**Journal of Information Systems Engineering and Management**
2025, 10(9s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

**Research Article**

# Exploring Face Detection Algorithms: A Comparative Overview

Arti Deshpande[1], Bhushan Jadhav[2], Anshul Parkar[3], Dhruv Mehta[4], Devansh Motwani[5], Vishal Mishra[6]
[1]*Associate Professor, Department of Computer Engineering, Thadomal Shahani Engineering College, Mumbai, India.*
*arti.deshpande@thadomal.org*
[2]*Assistant Professor, Department of Artificial Intelligence and Data Science, Thadomal Shahani Engineering College, Mumbai, India.*
*bhushan.jadhav@thadomal.org*
[3]*Department of Computer Engineering, Student, TSEC, Mumbai*
*anshulparkar@gmail.com*
[4]*Department of Computer Engineering, Student, TSEC, Mumbai*
*dhruvashmm1@gmail.com*
[5]*Department of Computer Engineering, Student, TSEC, Mumbai*
*motwanidevansh5@gmail.com*
[6]*Department of Computer Engineering, Student, TSEC, Mumbai*
*vishalmishra0427@gmail.com*

**ARTICLE INFO**

Received: 02 Nov 2024
Revised: 22 Dec 2024
Accepted: 05 Jan 2025

**ABSTRACT**

There are diverse approaches in face detection and recognition, highlighting algorithms, datasets, and practical applications. The Haar Cascade Classifier is a popular technique for fast and simple detection, though it is hindered by false positives and reduced accuracy in complex scenes. Advanced methods, such as combining machine learning algorithms with feature extraction techniques like Principal Component Analysis (PCA) and Three-Patch Local Binary Pattern (TPLBP), are examined for their higher recognition rates and improved accuracy. The use of datasets like ORL and Sheffield demonstrates robustness under various conditions, including changes in pose, lighting, and expression. Hybrid methods like Scale-Invariant Feature Transform (SIFT) and Principal Component Analysis integration, achieve remarkable accuracy on multiple datasets despite image variances. Notable contributions include the Dual Shot Face Detector algorithm for addressing challenges like scale variation and occlusion and the Viola-Jones algorithm, which is used for real-time detection with minimum computational overhead. The study also evaluates the performance of different algorithms across datasets and real-world applications, including automated attendance systems and emotion detection. This study conducts a comparative analysis of existing face detection methods, evaluating their accuracy, efficiency, and scalability across diverse environments. By identifying strengths and limitations, the study aims to provide insights and propose potential enhancements to guide the development of more robust and adaptable face detection solutions.

**Keywords:** Face Detection, Machine Learning, Sliding Window Approach, DeepFace, Integral Image Processing, Gradient-Based Features

## I. INTRODUCTION

By analysing facial features like the eyes, nose, mouth, face recognition technology allows algorithms to recognize or validate human faces from pictures. It integrates artificial intelligence and image processing to produce precise results, and it is widely used in applications like cell phones security and attendance systems etc. From simple template-based techniques to sophisticated strategies powered by AI and computer vision, face detection and recognition technologies have changed throughout time. Early techniques like geometric modelling and pixel intensity comparisons hampered the ability to handle complex real-world scenarios. Nowadays, these systems are essential in fields including human-computer interaction, education, healthcare, and security. The Viola-Jones algorithm, a groundbreaking method that integrated integral image computation, Haar-like features, and AdaBoost [1] for feature selection, was a major turning point. This laid the way for subsequent developments by enabling real-time facial detection. However, problems like false positives, complicated backdrops, and changes in

scale or lighting frequently plagued conventional techniques like Haar Cascade Classifiers. These restrictions prompted the creation of sophisticated algorithms that combine deep learning and machine learning methods.

Previous methods for face identification involved techniques like CNNs (Convolutional Neural Networks) and SIFT (Scale-Invariant Feature Transform), and SVMs (Support Vector Machines). Deep learning models like AlexNet, GoogleNet, and SqueezeNet are used in modern systems and provide impressive gains in accuracy and efficiency. The use of transfer learning has further revolutionized this field by enabling the reuse of pre-trained models, reducing computational demands, and training times.

