satisfy customer services and needs at the highest level. The improvement based on these service companies is based on the agents' performance, which will favorably impact customer satisfaction. In past years, researchers have revealed numerous factors about this background nature of support centers and here from this ultimate task seeks to predict the employee performance as the target. It is suggested to implement a detective system based on image processing and audio processing. The support centers fall into complications in supervising many employees in a specific period to measure the current status of the work and the progress level. During the process, the system uniquely identifies each employee to monitor their availability at a specific time, to assess his/her emotion levels, motions accurately. The proposed system will provide a better service to optimize each employee's efficient workflows, investigate agents' satisfaction level, and confirm whether the customers are treated humanely. The facial features, emotional aspects, motion patterns, and voice frequencies are vital in this research.

General Terms

CNN – Convolutional neural network, API – Application Programming Interface, SVM - Support Vector Machine, HOG – Histogram of oriented gradients, FER - Facial Expression Recognition

Keywords

Face Detection, Emotion Detection, Motion Detection, Voice Detection

1. INTRODUCTION

Image processing is a method to perform certain operations on images to generate enhanced images or retrieve remarkable statistics. This is a kind of signal processing; the input is an image, and the output may be an image or features associated with that image [1]. Image processing is one of the fast-

growing innovations in the present and constitutes the central field of engineering and computing research. Mechanisms for audio processing can be used to verify and monitor audio signals to make crucial decisions based on output matrices. Image processing and audio processing can be applied to detect employee work motivation through the face, emotion, motion, and voice of an employee in a customer support center. The research's main objective is to provide a straightforward approach for accurate decision-making that supervisors can handle within the planned timeline.

The proposed system implements a detective system, a standalone application that will record the employee's daily behaviors as a video and auto uploaded to a drive. The system will generate two reports of motivation levels summary of each employee based on image processing and audio processing separately, and that will be sent to the management through an email daily. The web cameras and headphones with microphones are used to monitor the work motivation of each employee remotely.

Face detection can determine a human face's location and size in a digital image [2]. This can be used to accurately track employees' attendance and working style without depending on their current poses, attribute changes, age variations, and even where they are in a vast gathering. The face detection systems are targeted to decrease the false-positive rate and increase detecting face precision, especially in detailed background images [3]. Three cameras will be used to detect his/her face, and it will capture their availability, emotions, and motions. A new face log in system is used to give access to own profile, which is more secure and much more accessible from in user perspective and the management perspective, to have a summarized detail of each one, including attending time and profile accessing time. Cameras are connected with the system to monitor employees and identify uniquely to take attendance at his arrival when seated in his seat, and meetings and gatherings.

Emotion plays an essential part in expressing one's thoughts and values effectively. Consequently, the identification of emotions has been the subject of many studies recently.

Emotion analysis is essential in a Customer Support Centers environment since low agent handling of interactions could lead to customer frustration and even business loss. A real-time emotion recognition system is being introduced that uses facial landmark analysis of Customer Support Centers agents and clients when conversing. Managers can track employees' emotions through the proposed system, and that would be supportive to get an idea of actions to be taken to motivate employees, knowing his satisfaction levels. Emotions (happy, sad, angry, fear, surprised, neutral, Etc.) will be detected, analyzed, and predicted. The system will identify if a Customer Support Centers agent is new to the working environment and predict how long it will take him to adapt to the environment [7].

Apart from emotions, motions can be tracked while monitoring their performance from an operational perspective. The time breaks they happen to consume and the variety of motion which would affect the progress will be highly analyzed. The camera tracks the motion of seated employees. The manager is notified when a person is engaging in activities other than working in his allocated working hours. That will be supportive of taking actions for employees who waste their time and the time of peer workers during working

hours.

The system also uses audio surveillance technology to record employees' and customers' telephone calls, analyze the data, and summarize them. When considering customer satisfaction, the verbal behavior of employees is highly essential to be monitored. Technology for voice recognition does not just about understand what is being said. It is about translating the information into a functional medium, like a text format or an audio file that can be studied later. This process has many applications that can be used to boost service to customers and the way center does business. The technology will also be used to measure the employee satisfaction level via the frequency and tone of speaking. This can be used as a behavioral analysis to monitor employees whether their service is done happily, with satisfaction and whether they are supportive, polite, and pleasant for customers or not.

