International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Conference Proceedings: Volume: 05 ACME – 2026 DOI:10.55041/ISJEM.ACME083
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

# A Robust Design of Real-Time Resilient Smile Recognition System

**Mr. T. KATAIAH** [1]
Assistant Professor, Department of AIDS
Annamacharya Institute of Technology and Sciences, Tirupati – 517520, A.P.
kata09573@gmail.com

**K. RAVI SHANKAR** [2]
UG Scholar, Department of AIDS
Annamacharya Institute of Technology and Sciences, Tirupati – 517520, A.P.
shankerr270@gmail.com

**S. NIKHIL** [3]
UG Scholar, Department of AIDS
Annamacharya Institute of Technology and Sciences, Tirupati – 517520, A.P.
sankarapunikhilkumar@gmail.com

**K .YAMINI** [4]
UG Scholar, Department of AIDS
Annamacharya Institute of Technology and Sciences, Tirupati – 517520, A.P.
konayamini5@gmail.com

**M. VARSHA** [5]
UG Scholar, Department of AIDS
Annamacharya Institute of Technology and Sciences, Tirupati – 517520, A.P.
varsha1232004@gmail.com

***Abstract*—** Facial expressions are an integral part of human-computer interaction. Smiles are used to convey emotions such as happiness, relaxation and comfort in daily-life situations. So, building a system that can detect smiles accurately is useful for many real-life applications. In this work, we propose a real-time smile recognition system by using both deep learning and classical machine learning models. The proposed system first uses a Haar Cascade-based face detector and then image preprocessing such as grayscale conversion, resizing and normalization. Classical machine learning models extract features using the Histograms of Oriented Gradients (HOG) method, whereas deep learning models learn features directly from the data. Several models such as CNN model, MobileNetV2 model, and SVM, Random Forest and KNN algorithms are used, so that we can compare the performance and understand the pros and cons. A unified-prediction-setup is designed, which selects among the different models to predict whether a person is smiling or not. Additionally, a confidence score is added to the proposed output. Effectiveness of the different models is assessed using standard performance metrics such as accuracy, precision, recall, F1-score and confusion matrix. From the obtained results, we advise that deep learning models have higher accuracy, while traditional models are faster and require fewer computational resources. Finally, the system is implemented for real-time use using webcam input, which is useful for smart cameras, monitoring systems and interaction platforms. Overall, the proposed system seeks to find a balance between accuracy and efficiency, while having a good performance in real-life scenarios.

The study concludes: Smile Recognition, Face Detection, Human–Computer Interaction, Convolutional Neural Network (CNN), MobileNetV2, Support Vector Machine (SVM), Random Forest, K-Nearest Neighbours (KNN), Histogram of Oriented Gradients (HOG), Deep Learning, Machine Learning, Real-Time Detection, Computer Vision.

## I.INTRODUCTION

The field of face expression recognition is a crucial and a fast developing field in computer vision and human computer interaction. It is concerned with making machines capable of recognizing the human feelings based on what they observe. Of all the possible facial expressions, smile detection is by far one of the least researched, though being useful in understanding the emotional state of the person and potentially useful in user experience analysis fields, as well as smart surveillance and interactive applications. As the deep

ISSN: 2583-6129
DOI:10.55041/ISJEM.ACME083

learning advanced, smile detection systems have become more trusted, quicker, and they can be applied in real-time.

The paper presents a web-based smile detector system in this paper, where both the deep learning and conventional machine learning are integrated into one system. The system will be based on the Django web framework, thus the user can easily communicate via the browser by uploading pictures or by using a live video stream. It is accompanied by popular libraries like OpenCV, which is used in face detection, and TensorFlow/Keras, which is used in deep learning model building and deployment.

The system will be constructed on the basis of a variety of models in order to enhance flexibility and performance comparison. These consist of a Convolutional Neural Network (CNN), MobileNetV2 (a trained deep learning model), and conventional machine learning algorithms, including Support Vector Machine (SVM), Random Forest (RF) and K-Nearest Neighbors (KNN). CNN model operates using grayscale images to lower the computation power whereas MobileNetV2 operates using the transfer learning model using the RGB images to capture more complex features. In the case of the classical models, features of Histogram of Oriented Gradients (HOG) are also extracted to convey the significant patterns in facial images.