This paper examines the evolution of face detection and identification systems, highlighting advancements, real-world applications, and challenges such as masked face recognition. It contrasts several approaches, such as deep learning, rule-based detection, and Haar Cascade Classifiers. The study suggests enhancements for scalable and dependable solutions in domains like emotion detection and attendance management.

## II. A STUDY ON THE VIOLA-JONES ALGORITHM FOR FACE DETECTION

This method is a classic approach to face detection, introduced by Paul Viola and Michael Jones [17]. This approach enables fast and accurate detection using three major components [9]. First is an integral image for feature computation. It is a preprocessing step where a rectangular Haar-like feature is used and also captures the contrast between different areas of the picture or image, thus reducing the computational complexity, Second is AdaBoost for feature selection, machine learning algorithm is used to select similar parts of thousands of capable features from the image, combine the feature into the strong classifiers and give the difference between faces and non-faces, thus significantly improving the detection power. The third component is the Attentional Cascade for Efficient Computational Resource Allocation, a series of increasingly complex classifiers that are arranged in pipelines. This complex classifier handles the difficult case after it has gone through a simple classifier.

A common issue with the Viola-Jones algorithm is that it can detect multiple faces at a time; that is, it overlaps the boundaries of the boxes of the faces. To solve this issue, a post-processing step [9] was introduced that retains the bounding boxes with the help of confidence, so the post-processing step is also important to reduce detection redundancy. Later, the use of block features in the Viola-Jones algorithm was identified as one of its limitations [10]. However, it is not designed to process entirely rigid objects like sticks or cups. Therefore, if such objects are present in the image, the algorithm may face limitations. To address this issue, a face detection method based on the Viola-Jones algorithm with composite features was developed, aiming to preserve the face recognition rate and maintain a reasonable level of observability. The composite feature consists of multiple types of feature descriptors, which are Haar-like features, edge-based features, and gradient features to gather more data about the face. This minimizes the false positive detection by enabling it to differentiate between faces and rigid objects.

## III. FACE DETECTION USING HAAR CASCADE CLASSIFIERS

Detection of human faces has become a crucial aspect of human-machine interaction and applications based on computer vision. It is a system in which an algorithm is used to analyse an image and determine which part of the image contains the human face. The human face has various unique features that make it distinct from one another; still, human face detection continues to be a challenging task in real-world applications due to factors like varying lighting, facial expressions, backgrounds, and other uncertainties.

The Haar Cascade Classifier algorithm is proposed by Viola-Jones [2], which uses various haar-like features to analyse image regions based on intensity contrasts. Whereas the cascade structure is employed to sequentially reject non-face regions while refining the search for faces the initial step of this algorithm is facial extraction with haar-like features; these are rectangular features used to detect patterns such as edges, lines, and textures in images. Each feature computes the difference in pixel intensities between adjacent rectangular regions. Then the classifiers work as multiple cascading stages where they analyse the window in the image, and the window that fails at any stage is discarded. While the one that passes all the stages is considered as a face.

These stages consist of weak classifiers too, which contain a single Haar-like feature. These classifiers are therefore combined into a strong classifier using the Adaboost algorithm [1], which focuses on harder-to-classify examples by assigning them higher weights during training [1][2]. This process of window detection can be done by using an approach called the sliding window approach. The training phase follows, during which the model is taught using both classifier tuning and samples that are both positive and negative. In positive and negative samples, the system

is trained on a dataset of positive samples (images including faces) and negative samples (images not including faces). In cascade tuning, all stages are trained to maximize the detection rate and minimize false detection rates. In the initial stages it is easier to distinguish the non-face regions and face regions from the image, but in the later stages it becomes more complex. So, the cascade structure ensures that the non-face regions are rejected and computational overhead is reduced, thus improving the efficiency of the Haar cascade classifier and thus making it adaptable for real-time face detection.