2. METHODOLOGY

The proposed system contains four functions that collaborate to be working as a system. How the main functionalities of the proposed system are connected is shown in Fig 1.

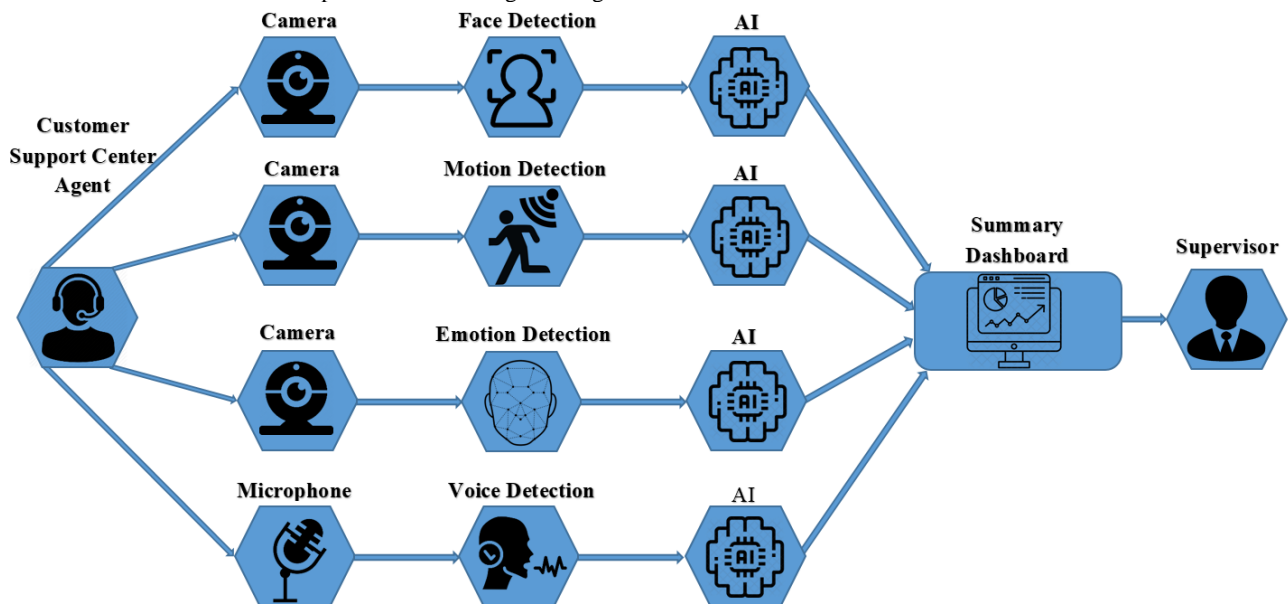


Fig 1: System Overview Diagram

2.1 Face Detection

This function is used for facial identification of people in recorded video. A python algorithm is used with an inbuilt CNN, and that parameter is used for classification. The data set is trained to create a new model, combining pre-trained data sets and extracted unique training data sets [5].

The pictures are rendered black and white as there is no need for color data when identifying faces, and every single pixel in images is tested at one time, and the pixels that are directly surrounded are tested for each pixel. Relative to the pixels immediately surrounding it, how dark the current pixel is relative, and arrows are drawn in which direction the image becomes darker. Looking at this one pixel and the pixels that enter it, the photo becomes darker to the top right [5].

When this is repeated for every pixel in the image, it replaces each pixel with an arrow. The arrows are called gradients and represent the flow from light to dark over the whole picture.

Saving the gradient for every single pixel gives unnecessary detail. The basic flow of lightness/darkness is considered to see the picture's basic pattern at a higher level. The picture will be divided into tiny 16x16 pixel squares. In each square, the gradients are counted in and replaced by the most potent arrows. The result will be to turn the original image into a straightforward representation that captures the face's basic structure. The original image is transformed into a HOG representation, capturing the image's major features irrespective of the image's brightness. The part of our image, which looks the most similar to a known HOG pattern extracted from a bunch of other training faces, should be found in this HOG picture. The face landmark estimation algorithm is used to distort each image such that the eyes and lips are in the sample position all the time in the image, and simple image transformations such as rotation and scale, which maintain parallel lines, are used [5].

No matter how the face turns, the system can focus the eyes,

and the mouth in the picture is approximately the same location. Here, a way of extracting a few fundamental measurements from each face is needed. This process could measure unknown faces and find the known face with the nearest measurements. One hundred twenty-eight measures reliably for each person is produced by the neural network. Nevertheless, measurements can be produced for any face once the network has been trained. Therefore, to get the 128 measurements for each face must run the face images through their pre-trained network. Finally, the Haar cascade classifier is trained to find the person who has the closest measurements of the test image in the known persons' database. That can take from a new test image in the measurements and tells which known individual is the closest match. It takes milliseconds to run the classifier. The output of the classifier is the person's name [5].



Fig 2: Result of locating the 68 face landmarks on our test image

2.2 Emotion Detection

Emotions (happy, sad, angry, fear, surprised, neutral, Etc.) will be detected, analyzed, and predicted by an employee in a support center. It aims to extract client conversations to simplify the tracking and management of emotions. The new proposed system will identify if a support center agent is new to the working environment; it will predict how long it will take him to adapt to the environment. The system comprises four modules: Emotion Trainer, Emotion Detection, Emotion Analysis, and Report Generation. The Emotion Detection module constructs CNN with Keras using TensorFlow backend. The system is comprised of a standalone application [6].

First, a model is created by using a data set. The processes of both training and assessment will be done with data set Fec2013. The compressed version of the dataset takes space of 92 MB while the uncompressed version takes space of 295 MB. The dataset contains 28 K of training and 3 K of testing images. Each picture has been stored as 48 pixels. The pure dataset consists of pixels of an image ($48 \times 48=2304$ values), emotion of image, and form of use (as an instance of train or test) [7].

After training the model, the system captured emotions every 2 minutes when the video is played. Current emotion is identified by using the CNN model and store in DB. Next, the collected emotions will be checked with the emotions on the list. Finally, two values as final positive and final negative will be stored in a CSV file. The system will identify if a Customer Support Centers agent is new to the working environment. It will predict how long it will take him to adapt to the environment. Employees' future bonds will be predicted using Prophet Algorithm.

2.3 Motion Detection

Motion detection is a vital process in which data on moving objects is collected, and stabilization is used in functional areas such as monitoring, classification, recognition, Etc. [9].

Here, this mainly focused on tracking a specific person by video surveillance and stated his/her records for further analysis. This task contains several steps that have been discussed below.

At first, a trained model is used to work our regards to deep learning-based facial recognition, including deep metric learning. The view of the output is a very valued vector of function.

This task is going with detecting movement tracking. Here the intention is to track the movements which are recorded against the shifts.

The images captured via the webcam will be subjected to identify the distracted actions [10-11]. The action will be as follows.

- Talking on the phone
- Drinking
- Talking with the next table

Here VGG16 model is used in order to identify and grab these distracting actions. This model name to be a convolutional neural network that achieves over 92.7% accuracy with the output. Here the task is going out with pre-trained images of the above actions. When the model is loaded, the predefined images will be compared with the current frame, captured by the webcam or through video surveillance. Then it will update by its log in every second, and a record will be generated related to the actions captured.

All these records from the above tasks will be taken as the data entries to perform the monthly report's overall prediction. Once every month, the employee will receive a report with its performance on behalf of the company, for the easiness organization will provide the pictorial graph monthly.

The next task describes the drowsiness measurement of an employee who is being allocated to some such shift plans. Only what is going here is the person will capture from a webcam in a particular time interval regarding action angles, and he/she will be testing out up to several frames. Furthermore, by using facial landmarks, it can be rectified whether he is in drowse or not.

The primary algorithm used here is the eye aspect ratio. After filtering the eye region, facial landmarks are applied to grab the eye's blink nature [8].

The distance between the p1 and p4 is kept constant for an average healthy person, while it will like zero or comparable value when it is identified as blink. The above graph shows clearly how the sudden fall happens when it goes to drowsiness [12].

2.4 Voice Detection

As in most AI programs, first, need data. Here is utilizing a marked literary dataset. Import the essential libraries and afterward read the dataset. Drop the rows with all unused labels to predict text emotion

This current one is a raw dataset with only four segments, the tweet ID, feeling portrayed by the tweet, the creator, and the tweet's content substance. The author column is not required.

Henceforth that can be dropped. Linear SVM algorithm is used for it—a pre-stored dataset used for emotion prediction.