The major advantage of the suggested system is that it is capable of first detecting the face and then classifying it. This helps in that only the part of face that is relevant gets looked at and this leaves less computation to be done. Upon face detection, the face is then preprocessed and fed into the chosen model to identify the smiling by determining whether the individual is smiling or not. The system also includes a confidence score of each prediction giving an impression on the reliability of the result. Moreover, webcam input is also supported in the form of real-time smile detector and thus the system can be put into practice in the real world.

Set of standard metrics like accuracy, precision, recall, and F1-score are used in order to assess the performance of the system. Confusion matrices as well as graphical comparisons are also produced so as to be able to visualise the performance of each model in a better way. The system is able to compare various approaches and select the best one upon combining the various models into a single platform, making the process of comparing these approaches very easy.

## II. RELATED WORK

The possibility to identify the facial expression and in particular smile detection is one of the most important research questions of the computer vision because of the amount of applications it has. Previous research was largely concerned with the conventional methods of machine learning which used handcrafted features. Usually, such techniques were Histogram of Oriented Gradients (HOG) and support vectors machine (SVM), k-nearest neighbors (KNN), and random forest (RF) classifiers. These methods were effective by deriving the bare essentials of the grayscale images of faces and offer fairly good results in controlled conditions.

Two of the most important advances in the introduction of deep learning are the Convolutional Neural Networks (CNNs). CNNs do not require manual feature extraction since important features are learned automatically by looking at the images, unlike the traditional methods. Since CNNs can extract intricate patterns, they have been extensively applied in other fields such as smile and non-smile classification and incurred greater accuracy than previous methods.

Moreover, the methods of transfer learning with the help of ready-made models like MobileNet V2 are gaining popularity. The models are trained on massive data such as ImageNet and can be trained on smaller tasks with reduced amounts of data and training time. The method is particularly effective in the context of smaller datasets, since it will assist in the reduction of overfitting and will serve to enhance the overall performance.

The combination of deep learning with standard machine learning techniques is also researched recently. In these kind of hybrid methods, the features obtained through CNNs are fed into the classifier, such as SVM or Random Forest to improve its accuracy further. Meanwhile, face recognition with OpenCV and Haar Cascade classifiers remains a trusted preprocessing phase of face recognition and extraction of facial features followed by face recognition.

© 2026, ISJEM (All Rights Reserved)

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Conference Proceedings: Volume: 05 ACME – 2026 DOI:10.55041/ISJEM.ACME083
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

The system suggested in this work works in a similar direction combining several techniques. It contains CNN model of grayscale images as well as the transfer learning model of RGB images based on MobileNetV2 and the traditional machine learning models which are based on HOG features. The given combination makes it possible to compare different approaches in the same framework. Moreover, the web interface system is interactive, and applied in real-time because it uses the Django platform.

## III. METHODOLOGY

The methodology used in this work to design the Hybrid Learning for Smile Recognition (HLSR) system is presented in this section. The proposed system is based on a hybrid method for real-time smile detection through a combination of a deep learning method for feature extraction and an ensemble method for smile classification. The proposed system includes real-time image acquisition, pre-processing, feature extraction, classification, and finally real-time output. Basically, the system architecture of the proposed system is described in the following fig 1.