## IV. FACE DETECTION USING DEEP TRANSFER LEARNING

Face recognition is one of the significant methods that has been used in many applications for security and surveillance purposes. K. Alhanaee et al. [5] have proposed that deep transfer learning is used on large datasets where it surpasses humans for face recognition tasks on difficult, unconstrained datasets. They have made use of transfer learning using three convolutional networks which were pretrained on the training dataset and the experimental study shows high performance in accuracy and training. A. P. Nivetha et al.[6] have suggested the success of deep transfer learning applications with small datasets could pave the way for low-cost and non-invasive facial screening and detection. Deep learning techniques are some of the most used techniques in different applications of computer vision. It is a technique that features automatic classifying and learning of images.

Accurate attendance management is crucial for organizational efficiency, especially in education. As suggested by the author [5], traditional methods like calling names or using paper sign-ins are time-consuming and prone to error. Face recognition offers a more stable and user-friendly alternative. With the advancements in deep learning, it has become essential to maintain secure and automated identification. It further enhances face recognition by reusing pre-trained models to obtain good accuracy and reasonable training time**.** Convolutional Neural Networks (CNNs) are the most widely utilized deep learning technique [7]. The CNN networks include various types of networks such as AlexNet, FaceNet, VGG-Face, ResNet, DeepFace, GoogleNet, and SqueezeNet, with a large number of public datasets present. These CNN models are trained for most suitable learning face representations automatically for better understanding of the computer. Transfer learning is a type of machine learning where for a particular task a model is built and is used again in the next task from the starting point to be modified, it is the optimal model for saving time, optimization, and attaining superior performance [5], the main aim is to classify the images and compare the performance metric with the other metrics

Pre-trained CNN models provide many advantages when developing a machine-learning device. They remove the need to build models from scratch, which can be computationally expensive and time-consuming. One of the earliest CNN architectures, AlexNet has been widely used in object classification tasks. It analyses images with the dimensions 227 × 227 × 3 (RGB) and gives labels to the objects that it recognizes as shown in figure 1. The network is known for its high accuracy and has been fine-tuned for various purposes [5].

GoogleNet is 22 layers deep with 5 pooling layers as shown in figure 2, there are total 9 initiation models which are stacked continuously, it uses 1x1 spiral filter, Because of the layer reduction and parallel network implementation, it has extremely strong calculational and logical efficiency, and the model is small-scaled model than others.[5]. SqueezeNet framework as shown in Figure 3 is a deep convolutional network which has 18 layers, it is pre-trained. It categorizes pictures, the network has gained an understanding of compound function representations. The aim of using this is to construct Convolutional Neural Network known as VGG16 is one of the best visions models it has a huge number of hyper-limits, utilize a homogenous padding and optimum layer of a 2x2 medium, and there are 16 layers this union is a huge bloc and has around 138 million(approx) limits [8].

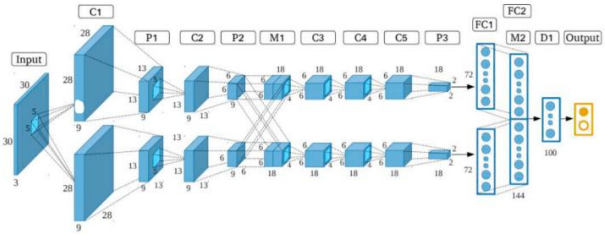



Figure 1: AlexNet Architecture [5]

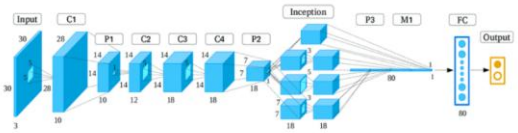



Figure 2: GoogleNet Architecture [5]

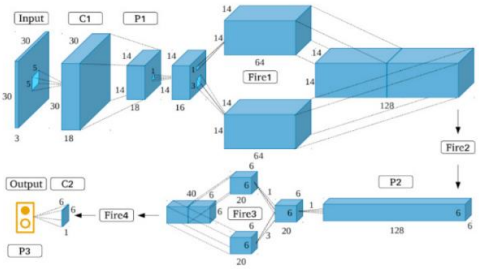



Figure 3: SqueezeNet Architecture [5]

The dataset used for the research [5] consists of 200 images, captured using an iPhone 12, the dataset was organized into 10 categories, each category contains 20 images, SqueezeNet along with AlexNet use 227 x 227, while GoogleNet utilizes 224 x 224 [5]. For VGG16 the whole dataset contained 350 images with 70 images for each sort of appearance, a total of 200 pictures, 40 of each type are used in the organising cycle, for examining the frameworks a total of 150 pictures, 30 of every type are used [8].