Here, proselyte all these textual data into mathematical data. First, make everything lowercase, eliminate accentuation, and stop words (like relational words). Get all the words to their root structure, i.e., the variations of a word inside the content (for instance, plural structures, past tense, Etc.) must all be changed over to the base word it speaks. The next thought is the possibility that if a word is showing up just a single time in the whole example of information, it doubtlessly has no impact in deciding the content's conclusion. Subsequently, this can eliminate all the infrequently happening words from the dataset, commonly formal people, places, things, and other immaterial words concerning the current setting.

Split the information into preparing and testing parts before performing highlight extraction & gives the general significance of a term in the information and is a proportion of how habitually and once in a while it shows up in the content.

Change the tweet into an exhibit having the check of appearances of each word in it. Straightforwardly use tweets as contributions for some classic AI models Prepared AI models utilizing tally vectors highlights Import all the entire libraries into our scratchpad. Libros and SciPy are the Python libraries utilized for handling sound signs to get text Model. They were using the Keras Sequential Conv1D to covert voice for text to speech conversion. Envision the sound sign in the time arrangement spaceComprehend that the testing pace of the sign. Comprehend the number of recordings for each voice order: Comprehend that the examining pace of the sign & peruse the sound waves and Split into train and approval set.

Import relevant dependencies and Load the Dataset to predict voice-based emotion MLP Classifier used to predict voice emotion out.

Download and convert the dataset to be appropriate for extraction. This cycle is tied in with stacking the dataset in Python, which includes separating sound highlights, for example, acquiring various highlights, such as force, pitch, and vocal lot setup from the discourse signal. The Libros library will be utilized to do that. After plan and burden the dataset, train it on a fit sklearn model. by estimating how great the model is getting along.

Naive Bayes classifier is used for Sentiment analysis. First, it is needed to prepare the data for training the Naïve Bayes Classifier class. Two data sets are prepared as Positive and Negative tweets. At long last, the system utilizes the Naïve Bayes Classifier class to assemble the model. Utilize the train() technique to prepare the model and the accuracy() strategy to test the testing information model.

Then analyze the frequency level will show 'High' and 'Low' as an output. Voice Signal Analysis (VSA) is the way toward examining the voice signal utilizing DSP's various boundaries. Pitch esteems the recurrence of the sound. Perfect pitch alludes to certain people's capacity to perceive a voice note's pitch with no discernable pitch standard.

3. RESULTS

The proposed system is a desktop application for a Customer Support Centers that uses image and audio processing to detect employee work motivation by tracking employees' faces, emotions, motions, and voice. Employees are monitored each day through live cameras, and the recorded video is automatically uploaded to the system drive. Accordingly,

Performance matrices are generated graphically in the system dashboard, and the supervisor can be notified and view each employee's performance to make better future decisions.

3.1 Face Detection

User account management is a crucial part of this system. Every employee maintains a unique user profile. Face unlock log in system is used for a more secure and efficient way to access the user profile. Here faces are captured, and users are automatically directed to the profiles to view details. Providing user credentials is replaced with an efficient and more secure login. Employee user profiles are linked to the main dashboard where supervisors can check each employee's performance through their profiles and the overall performance through the main page.

The system can uniquely and accurately detect individual employees in a single frame, and multiple employees face in a single frame using the input video.

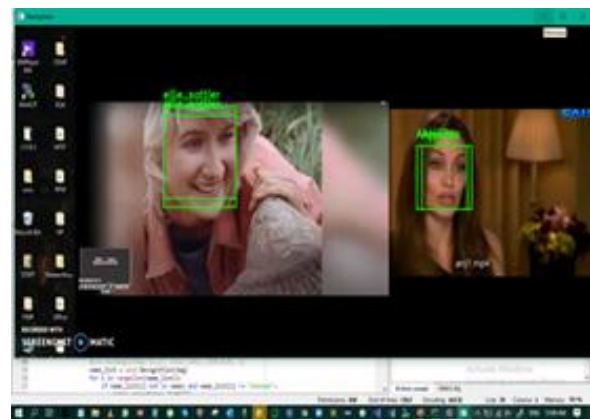


Fig 3: Capturing faces in playing video

Names are displayed when detecting individual faces in single frames. The captured data each day are saved in CSV files, and they are automatically uploaded to the office drive for future usage.

