

DOI: 10.5281/zenodo.124261121

A DEEP LEARNING FRAMEWORK FOR ROBUST FACE RECOGNITION IN UNCONSTRAINED VIDEO UNDER POSE, OCCLUSION, AND ILLUMINATION VARIATIONS

Naisargi B. Shah^{1*}, Dr. Sanjay Agal², Mrs. Kajol Patel³

¹Department of Computer Science and Engineering, Parul Institute of Engineering & Technology, Waghodia, India. Email: naisargi.1710@gmail.com

²Department of Artificial Intelligence and Data Science, Faculty of Engineering and Technology, Parul University Waghodia, India. Email: sanjay.agal32685@paruluniversity.ac.in

³Department of Computer Engineering, Faculty of Engineering and Technology - Diploma Studies, Parul University Waghodia, India. Email: kajol.patel12633@paruluniversity.ac.in

Received: 11/11/2025

Accepted: 28/01/2026

Corresponding Author: Naisargi B. Shah
(naisargi.1710@gmail.com)

ABSTRACT

Face detection is the first step in most of recognition tasks, where the challenges lies in separating facial regions from complex backgrounds. Variations in age, skin tone, facial expressions, and conditions such as brightness or image quality make this task difficult. Additional challenges arise from partial occlusion and the disguises use. An effective face detection system must therefore recognize faces under diverse conditions and backgrounds. Generally, the process consists of two major steps: classification, where an image is examined to determine if it contains a face, and localization, where bounding boxes are drawn to mark the detected faces. Since detection narrows down the search space to facial regions, it significantly improves the performance and the efficiency of recognition systems. Although, the variability of human faces in terms of pose, orientation, accessories and resolution makes reliable detection a challenging research problem.

KEYWORDS: Face Detection, Biometrics, Authentication, Frames, Stage, Variation, Embedding.

1. INTRODUCTION

A face recognition system is a form of biometric technology that identifies or verifies individuals based on their facial features extracted from images or video streams. Rather than using traditional identification & recognition methods such as passwords or ID cards, facial recognition depends on unique physiological/biometric traits that are quite hard in forging or replicating.

Methods of biometric authentication/s are increasingly utilized across security sector, law enforcement agencies, and for consumer applications. While biometric fingerprints are being widely utilized for decades, other biometrics methods like iris scans, voice & module recognition, and gait analysis are also gaining traction nowadays. Amongst these, facial recognition has come out as one of the most practical and viable options because to its non-intrusive nature and due the rapid advancement of computer vision techniques.

Modern face recognition systems work by analyzing the detailed & distinctive attributes of the human like eye distance, nose shape, jawline contour and converting these measurements into a computer understandable numerical representation, also known as a face template. This template is later compared with the stored data and determine identity or confirm a match within data. Unlike raw images, face templates are designed so it can capture only discriminative attributes, further leading in differentiating one person from another all the time even under varying conditions.

In computer vision Face detection technology that identifies human faces as per dataset provided by the algorithm. In many systems face detection is main application for security purpose as well as for user identification for location attendance.

With the adaptation of deep learning and high-performance computing power, face recognition has cam out with significant improvements in accuracy and speed with better processing. These developments have led to widespread deployment in areas such as mobile authentication system, surveillance systems, and smart city initiatives.

2. PROCESS



Figure 1. Process

The procedure of face recognition in video streams/motion frames generally begins with

detecting faces attributes and then processing recognition on the identified area. This process contains two important stages/steps: detection and localization. During the first stage or detection, the system will identify whether face exist in frame. In the second/localization stage, the precise area of the detected face is marked using bounded boxes and also defined by coordinates such as (x, y, width, height).

After detection, the facial area is isolated from rest of the image, allowing computational power to be focused only to relevant areas rather than the entire region. This helps to improve recognition efficiency and recognition inside the frames. The difficulties in this process comes from factors such as pose variation, facial expressions, differences in skin tone, occlusions (like glasses or scarves), illumination changes, and differences in resolution across frames.

The extracted focused face is then encoded into a computer understandable numerical representation, also referred as face embedding or template, that can be compared against stored templates available in the database. This multi-step procedure ensures that even in unconstrained situations like live videos, recognition can be achieved with high accuracy and speed by systematically filtering and refining visual information.

3. FACE DETECTION METHOD

a classification for face detection methods.

- 1. Knowledge-Based:** The knowledge-based method, human crafted rules such as the specific geometrical arrangement of the eye, nose and mouth are used to identified faces. The main limitation of this technique lies in complexity of defining effective rules. if the established rules are excessively broad or overly rigid, system is prone to generating high number of false positive errors.
- 2. Feature-Based:** The feature-based method recognizes face based on identifying and analyzing facial regions. This approach involves multiple processing steps and it achieve 94% accuracy ,including image with multiple faces.
- 3. Template Matching:** Template Matching technique detects faces by predefined or parameterized with input image through correlation analysis. A human face can be separated into parts like eyes, face contour, nose, and mouth.
- 4. Appearance-Based:** The appearance-based method construct face models from selected set of training face image. Compare to other approaches it more effective. Using statistical analysis and machine learning to find the relevant

characteristics of face images. This method commonly used for feature extraction.

4. DEEP LEARNING FOR FACE DETECTION

Deep learning is a subfield of machine learning system which relies on ANN artificial neural networks to model and solve complex problems in the process. These networks are developed to mimic like the functioning of the human brain, where layers of interconnected neurons process information in a hierarchical manner. Rather than traditional machine learning methods, deep learning method does not require manual feature engineering task, instead, it will automatically learn patterns from large volumes of data.

The recent increase in interests toward deep learning method is basically due to the availability of vast datasets and advancements in computational power, particularly introducing the use of GPUs. These resources have provided deep learning models that can outperform conventional methods in domain including image classification, speech recognition, and natural language processing.

While talking about face detection and recognition, deep learning method enables computers to distinguish & associate subtle variations in facial structure and expression by analyzing raw pictures pixels. CNN Convolutional Neural Networks, in particular, have come out as the backbone of modern recognition systems. CNNs learn hierarchical features by starting from edges and textures at lower layers of network, to complex and abstract features such as facial components at deeper layers. This feature extraction makes CNNs quite effective at distinguishing between individual faces, also under challenging conditions such as illumination changes, occlusions, or changes in pose.

The flexibility and scalability of deep learning make it an crucial technique for video based face recognition, as it ensures that recognition systems continuously improve over time with the availability of new training data.

5. COMPARISON OF METHODS / ALGORITHMS

A variety of algorithms techniques have been developed over the years to improve the performance of facial identification systems. Each methodology approach a unique balance of strengths and limitations:

1. Convolutional Neural Networks (CNNs): Accomplish high level of recognition accuracy by automatically learning discriminative facial

features. Perform well in both detection and recognition tasks, though accuracy may decrease in crowded scenes.

2. Support Vector Machines (SVMs): This approach combined with CNN-based features to improve disguised or partially occluded faces effectively. Overall Performance is weaker than end-to-end deep learning models.

3. