ISSN: 0975-833X

*Available online at http://www.journalcra.com*

*International Journal of Current Research*
**Vol. 17, Issue, 01, pp.31270-31276, January, 2025**
**DOI: https://doi.org/10.24941/ijcr.48375.01.2025**

**INTERNATIONAL JOURNAL OF CURRENT RESEARCH**

# RESEARCH ARTICLE

# OPTIMIZED VITALITY DETECTION IN MULTI-FACE IMAGES WITH ADVANCED CONVOLUTIONAL NEURAL NETWORKS

**[1*]Logeswari Saranya, R. and [2]Umamaheswari, K.**

[1]Research Scholar, Department of Information Technology, PSG College of Technology, Coimbatore, India
[2]Professor, Department of Information Technology, PSG College of Technology, Coimbatore, India

**ARTICLE INFO**

*Article History:*
Received 20th October, 2024
Received in revised form
17th November, 2024
Accepted 24th December, 2024
Published online 30th January, 2025

*Key Words:*

Pre-Processing; Face Detection; Encoding; Face Recognition; Vitality Detection; Convolution Neural Network.

***Corresponding author:***
***Logeswari Saranya, R***

**ABSTRACT**

Authentication is always required for the majority of systems in the rapidly evolving globe. Face recognition is a method for identifying or verifying someone based on an image captured by a camera or a single frame from a video. Such complex tasks are beyond the capacity of a computer to do alone. Advanced ideas like deep learning can be applied to the detection and recognition of faces. Face recognition is used in many different contexts, such as user identification, device unlocking, and more. They can also be crucial in identifying multiple locations where multiple people may enter or be present at the same time like student presence in seminar hall and entry cameras, etc. Relying on still photos from printed or digital images for verification poses security risks to users. Using the offered sample photographs as a lead, the multiple face identification and vitality detection methods locate various faces in the image, recognize them, and validate that the person in the frame is alive. This research work uses Encoding Convolutional Neural Network for the face detection, recognition and to verify vitality presence in the biometric system. This model can used in areas such as Indian Senior Pension Scheme for face detection which is essential now a days to monitor whether the person is really present in front of the camera, Human Tracking Systems, National Security Systemsetc.



**Citation: Logeswari Saranya, R. and Umamaheswari. K. 2025.** "Optimized Vitality Detection in Multi-Face Images with Advanced Convolutional Neural Networks". *International Journal of Current Research*, 17, (01), 31270-31276.

# INTRODUCTION

In the branch of research known as "facial image processing," information about human faces is extracted and analyzed with the goal of better understanding how social interactions including identification, emotion, and purpose are influenced by facial features. There has been tremendous growth in research over the past 10 years in the fields of face detection and tracking, facial feature detection, face identification, facial expression and emotion recognition, face coding, and virtual face synthesis (1). A variety of applications have been created recently in areas such as image retrieval, surveillance and biometrics, visual speech interpretation, virtual characters for e-learning, online marketing or entertainment, intelligent human-computer interaction, and others. This is made possible by the introduction of new, potent machine-learning techniques, statistical classification methods, and complex deformable models. However, more effort needs to be made to create more dependable systems, especially when it comes to adjusting for position and illumination differences in natural environments. Newer approaches might consider a range of inputs even if the majority of systems are made to process static images. Security, human-computer interaction, and video annotation are just a few examples of the vision-based human-oriented applications that are increasingly needed as video becomes more appealing and popular. The attractiveness of capturing 3-dimensional data is growing, and its processing can result in better systems that are more resilient to lighting effects and make it simpler to obtain discriminatory information (2). The paper is organized as follows: Section 2 details about the Literature survey and the gaps identified. The algorithm for proposed work, features of data set and algorithm for encoding are mentioned in Section 3.Result of image data set is demonstrated in Section 4.Performence analysis based on training loss and accuracy is discussed in Section 5.Conclusion is given in Section 6.

## RELATED WORK

**Face Detection:** Any facial analysis system that relies on the ability to recognize human face characteristics in photos uses face detection algorithms. Due to the growing need for face recognition systems in the past, face detection technologies are used in a variety of real-world scenarios.