The author [5] observed that the AlexNet model achieved a perfect validation accuracy of 100% at the cost of a longer training time (76 minutes). SqueezeNet was followed closely with a validation accuracy of 98.33% and a much-reduced training time (26 minutes). GoogleNet achieved an accuracy of 93.33%, taking 39 minutes to complete the training. The experimental study given by author [5] indicates that while AlexNet provides the best results in the terms of accuracy, SqueezeNet offers a balance between accuracy and efficiency, making it the best choice for systems with limited amounts of resources. Transfer Learning saves significant time and resources by re-using pre-trained models, often improving performances on smaller datasets but the performance is limited by the suitability of the model, with the risk of overfitting if the new data is too different from the original training data

## V. RULE - BASED FACE DETECTION IN FRONTAL VIEWS

The rule-based face detection approach relies on a hierarchical, knowledge-driven pattern recognition system, building upon the research conducted by G. Yang and T.S. Hang [4]. This face detection method utilizes multiresolution or mosaic images as its foundation. The algorithm works by dividing the image into low-resolution blocks called quartet images, where facial features like eyebrows, eyes, nose and mouth are detected as illustrated in figure 4. The idea proposed by Author [4] is very close to the approach used in rule-based algorithms for detecting human faces, but it is more computationally intensive. It is applied to an entire range of cells to determine the best cell dimensions out of the given range of cells.

The algorithm was evaluated using frontal view images obtained from the European ACTS M2VTS database [13] , which contains a set of frontal views of 37 different people. This algorithm gives 100 percent accuracy in identifying the correct facial candidates, whereas an accuracy of 86.5 percent is found for the strictest facial conditions, which include the facial features. Unlike earlier approaches, this algorithm gives more generalized rules and optimizes feature detection by focusing on the symmetry and structure of the human face. The system is hierarchical and ensures minimal false negatives.

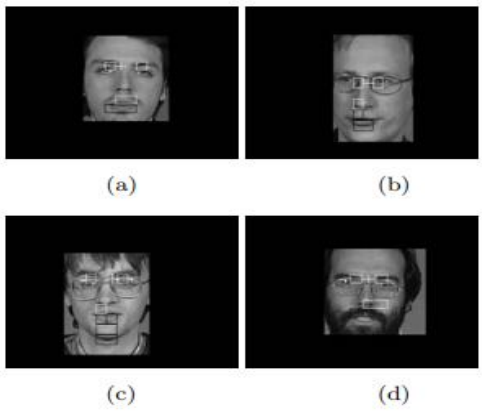
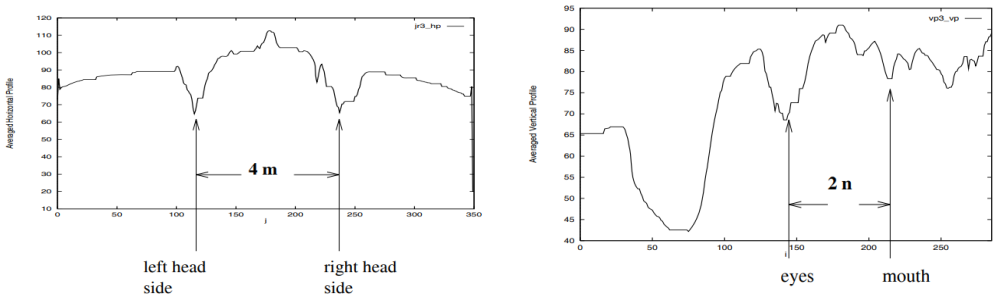



Figure 4: Problems in facial feature detection.[12]

Regardless of the algorithm suggested by G. Yang and T.S. Huang [4], a rule-based approach for face detection analyses the horizontal profile of an image as shown in Figure 5(a) and vertical profiles of an image as shown in 5(b). The horizontal profile is generated by averaging pixel intensities across each column of the image, while the vertical profile is created by averaging pixel intensities along each row. In this context, local minima in the horizontal profile indicate the left and right edges of the head, whereas significant local minima in the vertical profile correspond to facial features such as hair, chin, eyes, mouth, and eyebrows.




