**Evaluating CNN and Random Forest with HOG for Image Recognition**

*Abstract* ***-* Face recognition has become a pivotal technology in various applications such as security, surveillance, and authentication systems. This study presents a comparative evaluation of two distinct approaches for face recognition: Convolutional Neural Networks (CNN) and Random Forest combined with Histogram of Oriented Gradients (HOG). The dataset comprises images categorized for face recognition tasks. Our methodology encompasses a thorough preprocessing pipeline, including image augmentation and normalization, to prepare the data for model training. The CNN model leverages deep learning techniques to extract and learn features directly from the images, while the Random Forest with HOG method utilizes traditional machine learning approaches for feature extraction and classification. Both models are trained and evaluated using consistent metrics to ensure a reliable comparison. This comparative study underscores the effectiveness of CNNs in face recognition tasks due to their ability to learn complex patterns from raw data. It also highlights the potential of combining classical machine learning algorithms with feature extraction techniques like HOG for practical applications. Our findings contribute to the ongoing research in face recognition technologies and provide insights into the selection of appropriate models based on specific application requirements and performance trade-offs.**

***Keywords-* Face Recognition, Convolutional Neural Networks (CNN), Random Forest, Histogram of Oriented Gradients (HOG), Image Recognition, Machine Learning, Deep Learning, Feature Extraction, Comparative Study, Performance Metrics, Security Systems, Surveillance, Authentication Systems.**

1. INTRODUCTION

Face recognition is a critical application of pattern recognition capable of identifying individuals from digital images or video frames. It finds widespread use in various fields such as visual surveillance, demographic data collection, and security systems. Despite extensive research over the years, face recognition remains a challenging task in real-world scenarios due to factors like diverse facial expressions, lighting variations, occlusions, and partial visibility of faces. These challenges necessitate continuous advancements to improve the robustness and accuracy of face recognition systems.

Deep convolutional neural networks (CNNs) have emerged as the leading solution for computer vision and object recognition tasks, including face recognition. These networks have demonstrated remarkable ability to analyze and learn from visual data, leading to significant improvements in recognition performance. However, the practical implementation and use of CNNs can be hindered by issues such as the need for large training datasets and substantial computational resources. Collecting extensive datasets can raise privacy concerns, and the large memory and computational requirements can limit the deployment of CNNs on resource-constrained devices like smartphones and tablets.

In this study, we address the problem of face recognition by comparing two distinct approaches: Convolutional Neural Networks (CNN) and Random Forest combined with Histogram of Oriented Gradients (HOG). Our goal is to evaluate the effectiveness of these models in terms of accuracy, precision, recall, and F1-score. The CNN model leverages deep learning techniques to automatically learn features from the images, while the Random Forest with HOG method employs traditional machine learning for feature extraction and classification.

To ensure efficient processing and resource utilization, we implement strategies to optimize memory usage and reduce computation time. The CNN model's architecture is designed to balance performance and resource demands, making it suitable for deployment on devices with limited computational power. For the Random Forest with HOG approach, the feature extraction process is streamlined to ensure quick and accurate face recognition.

The implementation of these models involves a comprehensive preprocessing pipeline, including image augmentation and normalization, to prepare the data for training and evaluation. Our results indicate that while CNNs offer high accuracy and robust feature learning capabilities, the combination of Random Forest with HOG provides competitive performance with advantages in training speed and interpretability.

This paper is organized as follows. The second section provides an overview of related feature extraction and classification algorithms. The third section presents the details of our models, including the architecture and training process. Section four illustrates and discusses the experimental results, comparing the performance of CNN and Random Forest with HOG. Finally, the last section summarizes the findings and highlights the contributions of this study to the field of face recognition.

By conducting this comparative study, we aim to provide insights into the selection of appropriate models for face recognition applications, considering performance trade-offs and specific application requirements. Our findings contribute to the ongoing research in face recognition technologies and offer valuable guidance for the development and deployment of efficient and accurate recognition systems.

1. RELATED WORKS

There are many pattern recognition applications implemented on different platforms, with face recognition emerging as a predominant research field in recent years. Face recognition typically involves two main steps: face detection and face identification. Numerous classical face recognition methods have been explored in the literature, such as Support Vector Machines (SVMs), which are favoured for their fast solving and strong generalizing capabilities. However, these traditional methods often fall short of achieving significant recognition rates [16].

Several classification algorithms, including linear regression and SoftMax, have been suggested for face classification. These methods have demonstrated noteworthy performances in various studies. For instance, a face recognition model using Principal Component Analysis (PCA) and Adaptive Neuro-Fuzzy Inference System (ANFIS) has been proposed, showing promising accuracy and discrimination capabilities [15]. Nonetheless, the challenge in facial recognition often lies in balancing speed and precision. Traditional methods may not achieve satisfactory speed on devices without extensive parallel computing capabilities, limiting their real-time application effectiveness.

In our study, we aim to address both feature extraction and classifier design for face recognition, focusing on reducing running time and memory requirements. Among the most common dimension reduction techniques used for face feature extraction are Independent Component Analysis (ICA), Linear Discriminant Analysis (LDA), and PCA. The efficiency of PCA as a feature extractor for face recognition has been well-documented [15]. Additionally, hashing algorithms like Locality Sensitive Hashing (LSH) can reduce data dimensions by generating binary codewords for high-dimensional data, although they often require significant storage space and computational cost [14].

Utilizing principal components for face detection and recognition has proven beneficial as it maximizes data variance and minimizes reconstruction error [14]. This approach reduces time complexity by lowering the dimension of the feature space compared to the original image space.

For classifier design in face recognition, several effective methods have been proposed, including SVMs, nearest neighbour classifiers, and minimum distance classifiers. Recently, CNN-based face recognition models have shown superior performance. Applications such as ArcFace and VarGFaceNet have achieved impressive results in uncontrolled environments [17][18]. The success of these applications can be attributed to the availability of large datasets and the depth of the networks.

Parallel and distributed algorithms are commonly used to mitigate high training times in CNNs, significantly benefiting from massively parallel processing for tasks like applying filters on images. Many researchers have successfully ported their networks to GPUs, reaping the advantages of immense parallel processing capabilities [13].

In this work, we address the issues of memory and running time in face recognition. We utilize the PCA technique to reduce data dimensions and implement a CNN-based classifier optimized for parallel computing on GPUs. Unlike previous studies, we minimize the number of layers to reduce computation blocks and weight requirements, thereby balancing speed, and precision. Our model is designed to be efficient in both running time and memory usage, offering a robust solution for face recognition applications.

1. DATASET

The face recognition dataset is meticulously designed to facilitate tasks in computer vision, specifically for training and testing face recognition models. It addresses the need for comprehensive datasets in this field. The dataset comprises a total of 244 images, systematically divided into two primary folders: one for training and another for testing. This organization ensures a clear separation between the data used for model training and evaluation.

The training set contains 80% of the total images, providing a robust foundation for models to learn and generalize the distinguishing features of various individuals. These images capture a diverse range of facial features, expressions, and conditions such as lighting and occlusions, which are crucial for developing models that can perform well in real-world scenarios. The remaining 20% of the images are allocated to the testing set, used to evaluate the performance of the trained models.

Each image in the dataset is labelled with the corresponding individual's identity, enabling supervised learning tasks. These labels are essential for the models to learn the association between images and identities during the training phase. The images are provided in standard formats like JPEG or PNG, making them suitable for input into convolutional neural networks (CNNs) and other image processing algorithms.

The dataset is available for academic research and can be requested for use in developing face recognition systems. It supports significant advancements in computer vision, particularly in the domain of facial recognition.

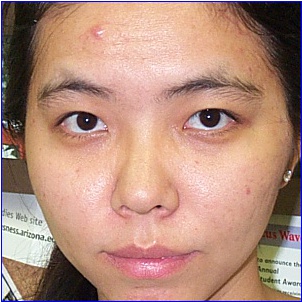
  

Fig.1 Sample Images from Dataset

1. DESCRIPTION MODEL

In our face recognition project, we develop two models: a Convolutional Neural Network (CNN) and a Random Forest classifier utilizing Histogram of Oriented Gradients (HOG) for feature extraction. Each model has its unique approach and advantages, providing a comprehensive comparison in the context of face recognition tasks.

The CNN model is designed to automatically learn and extract relevant features from the input images through multiple layers of convolutions, activations, and pooling operations. The architecture consists of several convolutional layers with 3×3 filters, each followed by ReLU activation functions to introduce non-linearity and prevent overfitting. Max pooling layers are employed to downsample the spatial dimensions of the feature maps, reducing computational load and focusing on the most salient features. After the convolutional and pooling layers, fully connected layers are used to integrate the extracted features and prepare them for classification. The output layer uses a Softmax activation function to produce a probability distribution over the classes (identities). Training the CNN is performed on a GPU to leverage its parallel processing capabilities, significantly reducing the execution time compared to CPU-based training.