Cascade Classifier, Dlib HOG (3), DlibCNN (4) and MTCNN (5) are among the approaches that have been proposed to solve this problem.

**Face Recognition:** The face recognition technique for in-depth learning is highly accurate despite the complexity of the model and the slow recognition speed. Despite the face recognition method's great accuracy for in-depth learning, the model is complex, and the recognition speed is slow (6) (7). The researchers give tried-and-true methods and technologies for every stage of the creation of a recognition system because the recognition field has produced such a wide variety of distinctive solutions. The support vector machine (SVM) method, which can greatly speed up the recognition process, is also provided by academics as a way of recognition. Techniques for estimating face landmarks have been described by researchers. By lining up the face for easier recognition, it helps you raise the system's quality (8).

**Vitality Detection:** According to these features, the face is the most significant; hence it is also utilized for recognition systems and is widely used in areas that require security. Vitality detection in the identification system is required to protect against attacks. Vitality detection, also known as spoof detection, improves the robustness of the recognition system by determining whether the input image depicts a real human or not (9). For presentation attacks, authors have created a variety of IQA-based approaches. Static and dynamic are two more texture-based methods. Face Net with an improved loss function is used for face recognition once face detection and alignment are completed using Multi-task Cascaded Convolutional Neural Networks (MTCNN) (10).

**PROPOSED METHODOLOGY FOR VITALITY DETECTION:** To effectively leverage a Convolutional Neural Network (CNN) for multifaceted tasks such as face recognition and vitality detection, it is essential to ensure precise face identification. Proper camera positioning is critical to guarantee that the subject's face is clearly visible. Upon successful face detection, the facial recognition module processes each image on a frame-by-frame basis, enabling the identification of individual faces. Subsequently, the vitality detection process is initiated. The CNN-based methodology involves three primary phases: Face Image Detection, Encoded Face Image Recognition, and Vitality Detection. The model must be adept at recognizing faces from the provided dataset and differentiating between multiple faces within a single image frame. The final phase involves evaluating the vitality of each detected face to ascertain if it is alive.For real-time video capture, a camera or mobile device is utilized, with the video being stored on a laptop after processing on a frame-by-frame basis. The captured image is prepared for face detection, after which the facial recognition model identifies the individual. Finally, the vitality detection model determines the aliveness of the detected face.

**Proposed Vitality Detection Work Flow:** The primary goal of the proposed research is to develop a facial recognition algorithm based on Convolutional Neural Networks (CNN). This work employs the Multi-task Cascaded Convolutional Neural Networks (MTCNN) technique, which integrates face recognition and vitality detection models. Each video frame is sequentially processed to identify faces on a laptop. The system should enable real-time capture and upload of images from a camera or mobile device. The final output must be processed, packaged, and stored efficiently.The sequence work flow of proposed work is given in Fig.1.


**Fig.1. Proposed Design Work Flow for Vitality Detection**

**Work Flow of Dataset Preprocessing**

**Frame by Frame Video Processing:** The real-time video is captured using a mobile camera equipped with the IPWebcam application. The IPWebcam hosts the live video stream on the local network, to which the laptop is also connected, enabling the laptop to send and receive sequential frame-by-frame images. Consequently, the laptop processes the received images. The captured images are processed by the model in several stages after being transferred to the laptop. The Uniform Resource Locator library enables the processing of images from IPWebcam, which are then sent to the face identification model for further analysis. In the dataset pre-processing phase for vitality detection, extracted frames from the video are utilized to crop facial images. Specifically, the video is segmented into three distinct image frames during pre-processing: the beginning (at the 1st second), the middle (at the 5th second), and the end (at the 10th second) of a 10-second video. Subsequently, facial regions within these frames are cropped using the Multi-task Cascaded Convolutional Neural Network (MTCNN) technique as shown in Fig.2 and Fig.5. The faces are then trained.