(a) (b)

Figure 5: (a) Horizontal Profile (b) Vertical Profile. [12]

The rule-based algorithm highlights the advantages of the mosaic-based approach in terms of simplicity and computational efficiency. The algorithm's output has been effectively utilized to manage the grid's placement in dynamic link matching processes [3]. It avoids complex calculations while achieving robust results, allowing for the use of rectangular cells. The use of preprocessing steps to estimate dimensions ensures reliable performance for detecting key facial features. The algorithm supports multiple domains requiring multimodal verification techniques, such as teleservices and teleshopping applications."

## VI. FACE DETECTION USING SUPPORT VECTOR MACHINE (SVM)

SVM is a kind of regression technique, which is a two-class labelling system [11]. An object to be classified is interpreted as a position in an n-dimensional area, and are known as characteristics of SVM. It conducts classification by establishing a hyperplane, represented as a line in two-dimensional space or a plane in three-dimensional space, ensuring that points belonging to the same category lie on one side as shown in figure 6 and 7.

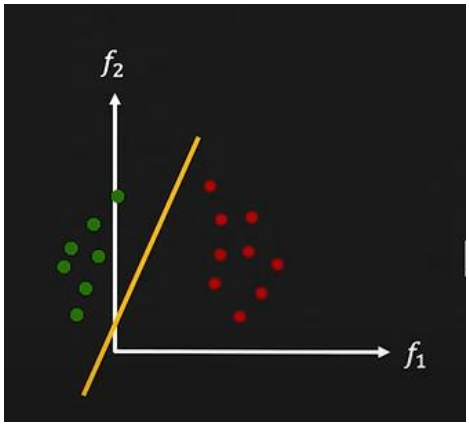



Figure 6: Linear Decision Boundary in 2D [11]

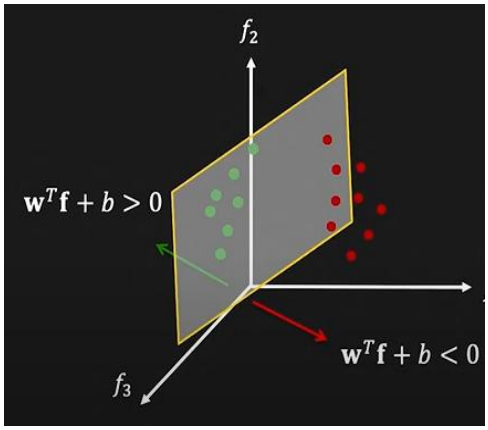



Figure 7: Linear Decision Boundary in 3D [11]

SVM points to the most dominant specimens, known as Support Vectors [11]. These support vectors are prototypes found nearest to the ruling surface being constructed; this area or part is called the margin. In this way, it enhances the distance to a point in any one of the categories; this interval is bounded, and points that fall exactly on them are known as supporting vectors shown in figure 8.

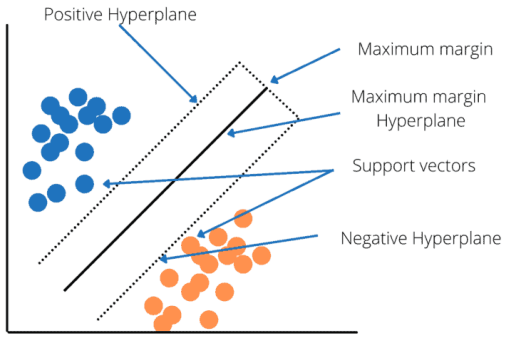



Figure 8: Support Vector Machine (SVM) Visualization [11]