	A	B
1	Date and Time	Names
2	07/11/20_18:34:06	['Anjelina', 'ellie_sattler']
3	07/11/20_19:41:11	['Anjelina', 'ellie_sattler']
4	07/12/20_00:48:15	['Anjelina', 'ellie_sattler']
5	07/12/20_00:48:59	['Anjelina', 'ellie_sattler']
6		

Fig 4: Data storage in CSV file

Performance measurements related to attendance are taken accordingly, and the result is displayed as a graph in the dashboard. Later, this can be used to generate the overall Employee performance report that would support the supervisor's perspective in making better future decisions to motivate the workers and enhance its progress.

3.2 Emotion Detection

The result is a system to identify emotions in a played video. Image processing is used for facial expression recognition. With the help of image processing, useful information from video can get extracted. The system captured emotions every 2 minutes when a video is played. Current emotion is

identified by using the CNN model and store in DB.

After that, the collected emotions will be checked with the emotions in the list, and finally, two values as final positive and final negative will be stored in the CSV file.

Capture the emotions through the video file –

The current emotion is identified by sending the detected face of the particular time through the CNN model. Emotions are captured in each 2 min when the video is played and stored images in another folder.

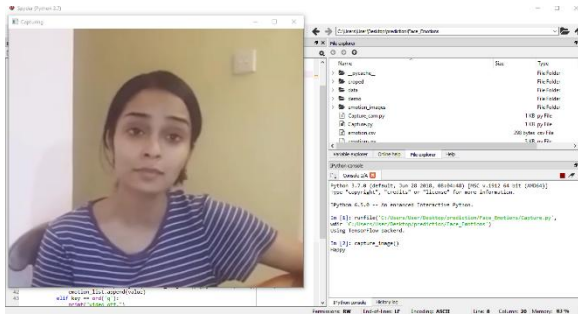


Fig 5: Identify current emotion

Emotion detection – overall prediction for the employee days

Collected emotions from the emotion list will be checked for two values as final positive and final negative, and the same will be stored in the emotion.CSV file.

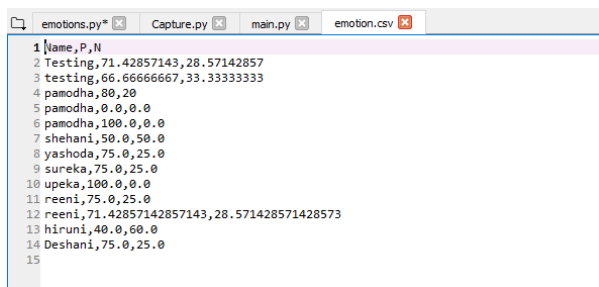


Fig 6: Emotion .csv file

Finally, the system will identify if a Customer Support Centers agent is new to a working environment. It will predict how long it will take him to adapt to the environment. Employee s' future bond will be predicted using Prophet Algorithm.

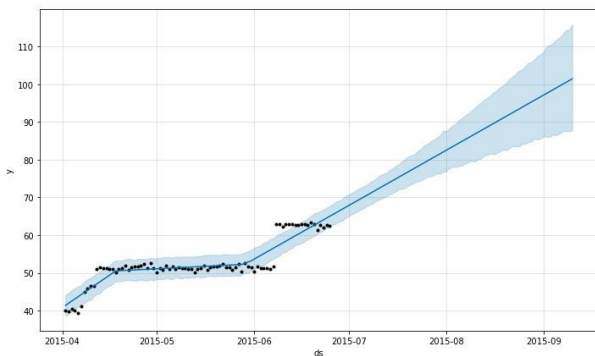


Fig 7: Performance plot graph

3.3 Motion Detection

Here the video will capture the persons as in frames, and the related log will be updated in every millisecond from its

action. It will create 30 frames for one second, and the log will be built up concerning the date and the time. In detecting drowsiness, the log will be built up with its status true or false. The count will be updated along with its log with regards to faces identified in the video. As the function is working out for a single employee, it will count as 1.

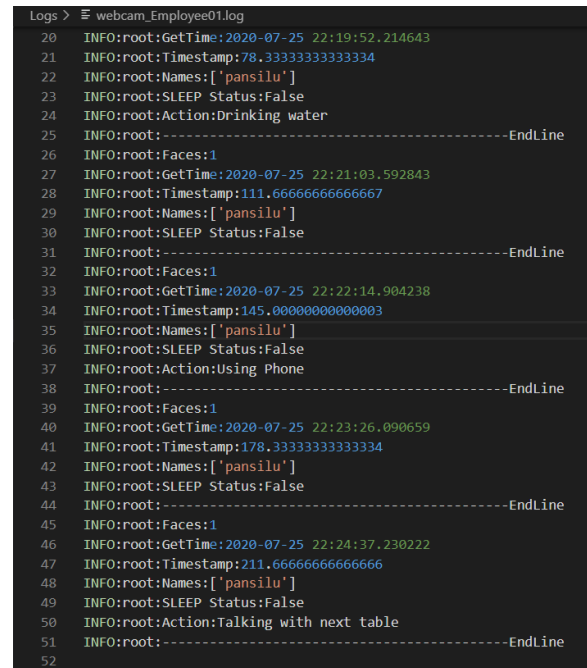


Fig 8: Updated Log

The recorded time and the date will be passed for further prediction, and it will store the data separately, identifying the employee in the working mode and absence mode. Also, it is capable of storing the data which detect the blink nature and typical nature. The breaks consist of the time he spends with other works with aspect to its regular work. It will then produce a record with the employer's information, and it will pass to related authorities in case of further analysis as an email notification.

hi@sarvodayafusion01.onmicrosoft.com
to kavilya123, me

The following are the records for Employee: motion\MotionDrowsiness\Logs\webcam

Breaks	
Start Time	End Time

Sleeps	
Start Time	End Time
08:00:05	08:00:07

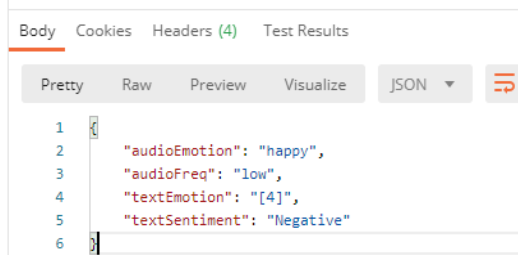
Fig 9: Records send via email

The figure depicts that how long an employee engaged with his work in his allocated shifts.

3.4 Voice Detection

The system can convert a saved audio clip to a text output (per call), detect employee text emotion & voice emotion individually (per call). Sentiment analyzes and plots a frequency graph from the saved audio (per call). After that, compare the graph with emotion to increase the accuracy.

User-wise report generation and send it through an email service.



```
1  
2 "audioEmotion": "happy",  
3 "audioFreq": "low",  
4 "textEmotion": "[4]",  
5 "textSentiment": "Negative"  
6
```

Fig 10: Emotion and Frequency output

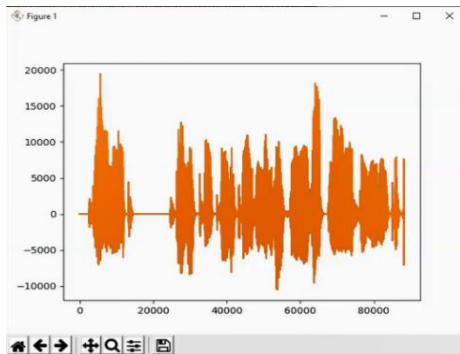


Fig 11: Plotted frequency graph

4. DISCUSSION & CONCLUSION

The main objective of this paper is to introduce a better solution to detect. Employees work motivation in a Customer Support Centers The system capture faces, emotions, motions using image processing techniques, and detects voice using audio processing techniques in the daily recorded video, and relevant reports are generated by analyzing data.

Face detection can be used to track employees' attendance accurately and working style without depending on their current poses, attribute changes, age variations, and even where they are in a vast gathering. The system captures faces in the daily recorded video, and using image processing techniques, reports are generated, and performance is summarized in a dashboard by analyzing data.

The face recognition algorithm that has been used here as the main algorithm gives accurate results using a lesser amount of trained data, unlike LBPH algorithms that came to usage in recent years. Face recognition python algorithm is used with an inbuilt CNN, and that parameter is used for classification.

HOG algorithm is used to encode a picture to generate a shortened version of the picture, and the part of the image that almost looks like a generic HOG encoding of a face is found using the simplified image. Then the prominent landmarks are used to find the poses. The image is wrapped so that the eyes and mouth are centered. A neural network is used to measure the features of the central image saving 128 measurements. The faces that are being measured in the past are compared with new measurements to find the closest match and display the name as a result. Attendance management is also done in the same procedure. After processing the input face images using advanced image processing methods such as Face recognition, face landmark estimation to have better image features and the same advanced image processing procedures will be applied to train face images to ensure high quality. The system can uniquely and accurately detect individual employees in a single frame and multiple employees' faces in

a single frame using the input video. Names are seen when individual faces are identified in single frames. The data collected each day is stored in CSV files and immediately uploaded to the office drive for future use. Performance measures on attendance are taken accordingly, and the result is shown in the dashboard as a graph.