**Detection of Facial Images**

**Algorithm1** Face Detection Process

The process begins with the input of an RGB image. The BGR color space is subsequently extracted from the RGB color space. The extracted facial image is then processed through a Multi-task Cascaded Convolutional Neural Network (MTCNN) model, which operates as follows:

- **Image Pyramid Generation:** Multiple resized versions of the input image are created to facilitate face detection at various scales.
- **P-Net (Proposal Network):** A sliding window approach is applied to each image in the pyramid, generating candidate face regions. Non-Maximum Suppression (NMS) is employed to eliminate overlapping candidates.
- **R-Net (Refinement Network):** This network refines the candidate bounding boxes generated by the P-Net. It crops and resizes the regions for further processing, applies the R-Net, and performs NMS again to filter the results.
- **O-Net (Output Network):** At this final stage, the O-Net further refines the bounding boxes, classifies the regions as faces, and detects facial landmarks. A final NMS is applied to obtain the definitive detections.
- **Main Function:** This function integrates the three stages and returns the final detected bounding boxes and facial landmarks.

**Output:** The detected face coordinates are then cropped from the image, resulting in the extraction of distinct facial regions.

**Processing of Key points in Detected Faces:** Key facial landmarks, including the left eye, right eye, nose, left corner of the mouth, and right corner of the mouth, are detected as illustrated in Fig. 3.

The detection process returns a list of dictionaries containing the coordinates of the identified facial features. Each dictionary entry corresponds to one of the aforementioned landmarks. In addition to the coordinates, the output includes a confidence score indicating the accuracy of the face detection. The detected face is enclosed within a bounding box, where the coordinates (x,y) represent the top-left corner and (x+width, y+height) represent the bottom-right corner of the face.

**Recognition in Facial Images**

**Proposed Encoding Process for Face Recognition**

**Algorithm2** Facial Feature Extraction and Embedding

**Input Image Processing**

- Input: A set of facial images.
- Pre-process the images to standardize size and format for consistent input to the Convolutional Neural Network (CNN).

**Feature Extraction**

- Apply a Convolutional Neural Network (CNN) to each pre-processed facial image.
- Extract high-dimensional facial features that represent the unique characteristics of each face.

**Embedding Generation**

- For each extracted set of facial features, generate a 128-dimensional (128-d) embedding.
- This embedding serves as a compact and discriminative representation of the facial features.

**Label Mapping and Storage**

- Map each 128-d embedding to its corresponding labelled name.
- Store the mapped embedding's and labels in an encoding pickle file for efficient retrieval and future use.

**Data Encoding and Storage**

- For each facial image, save the encoded 128-d embedding along with the corresponding label (i.e., the person's name).
- Ensure that each entry is correctly labelled and stored for subsequent identification and verification tasks.

**Output:** A structured dataset containing 128-d embedding's mapped to labelled names, stored in a pickle file for later retrieval and use in facial recognition tasks.

**Working of Face Recognition**

**Algorithm 3** Face Recognition Process

**Face Image Loading and Detection**

- Load the face image from the backend system.
- Detect the face within the image using a face detection algorithm.

**Face Comparison**

- Compare the detected face against pre-stored encoded data (e.g., 128-dimensional embedding's).
- Calculate the similarity between the detected face's encoding and the stored encodings.

**Match Counting**

- For each successful match between the detected face and the stored encodings, increment the corresponding match count.

**Label Determination**

- Identify the label with the highest match count.
- This label is considered the most likely identification of the detected face.

**Final Prediction**

- If the highest match count exceeds a certain threshold, assign the corresponding label as the final prediction.
- If no label meets the threshold, classify the face as "unknown."

**Output:** The algorithm outputs the most likely label associated with the detected face or identifies the face as "unknown" if no significant match is found.

**Vitality Prediction of Facial Images**

**Algorithm 4** Vitality Detection Model Training Protocol

**Initialization**

- o Set the number of training epochs and define the data split ratio for training and testing.
- o Initialize the model specifically designed for vitality detection.

**Loading and Preprocessing**

- o For each epoch, iterate over the input images.
- o Load each image and resize the facial region to 32x32 pixels to standardize input dimensions.

**Region of Interest (ROI) Extraction**

- o Extract the ROI from the resized image, focusing on the part of the image that is most relevant for vitality detection.