Face detection using SVM classifies whether a given region of a photo or video contains a face or not. It can be explained step by step as follows: First, data preparation is done in which data is collected containing faces. In the same step, image preprocessing is done in which the image is converted into grayscale and resized to a fixed size. After this feature extraction is done, instead of using raw pixel data, extract features like HOG (Histogram of Oriented Gradients) [14] that capture edges orientation and local gradients. Similarly, PCA (Principal Component Analysis) [15] can also be used to reduce dimensionality while preserving important features. In the next step, the extracted feature (i.e., HOG feature) from the previous step is used in training the SVM classifier. There are two types of support vector machines, depending on the kind of data that is linear SVM, sometimes referred to as simple SVM, and nonlinear SVM, which is resolved by the kernel approach. SVM finds an optimal hyperplane that maximizes the margin between two classes (i.e., face and non-face). To detect an exact face in an image Sliding window approach [16] is used. It slides a fixed size of window across the entire image at various scales and positions. This small window extracts features from the image and uses it in the trained SVM. Then the post-processing step is done if multiple detections overlap each other for the same face or a different face. NMS (Non-Maximum Support) is applied, which retains only the detection with the highest confidence score.

One of the major advantages of the SVMs is ease of use, implementation, interpretation, and understanding. It can be as trivial as loading a library in Python, assembling your training data, providing it to the fit function, and calling predict to give the accurate category to a latest object. In many practices, points are incapable of being separated by hyperplanes; that is, SVM can struggle with non-linearly separable data, but techniques like kernel tricks help to manage this limitation. These methods allow for efficient separation in high-dimensional space.

## VII. COMPARISON STUDY OF FACE DETECTION METHODS

| Methods<br><br>Parameters | Haar Cascade [1][2] | Deep Transfer Learning [5][6] [7][8] | Viola Jones [9][10] | Rule Based face detection [12] [13] | Support vector Machine [15] [26] |
|---|---|---|---|---|---|
| **Dataset** | FERET, LFW | FERET, KREMIC | FERET, CMU, MIT | European ACTS M2VTS | FERET, LFW |
| **Algorithm** | Haar-like Features, AdaBoost | Deep Transfer | AdaBoost | Rule-Based Detection | SVM |
| **Approach used** | Sliding Window, Integral Image | Data Augmentation | Sliding Window, Feature Selection | Mosaic-Based Approach | Feature-Based Classification |
| **Computational Complexity** | Low computational requirements; efficient for simple tasks | High computational requirements due to CNNs. | Moderate; suitable for real-time systems with optimizations. | Computationally intensive due to mosaic-based segmentation. | Moderate, with kernel tricks adding complexity for non-linear problems |
| **Feature Representation** | Haar-like features (edges, lines, textures). | CNN-based features learned from pre-trained models (e.g., AlexNet, GoogleNet, SqueezeNet). | Haar-like features with integral image computation | Horizontal and vertical profiles for facial structure | Histogram of Oriented Gradients (HOG) and PCA for feature extraction. |
| **Advantages** | Fast, lightweight, works well with static images | High accuracy, efficient use of data augmentation to enhance training | Efficient, real-time processing | Simplicity, Generalised rules, Preprocessing | Effective for non-linear problems |
| **Disadvantages** | Struggles with pose/lighting variations | Custom CNN models perform less optimally, some CNN models require high computational power | Struggles with occlusions | Computationally Intensive | Requires careful feature engineering |

## VIII. CONCLUSION

The study analyses face detection methods, emphasizing traditional and modern approaches. It emphasizes the evolution from simpler methods, such as Haar Cascade and Viola-Jones algorithms, to advanced strategies integrating deep learning and transfer learning. Early algorithms like Haar Cascade classifiers are lightweight and suitable for real-time applications, with challenges such as occlusions, pose variations, and complex lighting

conditions. In contrast, deep learning methods, including CNNs and transfer learning with pre-trained models like AlexNet, GoogleNet, and SqueezeNet, deliver superior accuracy and scalability but require significant computational resources. Hybrid approaches, combining feature extraction techniques like SIFT and PCA, demonstrate robustness in handling diverse datasets under varying conditions. Our analysis emphasizes how crucial it is to balance face detection systems' accuracy, computational effectiveness, and flexibility. Addressing issues like overfitting in transfer learning and streamlining resource-intensive algorithms to make them more practical for real-world applications are instances of potential future research.