Emotion detection mainly focuses on the system will identify if a Customer Support Centers agent is new to the working environment and will predict how long it will take him to adapt to the environment. Employee s' future bond will be predicted using Prophet Algorithm. The system captures faces in the daily recorded video, and using image processing techniques, reports are generated, and performance is summarized in a dashboard by analyzing data. After preprocessing the input face images using advanced image processing techniques such as Face recognition, face landmark estimation, better image features, and the same advanced image processing techniques will be applied to train emotion images to ensure high quality. After training the model, the system will capture emotions every 2 minutes when the video is played. Current emotion is identified by using the CNN model and store in DB. The emotion recognition part generates a report of the facial expression classifiers' performances, and collected emotions from the emotion list will be checked for two values as final positive and final negative, and the same will be stored in the emotion.CSV file.

In motion detection, the primary intention is to introduce a system where employees can identify their downfalls overcome with several effects. In a motion tracking system, it can track the employee face at the 1st stage, and through a sequence of images taken from the webcam, will be subjected for further analysis to calculate the breaks he has taken. The time breaks will be identified by frames that capture the employee in the seat and the frames where they are not in the seat. It will record the time where the employee is out of the seat. This is going to be the 1st phase that the system cover-up. In the next phase, it will detect the drowsiness of an employee during working hours. This detection is done using predefined formulae introduced by extracting the eye region of the employee. As identified the facial landmarks from the first phase here, it is extracting only the eye region for the detection and analysis. The final phase system can detect an employee's action, and here it should be said only declaring an employee's hostile actions. The prediction is made if any employee is a track out with these negative actions, it will be identified as a factor to downfall the performance. The action here is to check out drinking water and answer mobile phones for a long time, perhaps in working time. By considering the complexity of handling more than two actions, it was just limited up to track two actions to take the performance level. The complexity means that the system will not identify those actions if the predefined angle might be change up for any reason. Also, it takes too much effort to train the data set to have accurate output. By taking all these output data, it is graphically viewing an employee's performance level where he and the organization can get a clear view of his working nature.

Voice detection is used for facial identification of people in recorded audio. Voice recognition python algorithm is used with an inbuilt CNN, and that parameter is used for classification. The data set is prepared to make another model, joining pre-prepared informational collections, and removed uncommon highlights of preparing informational collections. This part's fundamental goal is to acquaint a superior

arrangement with distinguishing Employees' work inspiration in a Customer Support Centers. In the capacity 'voice detection' system catches employees' emotions in the call by call recorded sound and utilizing voice acknowledgment strategies, reports are produced, and execution is summed up in a dashboard by dissecting data. Speech Emotion Recognition (SER) is an examination zone issue that attempts to surmise the speech signals' emotion. Proposed work plans to build up a framework that needs to convert verbally expressed word into text utilizing SVM demonstrating strategy utilizing acoustic component to be specific MFCC. In this work, the worldly encompass through RMS vitality of the sign is determined for isolating individual words out of the persistent addresses utilizing voice movement identification technique. Highlights for each disengaged word are separated, and those models were prepared. During preparing measure, each included word is isolated into 20ms covering windows for extricating 13 MFCCs highlights. SVM displaying strategy is used to show every individual expression. Subsequently, each disconnected word fragment from the test sentence is coordinated against these models for finding the semantic portrayal of the test input discourse. Here, a report on voice behavior is generated [4].

The generated four reports will update the employee profiles and the system dashboard. The final output is the overall employee performance summary, displayed in graphs that support the supervisor's perspective in making better future decisions to motivate the workers to enhance the company's progress.

The limitations of face detections concluded as the inability to use more cameras to detect every space for satisfying coverage and accurate output owing to the standard integration issues. Also, the inability to capture employees unless they are around his machine since a webcam's detection area is limited. Furthermore, the inability to detect one employee's machine's working history through his profile if another one is working in his place using that machine as the system will make the new person log in to his profile with his new face detection.