**Encoding**

- 
  - o Encode the extracted ROI into a format suitable for model training. This encoding translates the

visual features into a data structure that the model can learn from.

### Model Training and Testing

- Train the model using the encoded data and the predefined train-test ratio. The model is updated based on training results during each epoch.
- Test the model during each epoch to evaluate its performance and adjust as needed.

### Model Storage

- Once training is complete, save the final version of the trained model for future use in vitality detection tasks.

**Output:** This algorithm systematically addresses the steps required to train a vitality detection model, ensuring that the data is processed, encoded, and utilized effectively to create a reliable and accurate model.

## FEATURES OF DATASET AND RESULTS OF PROPOSED WORK

### Features of Kaggle Dataset considered for Face Detection and Recognition

Face Recognition is performed using the Kaggle dataset "100-Bollywood-celebrity-faces.

**TABLE 1. 100 Bollywood star faces in the dataset**

| | |
|---|---|
| **Images Considered** | 12,400 |
| **Persons Estimated** | 100 |
| **Images/person Predicted** | 100-130 |
| **Size** | 2.0 GB |

### Sample Preprocessed Image from Dataset

(a)

(b)

**Fig.2. (a) and (b) Cropped Face from the Original Image**

### Key Point Detected in Facial Image

**Fig.3. Result of Sample Real Live Image for Key Points Detection**

### Features of ROSE Youtu Dataset considered for Vitality Detection

**Table 2. ROSE Youtu Dataset for Vitality Detection**

| | |
|---|---|
| **Total Videos** | 3350 |
| **Subjects Estimated** | 25 |
| **Total Size** | 5.45 GB |
| **Total Video Clips** | 150-200/subject |
| **Each Video Duration** | 10 secs |
| **Standoff Distance** | 30-50 cm |

The ROSE Youtu Dataset is categorized into four classes of images:

- Paper Print
- Digital Photo
- Actual
- Mask

For the purpose of classification, the dataset is divided into two categories: real and fake. The "real" category consists of images from the "actual" class, while the "fake" category includes images from the "paper print" and "digital photo" classes. Masks are treated as paper prints and are therefore excluded from the analysis.In total, 3,350 videos were initially considered, which were then refined to 1,781 videos classified as either real or fake. The "real" category contains 2,688 actual images, while the "fake" category comprises 1,315 paper prints and 1,430 digital photos.A sample image from the dataset is shown in Fig.4.