The Random Forest classifier is combined with HOG features for face recognition. HOG is a robust feature extraction method that captures the edge and gradient information of the image. The process begins with computing the HOG features for each image, which involves dividing the image into small connected regions (cells) and computing a histogram of gradient directions or edge orientations for the pixels within each cell. These histograms are concatenated to form a feature vector that represents the image, capturing essential structural information while being invariant to geometric and photometric transformations. The Random Forest classifier, an ensemble learning method, is then trained on these HOG features to perform the face recognition task. By combining the strengths of CNN and Random Forest with HOG, our project aims to evaluate and compare the effectiveness of these models in face recognition, highlighting their respective strengths and potential applications.

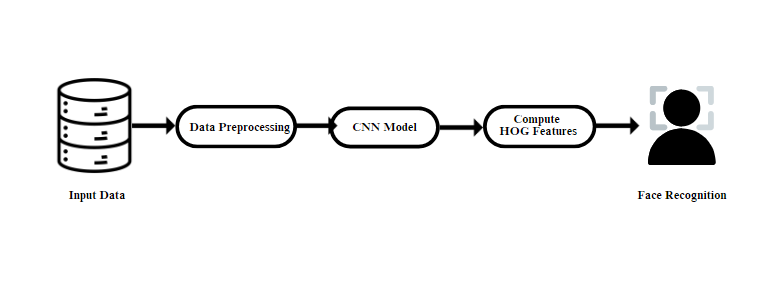


Fig 2. Architecture of the Proposed Approach

V. RESULTS AND DISCUSSIONS

*A. PERFORMANCE ANALYSIS*

TABLE I

TEST SET PERFORMANCE METRICES

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Accuracy** | **Precision** | **Recall** | **F1-Score** |
| CNN | 0.9324 | 0.9552 | 0.9324 | 0.9321 |
| **RF with HOG** | **0.9865** | **0.9892** | **0.9863** | **0.9864** |

The CNN model demonstrated strong performance with a recall of 0.9324 and a precision of 0.9552, indicating its effectiveness in identifying true positives while maintaining a low false positive rate. The F1-score of 0.9321 reflects a well-balanced trade-off between precision and recall, making it a reliable option for the image recognition task. The accuracy of 0.9324 further supports its overall robustness in classification tasks.

In comparison, the Random Forest model, utilizing Histogram of Oriented Gradients (HOG) for feature extraction, outperformed the CNN model with a recall of 0.9865 and a precision of 0.9892. The higher F1-score of 0.9863 suggests a more consistent performance in terms of precision and recall, emphasizing its ability to manage the classification task effectively. The accuracy of 0.9865 further highlights its superior performance, making it a highly efficient model for the image recognition task.

This analysis shows that while the CNN model provides strong results, the Random Forest model with HOG features offers a slight edge in performance across key metrics.

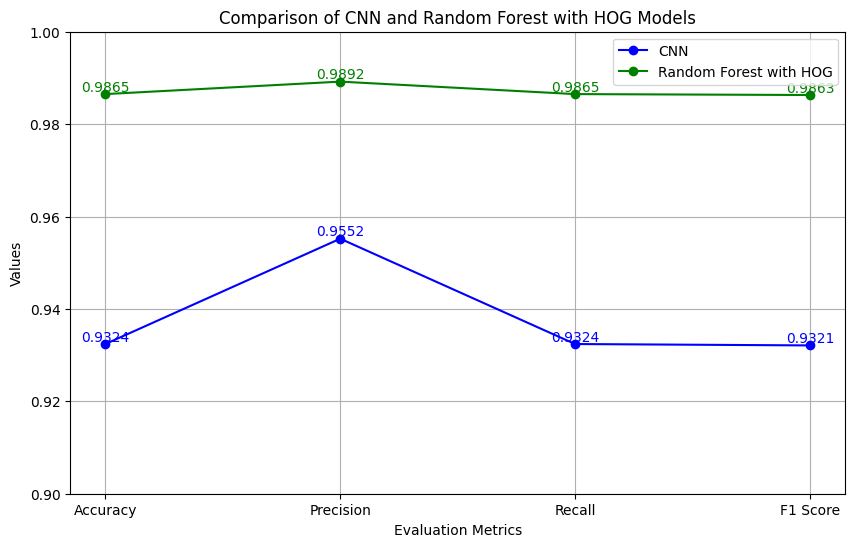


Fig. 4 Performance Analysis of the two models

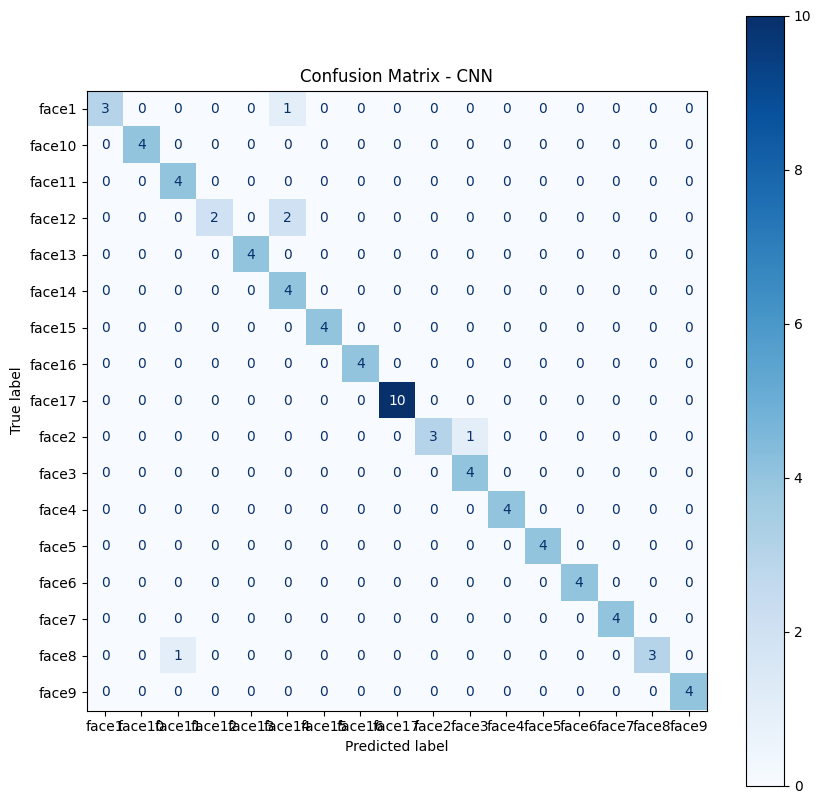


Fig 4(a). Confusion Matrix – CNN

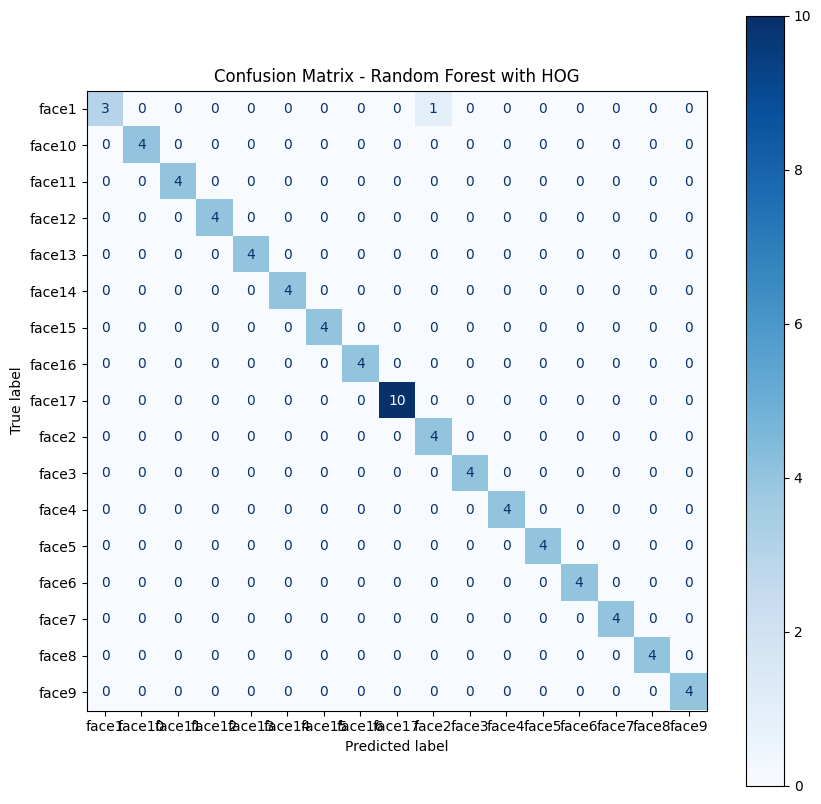


Fig 4(b). Confusion Matrix – Random Forest with HOG

*B. COMPARISON WITH OTHER MODELS*

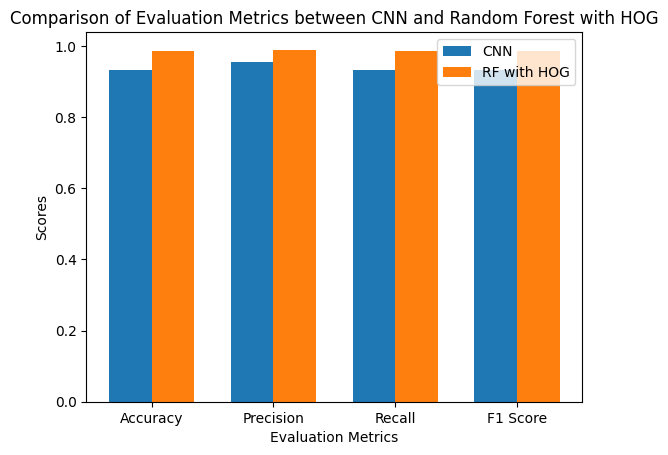
****

Fig 5. Model Performance Metrices Comparison

In comparing the performance of different models for image recognition, several key observations and insights can be drawn from the results of the CNN and Random Forest with HOG models, as well as the more variable results obtained from models like GoogLeNet, MobileNet, and VGG16.

The CNN results indicate that the CNN performed well, showing a good balance between precision and recall. The CNN's strong performance can be attributed to its ability to effectively capture spatial hierarchies in images through convolutional layers, which helps in distinguishing between different features and patterns. This feature extraction capability allows the CNN to achieve relatively high accuracy and consistency in classification tasks.

On the other hand, exceptional results from Random Forest Model highlight the effectiveness of combining Random Forest classifiers with HOG descriptors. HOG features are particularly useful for capturing object shapes and structures, while Random Forests, as an ensemble learning method, improve robustness by aggregating multiple decision trees' predictions. This combination enhances the model’s ability to generalize well across different images, contributing to the high-performance metrics observed.