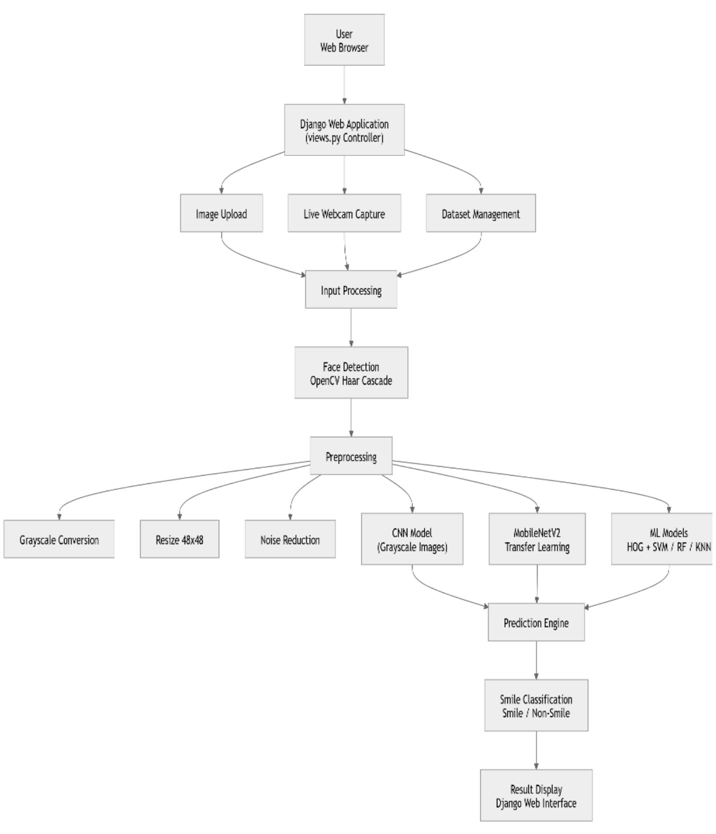



Fig 1: Proposed system architecture

### *A. System Overview*

The smile detection system involves a series of organized steps, such as image capturing, face recognition, preprocessing of images, feature extraction, training and models of classification. The implementation of the system has been carried out utilizing a web interface whereby the user is enabled to engage the system through the use of the web interface whereby the user can post images or apply the webcam to take pictures. The interface will provide the output to the user.

### B. Image Acquisition

The sources of the images that are being acquired by the system are different. The images are discovered via the images uploaded by the user via the interface as well as the images taken using the webcam. The system enables one to get in real time images through a webcam and using the interface. The face detection module is used to achieve the images.

### C. Face Detection and ROI Extraction.

Open CV Haar Cascade classifier is applied on face detection module. The pictures are turned into black and white and face scanned. The biggest face identified is selected and applied as the Region of Interest (ROI).

### D. Preprocessing

The face ROI is extracted by undergoing the image preprocessing where images are transformed into grayscale, rescaled, and normalized depending on the model requirements. In CNN models, the images are rescaled to a size of (32x 32) pixels and brought to the range [0,1]. In case with MobileNet models, images are re-sized to (224×224) pixels and are further processed. In classical models, images become converted into grayscale, downsizing them into (64x64) pixels and then manipulated.

### E. Feature Extraction and Training a Model.

It has several models on feature extraction and classification:

1. Convolutional Neural Network (CNN):

Image classification and feature extraction is defined and implemented in a sequential CNN. The network consists of convolutional layers, ReLU activation layer, max pooling layers as well as fully connected layers. The

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Conference Proceedings: Volume: 05 ACME – 2026 DOI:10.55041/ISJEM.ACME083
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

dropout functionality will be applied to avoid overfitting when creating a model. Model compilation is done using Binary Cross- Entropy loss and Adam optimizer.

2. MobileNetV2:
Transfer learning involves the use of the pre-trained model to classify the smile of a person in an image. The base model is applied directly; to distinguish between a smile and a non smile image, there are added more layers.

3. Classical machine learning models:
Histogram of Oriented Gradients (HOG) is employed to extract data of grayscale images and subsequently be trained on machine learning algorithms like SVM, Random Forest and KNN to either label the image as a smile or not a smile.

F. Model Evaluation
The quality of the machine learning model is measured with a set of different parameters, which include accuracy, precision, recall, F1 score. The classification of the model performance is demonstrated using the confusion matrix.

G. Output Classification
This is done using the machine learning model, which classifies the input image of the person face and comes up with a confidence score of the classification. The picture will be described as Smiling or Not Smiling by referring to the confidence score and the set confidence threshold according to which the classification is made. Confidence score is equally applied to demonstrate the dependability of the classification output.