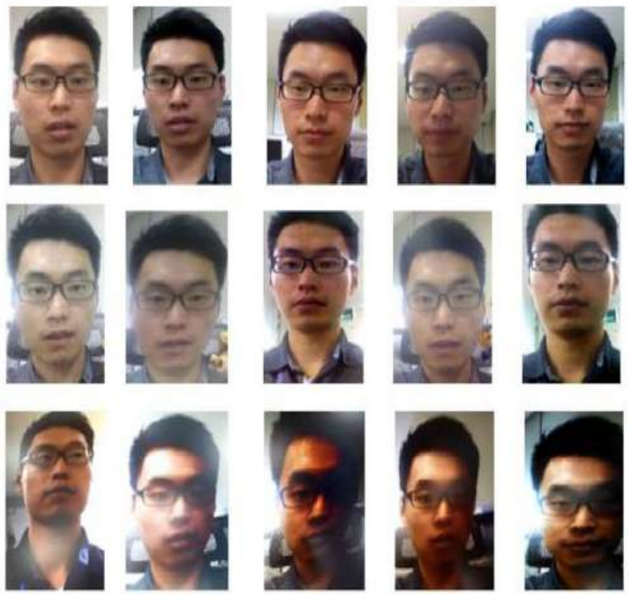

**Fig.4. Sample Images in ROSE Youtu Dataset of a Single Person**

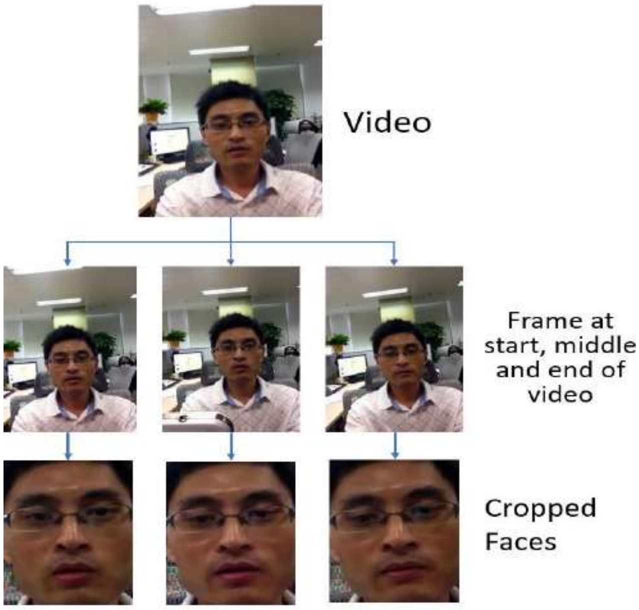


**Fig. 5. Video Preprocessing Phases of ROSE Youtu Dataset**

**Face Detection of Static and Dynamic Images:** Results of face detection in static is given in Fig. 6 and dynamic images is given in Fig. 7.

**Fig. 6. Face Detection in Static Image**

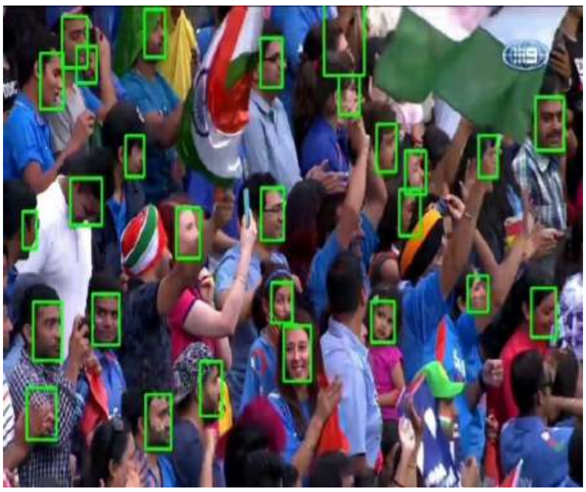

**Fig.7.Face Detection Result in Dynamic (Live Footage) Image-Single Frame**

**Face Recognition:** Results of face recognition is given below in Fig. 8 and Fig. 9.

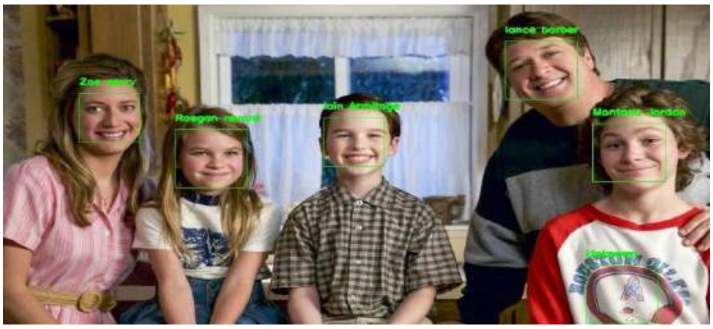

**Fig. 8. Results of Face Recognition – Young Sheldon**

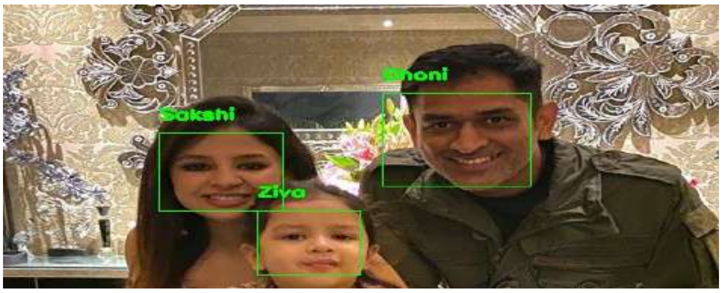


**Fig. 9. Results of Face Recognition -Dhoni Family**

**Vitality Detection:** Comparison of vitality detection between live person, pass port size photo graphic image and mobile image is given in Fig. 10 and Fig. 11.


**Fig. 10. Comparison of Vitality Detection of Photo Graphic Image and Mobile Image with Live Person**

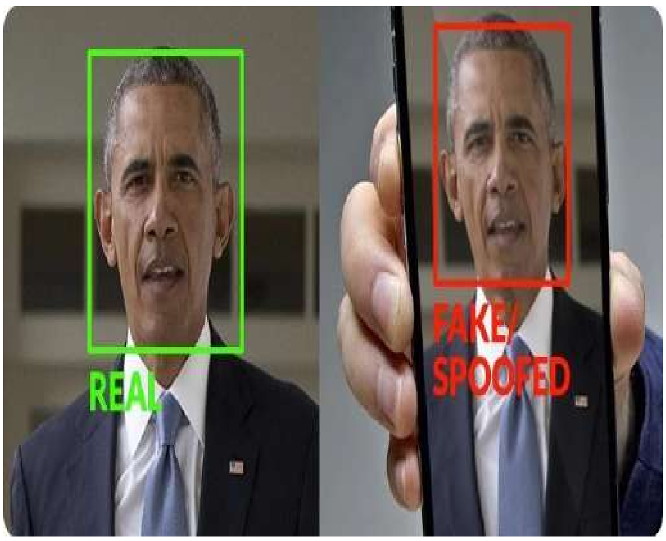


**Fig.11. Comparison of Vitality Detection between Real and Fake**

# DISCUSSION

### Analysis on Face Detection

**Performance of Face Detection based on Process Time:** Performance of model from Intel i7-3612QM CPU @ 2.10GHz, with a CPU-based Tensorflow1.4.1.Based on image size the image quality is determined (11) (12).

**TABLE 3.Performance of Face Detection using Single Frontal Face**

| Image size | Total pixels | Process time | Frame Per Second |
|---|---|---|---|
| 460x259 | 119,140 | 0.118sec | 8.5 |
| 561x561 | 314,721 | 0.227sec | 4.5 |
| 667x1000 | 667,000 | 0.456sec | 2.2 |
| 1920x1200 | 2,304,000 | 1.093sec | 0.9 |
| 4799x3599 | 17,271,601 | 8.798sec | 0.1 |

**Table 4.Performance of Face Detection using Multi-Frontal Faces**

| Image size | Total pixels | Process time | Frame Per Second |
|---|---|---|---|
| 474x224 | 106,176 | 0.185sec | 5.4 |
| 736x348 | 256,128 | 0.290sec | 3.4 |
| 2100x994 | 2,087,400 | 1.286sec | 0.7 |

**Accuracy obtained for Face Detection:** Accuracy based on performance against benchmark Dataset "UTK Dataset"

**Table 5. Accuracy for Face Detection using UTK Dataset**

| | Multi-task Convolutional Neural Network | Haar Cascade |
|---|---|---|
| Detected Faces in UTK Dataset (24,111faces) | 21666 | 19,915 |
| Accuracy | 95% | 89% |

### Analysis on Face Recognition

**Accuracy obtained and Comparison of models for Face Recognition:** Results based on various face recognition algorithms' performance against the "Labeled Faces in the Wild Home" Dataset.

**Table 6. Accuracy of Face Recognition using LFW Dataset**

| Algorithms | Accuracy |
|---|---|
| **LBP** | 85% |
| **Eigen Faces** | 88% |
| **Fisher Vector Faces** | 93% |
| **Proposed Encoded Convolutional Neural Network** | 99% |

### Analysis of Vitality Detection based on Loss and Accuracy

**Training Loss and Accuracy Graph for Vitality Detection with different Train and Test split up**

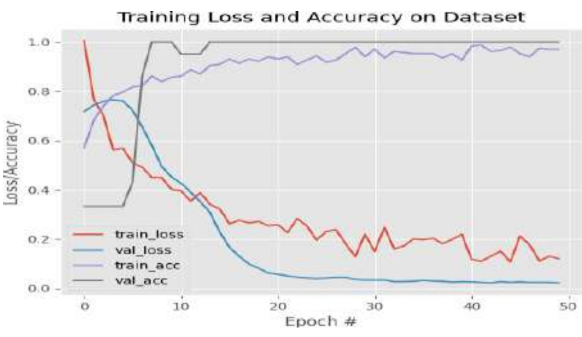


**Fig. 12. Accuracy and Training Loss of Vitality Detection when split up is 70% & 30%**

The training loss comparison between split up and accuracy is predicated based on image feature analysis and image quality measure (13) (14).