In contrast, models such as GoogLeNet, MobileNet, and VGG16 exhibited unstable results. The variability in performance across these models could be due to several factors. For instance, these models may require more extensive fine-tuning and hyperparameter optimization to stabilize their performance. Additionally, their performance can be sensitive to the size and quality of the training data. GoogLeNet and VGG16, being deep architectures, might also suffer from issues related to overfitting or vanishing gradients if not properly managed. MobileNet, designed for efficiency on mobile devices, might trade off some accuracy for computational efficiency, which could contribute to the observed instability.

Overall, the high performance of the CNN and Random Forest with HOG models suggests that they are well-suited to the specific characteristics of the dataset used in this image recognition task. The CNN’s ability to learn hierarchical features and the Random Forest’s robustness with HOG features provide a strong foundation for stable and accurate image classification.

VI. CONCLUSION

In this project, we evaluated and compared the performance of various image recognition models, specifically focusing on Convolutional Neural Networks (CNN) and Random Forest with Histogram of Oriented Gradients (HOG) features. The CNN demonstrated solid performance with an accuracy of 93.24%, precision of 95.52%, recall of 93.24%, and an F1 score of 93.21%. This performance reflects the model's capability to effectively learn hierarchical features from images, contributing to its overall robustness in classification tasks.

In comparison, the Random Forest with HOG features achieved outstanding results, with an accuracy of 98.65%, precision of 98.92%, recall of 98.65%, and an F1 score of 98.63%. This high performance underscores the efficacy of combining HOG feature extraction with the Random Forest classifier. HOG's ability to capture detailed object shapes and structures, coupled with the ensemble learning approach of Random Forest, proved to be highly effective in enhancing classification accuracy and consistency.