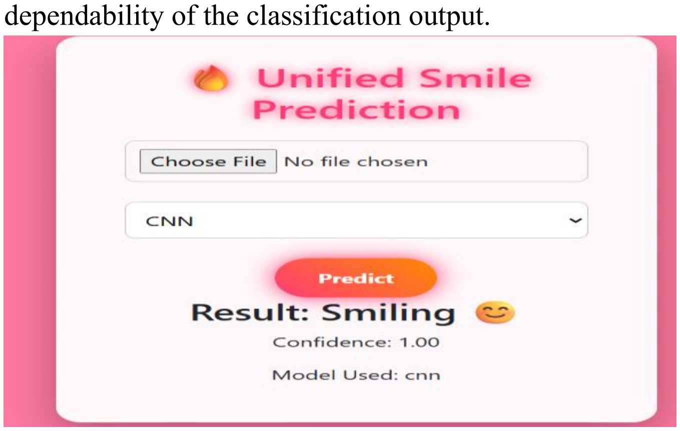



Fig 2: Unified Smile Prediction Result Interface

H. Real-Time Detection and Web Integration

The system is capable of providing real-time smile detection through video streaming. This is achieved by processing each video frame, and the predictions are displayed on the video with bounding boxes and labels. The Django web interface handles all user requests, displays output, and generates visualization outputs such as confusion matrix and accuracy graph.

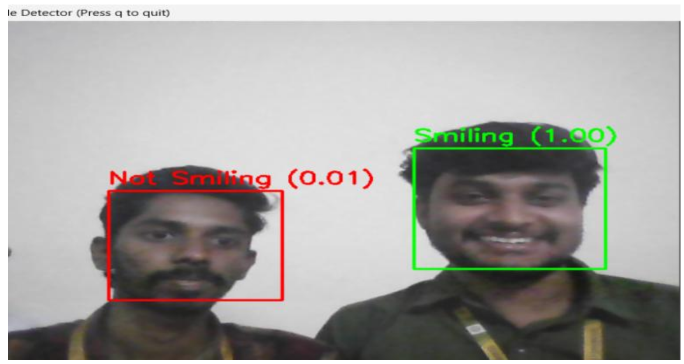



Fig 3: Real-Time Smile Detection Output

## IV. PERFORMANCE ANALYSIS

This section explains the performance of the proposed smile detection system using various machine learning and deep learning models. The evaluation is carried out using standard performance metrics such as accuracy, precision, recall, and F1-score. These metrics help in understanding how effectively the system can distinguish between smiling and non-smiling faces under different conditions.

A. Evaluation Metrics
To evaluate the performance of the models, the following metrics are used:

- Accuracy: Measures the overall correctness of the model.
- Precision: Indicates how many of the predicted smiles are actually correct.
- Recall: Measures the ability of the model to identify all actual smiling cases.
- F1-Score: A balanced metric that combines both precision and recall.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Conference Proceedings: Volume: 05 ACME – 2026 DOI:10.55041/ISJEM.ACME083
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