The implemented system can be improved to use more cameras to detect every space in the working area for satisfying coverage and accurate output. Apart from that, this can be improved to address occlusion as it cannot guarantee that the implemented system avoids this issue every time. The system can identify partially covered faces to some extent, but fully covered faces, faces with masks cannot be identified on the other hand.

Additionally, this can be improved to use more cameras to detect every space in the working area for a satisfying coverage and accurate output, use advanced methods to track sensitive motions, and train any other language other than English for voice detection.

5. ACKNOWLEDGMENT

It is a great honor to extend sincere gratitude to our supervisor for her dedicated support and aspiring guidance, which motivated us to move further. Her support has been invaluable during this research, and comprehensive insights into this area have made this an exciting experience for us. The meetings and conversations were crucial in encouraging us to think outside the box, to shape a detailed and critical critique from multiple perspectives. Finally, our warmest thanks and appreciations are extended to the co-supervisor for the encouragement given.

6. REFERENCES

- [1] Incipientinfoteam, "Introduction to image processing," IncipientInfo, 04-Mar-2019. [Online]. Available: <https://incipientinfo.com/what-is-image-processing-digital-image-processing/>. [Accessed: 10-Feb-2020]
- [2] Kumar, A. Kaur, A. and Kumar, M. 2018, "Face Detection Techniques: A Review. Artificial Intelligence".
- [3] Liu, R. Zhang, M. and Ma, S. 2010. "Design of face detection and tracking system", 3rd International Congress on Image and Signal Processing, China.
- [4] Analytics Insight, "Speech Emotion Recognition (SER) through Machine Learning," 27-Jul-2020. [Online]. Available: <https://www.analyticsinsight.net/speech-emotion-recognition-ser-through-machine-learning/>. [Accessed: 12-Feb-2020].
- [5] Geitgey A. "Machine Learning is Fun! Part 4: Modern Face Recognition with Deep Learning," Medium, 07-Nov-2018.[Online].Available: <https://medium.com/@ageitgey/machine-learning-is-fun-part-4-modern-face-recognition-with-deep-learning-c3cfc121d78>. [Accessed: 16-Jul-2020].
- [6] Serengil, S. I. 2020, "Facial Expression Recognition with Keras".[Online].Available: <http://sefiks.com/2018/01/01/facial-expression-recognition-with-keras/>. [Accessed: Apr 10, 2020].
- [7] Nguyen, B.T. Trinh, M.H. Phan, T.V. and Nguyen, H.D. "An efficient real-time emotion detection using camera and facial landmarks", IEEE Conference Publication. [Online]. Available: <https://ieeexplore.ieee.org/document/7926765> [Accessed Feb 10, 2020]
- [8] Maior, C.S. Moura, M. Santana, J.M and Nascimento L. "Real-time SVM Classification for Drowsiness Detection Using Eye Aspect Ratio". Conference: Proceedings of Probabilistic Safety Assessment and Management PSAM 14. Los Angeles. Retrieved from,04-Sep-2018. [Online]. Available: https://www.researchgate.net/publication/336106970_Realtime_SVM_Classification_for_Drowsiness_Detection_Using_Eye_Aspect_Ratio. [Accessed: 20-Jul-2020].
- [9] Kak, S.F. Mustafa, F.M and Valente P. "A Review of Person Recognition Based on Face Model." Retrieved from, 01-Jan-2018.[Online].Available: https://www.researchgate.net/publication/323611486_A_Review_of_Person_Recognition_Based_on_Face_Model . [Accessed: 21-Mar-2020].
- [10] Alotaibi, M and Alotaibi, B. 2019, "Distraacted driver classification using deep learning." [Online]. Available: https://www.researchgate.net/publication/336685346_Distraacted_Driver_Classification_Using_Deep_Learning
- [11] Kshirsagar, S.S. Ghonge, P.H. "Movement Detection Using Image Processing. International Journal of Science and Research (IJSR)," 09-Jun-2013. [Online]. Available: https://www.researchgate.net/publication/323611486_A_Review_of_Person_Recognition_Based_on_Face_Model . [Accessed: 15-Feb-2020].
- [12] Rosebrock, A. 2017, "Eye blink detection with OpenCV, Python, and dlib." [Online]. Available: <https://www.pyimagesearch.com/2017/04/24/eye-blink-detection-opencv-python-dlib> [Accessed: 17-Jul-2020].