However, models such as GoogLeNet, MobileNet, and VGG16 exhibited unstable results. This instability may be attributed to factors such as the need for more rigorous fine-tuning, sensitivity to training data size and quality, and potential overfitting or inefficiencies in deep architecture management.

Overall, the Random Forest with HOG features and CNN emerged as the most reliable models for this image recognition task. Their strong performance highlights their suitability for handling the dataset effectively. This project illustrates that while deep learning models like GoogLeNet, MobileNet, and VGG16 offer advanced capabilities, traditional machine learning approaches combined with robust feature extraction techniques can also deliver exceptional results.

VII. REFERENCES

[1] P. Chagas et al., "Evaluation of Convolutional Neural Network Architectures for Chart Image Classification," 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, 2018.

[2] S. Susan Jacob, "An Efficient Scheme for Facial Expression Recognition using Random Forest and CNN," *2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT)*, Kannur, India, 2022.

[3] T. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma, “A simple deep learning baseline for image classification?” IEEE transactions on image processing, vol. 24, no 12, pp. 5017-5032, September 2015.

[4] S. I. Yudita, T. Mantoro and M. A. Ayu, "Deep Face Recognition for Imperfect Human Face Images on Social Media using the CNN Method," 2021 4th International Conference of Computer and Informatics Engineering (IC2IE), Depok, Indonesia, 2021.

[5] S. S. Rajput and K. V. Arya, "CNN Classifier based Low-resolution Face Recognition Algorithm," *2020 International Conference on Emerging Frontiers in Electrical and Electronic Technologies (ICEFEET)*, Patna, India, 2020.

[6] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan, “Show and tell: A neural image caption generator,” IEEE transactions on pattern analysis and machine intelligence, vol. 39, no. 4, pp. 3156-3164, april 2017.

[7] G. C. Lee, Y. -C. Lee and C. -C. Chiang, "Low-Resolution Face Recognition in Multi-person Indoor Environments Using Convolutional Neural Networks," *2021 International Conference on Computational Science and Computational Intelligence (CSCI)*, Las Vegas, NV, USA, 2021

[8] E. Winarno, I. Husni Al Amin, H. Februariyanti, P. W. Adi, W. Hadikurniawati and M. T. Anwar, "Attendance System Based on Face Recognition System Using CNN-PCA Method and Real-time Camera," *2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, Yogyakarta, Indonesia, 2019.

[9] G. Guo and N. Zhang, "What Is the Challenge for Deep Learning in Unconstrained Face Recognition?" ,2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), Xi'an, China, 2018.

[10] K. Guo, S. Wu and Y. Xu, “Face recognition using both visible light image and near-infrared image and a deep network”. CAAI Transactions on Intelligence Technology, vol. 2, no.1, pp.39-47, 2017.

[11] R. Szmurlo and S. Osowski, "Deep CNN ensemble for recognition of face images," *2021 22nd International Conference on Computational Problems of Electrical Engineering (CPEE)*, Hrádek u Sušice, Czech Republic, 2021.

[12] K. He and J. Sun, “Convolutional neural networks at constrained time cost,” IEEE Int. conf. on computer vision and pattern recognition, pp. 5353-5360, 2015.

[13] H. Ben Frej, M. Ltaif, A. Ammar, C. Souani, “Parallel implementation of Sobel filter using CUDA”IEEE Int. conf. on Control, Automation and Diagnosis (ICCAD), pp. 209-212, 2017.

[14] V. Soni and V. P. Singh, "Copy-Move Forgery Detection Using HOG and LBP Features with Random Forest," *2023 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, Greater Noida, India, 2023

[15] R. Sharma, M.S.Patterh, “A new pose invariant face recognition system using PCA and ANFIS,” Optik-International Journal for Light and Electron Optics, vol. 126, n. 23, pp.3483-3487, 2015.

[16 A. Sugiharto *et al*., "Comparison of SVM, Random Forest and KNN Classification By Using HOG on Traffic Sign Detection," *2022 6th International Conference on Informatics and Computational Sciences (ICICoS)*, Semarang, Indonesia, 2022.

[17] J.Deng, J.Guo, N. Xue and S. Zafeiriou, “Arcface: Additive angular margin loss for deep face recognition,” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4690-4699, 2019.

[18] M. Yan, M. Zhao, Z. Xu, Q. Zhang, G. Wang, Z. Su, “Vargfacenet: An efficient variable group convolutional neural network for lightweight face recognition” In : Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019.