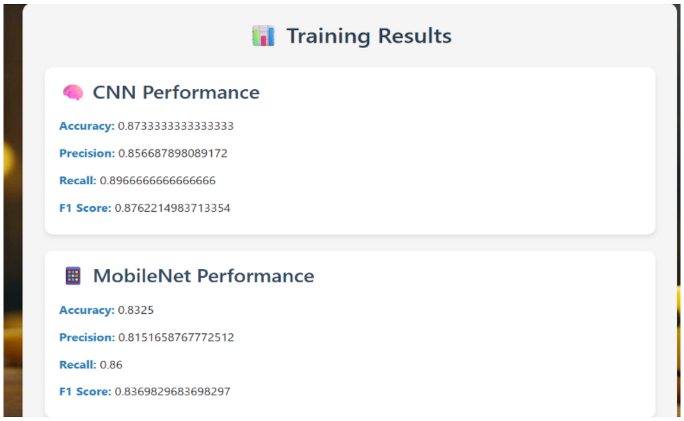



Fig 4.1: The Evaluation Metrics for Deep learning Models

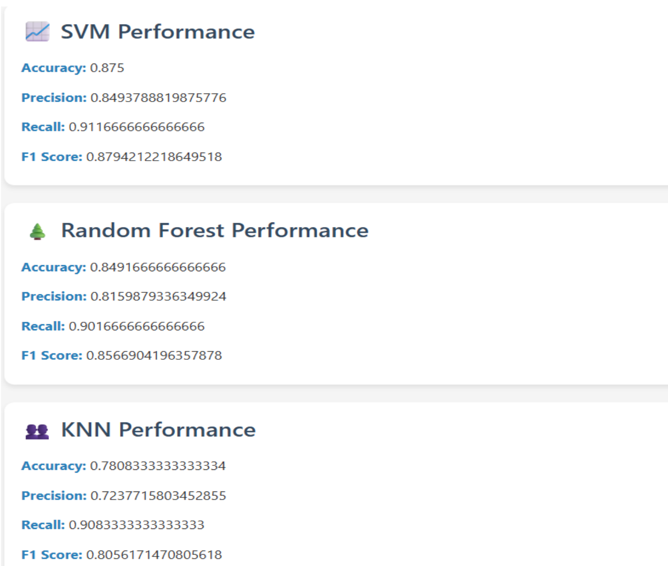
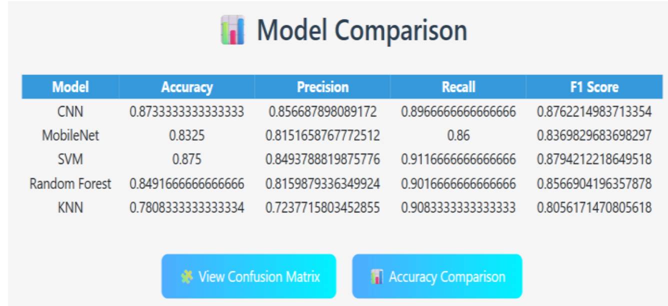



Fig 4.2: The Evaluation Metrics for Machine learning Models

These metrics are calculated using standard evaluation functions to ensure consistent and reliable results.
In addition, statistical methods such as correlation coefficients, t-tests, and p-values are used to compare the performance of different models.

B. Model-wise Performance Comparison
Various models are used to test the system, including CNN, MobileNetV2, and traditional machine learning models such as Support Vector Machine (SVM), Random Forest (RF), and K-Nearest Neighbors (KNN).


Model Comparison

| Model | Accuracy | Precision | Recall | F1 Score |
|---|---|---|---|---|
| CNN | 0.8733333333333333 | 0.856687898089172 | 0.8966666666666666 | 0.8762214983713354 |
| MobileNet | 0.8325 | 0.8151658767772512 | 0.86 | 0.8369829683698297 |
| SVM | 0.875 | 0.8493788819875776 | 0.9116666666666666 | 0.8794212218649518 |
| Random Forest | 0.8491666666666666 | 0.8159879336349924 | 0.9016666666666666 | 0.8566904196357878 |
| KNN | 0.7808333333333334 | 0.7237715803452855 | 0.9083333333333333 | 0.8056171470805618 |



Fig 5: The Model Comparison between models

1. CNN Model
The CNN model is trained using grayscale images of size 32×32 pixels. It consists of convolutional layers, pooling layers, and fully connected layers. This model is efficient and suitable for real-time applications due to its lower computational requirements.
2. Mobile NetV2 Model
MobileNetV2 is a pre-trained deep learning model that uses transfer learning. It processes RGB images of size 224×224 pixels. Due to its deeper architecture and pre-trained weights, it extracts more meaningful features and generally achieves higher accuracy compared to the CNN model.
3. Classical Machine Learning Models
Classical machine learning models such as SVM, Random Forest, and KNN use HOG (Histogram of Oriented Gradients) features extracted from grayscale images. Although these models are faster and require less computational power, they generally perform lower than deep learning models due to limited feature learning capability.