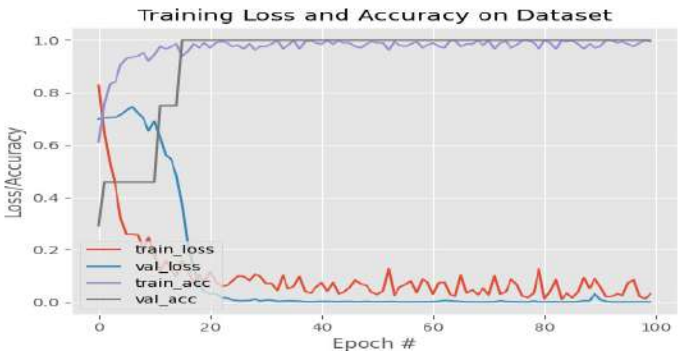


**Fig. 13. Accuracy and Training Loss of Vitality Detection when split up is 80% &20%**

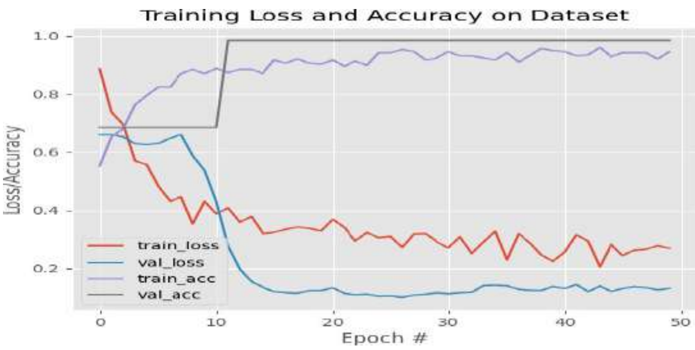


**Fig. 14. Accuracy and Training Loss of Vitality Detection when split up is 90% &10%**

**Accuracy and Loss Estimation:** The model is split into different training and testing data, including70-30, 80-20, and 90-10.The 80/20 split was also proven to be efficient. There were 100 Epochs in each divide (15).

**Table 7. Accuracy estimated for different Dataset split-up**

| **Train-Test Split up** | **70-30 Split Up** | **80-20 Split up** | **90-10 Split Up** |
|---|---|---|---|
| Training Accuracy | 94% | **99%** | 90% |
| Loss | 10% | **1%** | 25% |
| Validation Accuracy | 98% | **99%** | 97% |
| Loss | 5% | **3%** | 10% |

# CONCLUSION

The proposed model was developed and validated using pre-recorded video data, demonstrating its effectiveness. The integration of the camera, face detection and recognition components, along with the vitality detection module, was successfully achieved. The model underwent rigorous testing and evaluation using both existing and recorded video footage. To enhance the system's processing speed, employing a computer with higher computational power is recommended. This model is instrumental in Automated Attendance Management Systems, particularly in seminar halls, where the vitality detection feature is crucial for preventing proxy attendance. Additionally, it holds significant value in applications such as Senior Pension Schemes, where real-time presence verification is required.

Thus, the potential applications of the Face Recognition System extend beyond these use cases.

**FUTURESCOPE**

Face recognition technology is a critical component in advanced human authentication systems. Given that many organizations and offices utilize CCTV footage for surveillance, integrating face recognition with live video feeds significantly enhances the effectiveness of automated monitoring systems. When applied to traffic and sidewalk cameras, this technology can assist law enforcement in more efficiently identifying criminals and locating missing persons, providing real-time alerts on their presence. Additionally, face recognition can be employed to develop automated feedback systems based on facial expressions during guest lectures. Moreover, it can be utilized to effortlessly identify the cast in any film, highlighting its broad range of applications.

# REFERENCES

1. Garg, S., Mittal, S. and Kumar, P, “Performance Analysis of Face Recognition Techniques for Feature Extraction”, Journal of Computational and Theoretical Nanoscience, 16(9), p.3830-3834, 2019.https://doi.org/10.1166/jctn.2019.8257.
2. Chunming Wu, Ying Zhang, “MTCNN and FACENET-Based Access Control System for Face Detection and Recognition”, In Automatic Control and Computer Sciences 55, pp.102-112, 2021.https://doi.org/10.3103/S0146411621010090.
3. Amal Adouani, WiemMimoun Ben Henia, ZiedLachiri, “Comparison of Haar-like, HOG, and LBP approaches for face detection in video sequences”, 16th International Multi-Conference on Systems, Signals & Devices (SSD'19), 2019.DOI: 10.1109/SSD.2019.8893214.
4. Rosebrock, A. Facial landmarks with dlib, OpenCV, and Python (Electronic resource) Accessmode: https://www.pyimagesearch.com/2017/04/03/facial-landmarksdlib-opencv-python/.
5. Al-Imran, Baniamin Shams, Faysal Islam Nasim “Analysis of Face Recognition based on five different viewpoints of face I mages using MTCNN and Face Net“, 2019. https://www.semanticscholar.org/ paper/ Analysis-on-face-recognition-based-on-five-of-face-Hossain Golam/ af9554de5f4ad3eb 5d1b272e343 eadd74b 20fcd2.
6. F. Schroff, D. Kalenichenko, and J. Philbin, “FaceNet: A Unified Embedding for Face Recognition and Clustering”https://www.cvfoundation.org/openaccess/content_cvpr_2015/app/1A_089.pdf.DOI: 10.1109/CVPR.2015.7298682.
7. Sharma S, Karthikeyan Shanmugasundaram, Sathees Kumar Ramasamy “FAREC - CNN Based Efficient Face Recognition Technique using Dlib, International Conference on Advanced Communication Control and Computing Technologies (ACCT), 2016.DOI: 10.1109/ICACCCT.2016.7831628.
8. Support Vector Machines (Electronic resource) – Access mode: http://scikit-learn.org/stable/modules/ svm.html.
9. Rybchak, Z. and O. Basystiuk, "Analysis of computer vision and image analysis techniques" ECONTECHMOD: an international quarterly journal on the economics of technology and modeling processes, Lublin: Polish Academy of Sciences, vol. 6, no. 2, pp. 79-84, 2017.
10. Chan, P.P., Liu, W., Chen, D., Yeung, D.S., Zhang, F., Wang, X. and Hsu, C.C, “Face Liveness detection using a flash against a 2D spoofing attack”, IEEE Transactions on Information Forensics and Security, 13(2), pp.521-534, 2017.https://dl.acm.org/doi/abs/10.1016/j.patrec.2023.04.011.
11. Galbally, J. and Marcel, S., “Face anti-spoofing based on general image quality assessment” International Conference on Pattern Recognition (pp. 1173-1178), IEEE 2014.DOI: 10.1109/ICPR.2014.211.
12. Paul Viola Michael Jones, “Rapid Object Detection using a Boosted Cascade of Simple Features”, Conference on Computer Vision and Pattern Recognition 2001.DOI: 10.1109/CVPR.2001.990517.
13. Arashloo, S.R., Kittler, J. and Christmas, W, “Face spoofing detection based on multiple descriptor fusion using multiscale dynamic binarized statistical image features”, IEEE Transactions on Information Forensics and Security, 10(11), pp.2396-2407, 2015.DOI: 10.1109/TIFS.2015.2458700.
14. Raheem, E.A. and Ahmad, S.M.S, “Statistical Analysis of Image Quality Measures for Face Liveness Detection. In 10th International Conference on Robotics, Vision, Signal Processing and Power Applications” (pp. 543-549). Springer, Singapore 2019.https://link.springer.com/chapter/10.1007/978-981-19-0252-9_37.
15. A toolkit for making real-world machine learning and data analysis applications (Electronic resource)– Accessmode:https://github.com/ davisking/ dlib/ blob/master/ python_examples/face_detector.py

*******