## REFERENCES


[1] Maale, B. R., & Nandyal, S. (2021). Face detection using Haar cascade classifiers. *International Journal of Science and Research, 10*(3), 1179–1182. https://doi.org/10.21275/SR21306204717

[2] Sharifara, A., Rahim, M. S. M., & Anisi, Y. (2014). A general review of human face detection including a study of neural networks and Haar feature-based cascade classifier in face detection. In *2014 International Symposium on Biometrics and Security Technologies (ISBAST)* (pp. 73–78). IEEE. https://doi.org/10.1109/ISBAST.2014.7013097

[3] Kotropoulos, C., Pitas, I., Fischer, S., & Duc, B. (1997). Face authentication using morphological dynamic link architecture. In *Proceedings of the First International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA 97)*, Crans-Montana, Switzerland, 12–14 March. https://doi.org/10.1007/BFb0015993

[4] Yang, G., & Huang, T. S. (1994). Human face detection in a complex background. *Pattern Recognition, 27*(1), 53–63. https://doi.org/10.1016/0031-3203(94)90017-5

[5] Alhanaee, K., Alhammadi, M., Almenhali, N., & Shatnawi, M. (2021). Face recognition smart attendance system using deep transfer learning. *Procedia Computer Science, 192,* 4093–4102. https://doi.org/10.1016/j.procs.2021.09.184

[6] Nivetha, A. P., Suhail Razeeth, M. S., & Prasanth, K. (2023). A comparative study on face recognition using deep learning approach. *International Journal of Advanced Multidisciplinary Research and Studies, 3*(1), 552–559.

[7] Jin, B., Cruz, L., & Gonçalves, N. (2020). Deep facial diagnosis: Deep transfer learning from face recognition to facial diagnosis. *IEEE Access, 8,* 123649–123661. https://doi.org/10.1109/ACCESS.2020.3005687

[8] Bindu, B. R., & Renuka, M. (2022). Diagnosis of facial recognition using deep transfer learning. *International Journal for Research in Applied Science and Engineering Technology, 10*(7), 1631–1635. https://doi.org/10.22214/ijraset.2022.45519

[9] Wang, Y. Q. (2014). An analysis of the Viola-Jones faces detection algorithm. *Image Processing On Line, 4,* 128–148. https://doi.org/10.5201/ipol.2014.104

[10] Wen-yao, L., & Ming, Y. (2019). Face detection based on Viola-Jones algorithm applying composite features. In *Proceedings of ICRIS*. https://doi.org/10.1109/ICRIS.2019.00029

[11] Shavers, C., Li, R., & Lebby, G. (n.d.). An SVM-based approach to face detection. North Carolina A&T State University, Department of Electrical Engineering, Greensboro, NC, USA.

[12] Kotropoulos, C., & Pitas, I. (1997). Rule-based face detection in frontal views. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (pp. 2537–2540). https://doi.org/10.1109/ICASSP.1997.595305

[13] Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Computer Vision – ECCV 2001* (Vol. 2351, pp. 511–518). https://doi.org/10.1109/CVPR.2001.990517

[14] Ghorbani, M., Targhi, A. T., & Dehshibi, M. M. (2016). HOG and LBP: Towards a robust face recognition system. In *Proceedings of ICDM*. https://doi.org/10.1109/ICDIM.2015.7381860

[15] Phillips, P. J. (1999). Support vector machines applied to face recognition. National Institute of Standards and Technology, Gaithersburg, MD, USA.

[16] Chen, J., Cheng, S., & Xu, M. (2019). A face detection method based on sliding window and support vector machine. *Journal of Computers, 14*(7), 470–477. https://doi.org/10.17706/jcp.14.7.470-477

[17] Great Learning Team. (2022, December 12). Viola-Jones algorithm: Everything you need to know. *Great Learning Blog*. Retrieved January 3, 2025, from https://www.mygreatlearning.com/blog/viola-jones-algorithm