C. Confusion Matrix Analysis

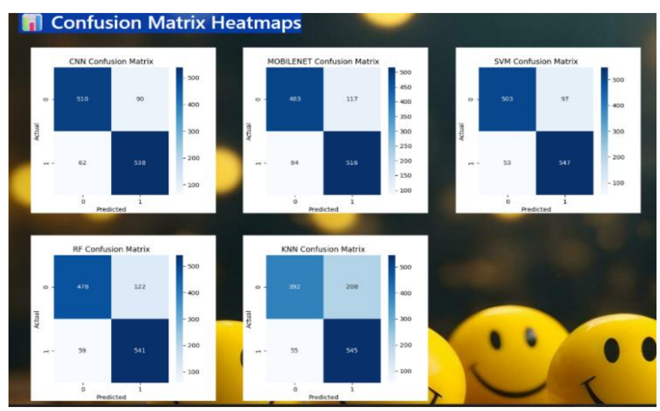



Fig 6 Model-wise Confusion Matrix Analysis for Smile and Non-Smile Classification

Confusion matrices are used to better understand the performance of each model in terms of correct and incorrect predictions. They show the number of true positives, true negatives, false positives, and false negatives. Heatmaps are also used to visually interpret the classification results and identify common errors.

D. Accuracy Comparison

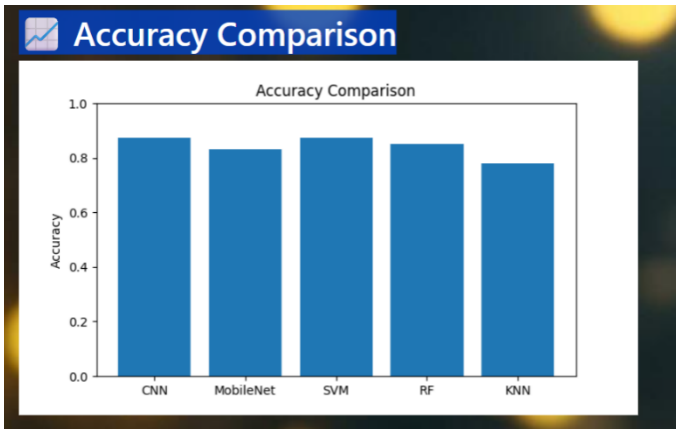



Fig 7: Model-wise Accuracy comparison for Smile and Non-Smile Classification

All the models are evaluated using an accuracy comparison graph. The findings indicate that deep learning, especially MobileNetV2, is more accurate. CNN model offers an ideal compromise between performance and speed and thus can be used in real-time applications.

*E. Real-time Performance Evaluation.*

The actual functioning of the system is tested in terms of responsiveness, load distribution as well as load testing. Webcam input is also used in testing the system with real-time smile detection. This is done through the CNN model since it has a higher processing speed. Haar Cascade classifiers are applied to perform face detection and the predictions are displayed as well as confidence scores. To enhance performance in the real-world scenario, an adaptive threshold (approximately 0.4–0.5) is applied.

V. RESULTS AND DISCUSSION

The smile detection system was also experimented using the uploaded facial images and input of the web camera to test the effectiveness of the system to recognize facial expressions as smiling or not smiling. The proposed system has a testing phase through which the efficiency of the machine learning and deep learning algorithms and usability of the proposed web interface developed using the Django framework are tested.

The proposed system takes inputs in the form of an input image and has to go through different phases to reach the prediction results through face detection, preprocessing, feature extraction and classification. The face detection algorithm of the proposed system is Haar Cascade algorithm of OpenCV, which recognized the facial part as the Region of Interest (ROI). The preprocessing phase of the proposed system entailed the conversion of the image into a grayscale and the process of downsizing the image to fit into the input image of the machine learning and deep learning algorithms.

Other models of classification that are supported by the framework include CNN, MobileNetV2, SVM, Random Forest, and KNN. The experiment achieved the indication that SVM was most accurate. Nevertheless, the CNN model also performed highly when automatic learning of the features was done. MobileNetV2 model provided effective predictions to the classification model. Another model that provided consistent performance was the Random Forest model which applied learning through ensemble. Nevertheless, KNN was a relatively poor model in terms of accuracy and high prediction time.

Using the interface, the user is able to upload the image and choose the model. The classification of the smile and the confidence level is shown on the output screen. Validity of the performance of the proposed system has been tested by computing the accuracy, precision, recall, and F1 score. The effectiveness of the proposed system to classify the facial expression was also indicated by the confusion matrix. The suggested framework was good in categorizing the facial expression with small mistakes in classification.

All in all, the findings demonstrate that the suggested system is robust and efficient when identifying the smiles in both the static and real time scenarios. This flexibility of the system is enhanced by the employment of a number of different models within one umbrella of a web application. Emotions can be recognized with the assistance of the suggested system. An improvement of the system in future is to enhance the training data and strength of the system.

VII. CONCLUSION

This research aimed to create a smile detection system, which was a combination of deep learning and traditional machine learning through a web environment. The system could use various models, namely Convolutional Neural Network (CNN), MobileNetV2, and traditional models, such as Support Vector Machine (SVM), Random Forest, and K-Nearest Neighbors, among others, to classify into two groups.

Using the Django environment, the implementation was able to work successfully when using image data and video data. Face detection on openCV could help to enhance efficiency and accuracy since the objects that were processed were limited to only those areas of interest.

The efficiency of the system was possible to demonstrate through various metrics such as accuracy, precision, recall and F1 score whereby the deep learning models, namely CNN and MobileNetV2, could perform better than the traditional models.

This research was also a strong one in that it enabled users to select models to use in prediction, consequently enabling them to deliver their results in real time with some confidence. The adaptive thresholding method could be permitted to make accuracy measures with real implementation.

Live video streaming smile detection allowed demonstrating how it could be used in real human to computer interaction.

Conclusively, this research was in a position to create an efficient, scalable and effective smile detection system and thus it can be applicable in practice. To improve on its effectiveness in other lighting conditions, its applicability to multi-class emotion recognition, as well as its ability to run on real devices could be used in future research.

REFERENCES

[1] I. Adjabi, A. Ouahabi, A. Benzaoui, and A. Taleb-Ahmed, “Past, present, and future of face recognition: A review,” *IEEE Access*, vol. 8, pp. 100 000–100 020, 2020.

[2] F. L. Malallah and A. A. Al-Jubouri, “Smile and non-smile emotion recognition using deep convolutional neural networks,” in *Proc. IEEE Int. Conf. Smart Computing*, 2020.

[3] T. Küntzler, T. T. A. Höfling, et al., “Automatic facial expression recognition under controlled and uncontrolled conditions,” *IEEE Trans. Affective Computing*, 2021.

[4] Y. Liu, Z. Liu, and K. Zhang, “A robust approach for smile recognition using deep convolutional neural networks,” in *Proc. IEEE Int. Conf. Image, Vision and Computing*, 2022.

[5] H. Wu, Y. Liu, and X. Zhang, “Fast facial smile detection using convolutional neural networks in intelligent environments,” *IEEE Trans. Image Processing*, 2020.

[6] L. Liu, W. Gui, and J. Zhang, “Real-time pose-invariant smile detection using regression forests,” *IEEE Access*, 2019.

[7] S. Jia, S. Wang, and L. Zhang, “Detection of genuine and posed facial expressions of emotion,” *IEEE Trans. Multimedia*, 2021.

[8] R. Syaputra, D. Syamsuar, et al., “Smile detection using histogram of oriented gradients and support vector machines,” in *Proc. IEEE Int. Conf.*, 2021.

[9] M. Nascimben and T. Zoëga Ramsøy, “A minimal setup for spontaneous smile quantification,” *IEEE Trans. Affective Computing*, 2020.

[10] K. Tanaka, S. Ogura, and others, “Flexibility and productivity in IoT programming: A case study,” *IEEE Internet of Things Journal*, 2024.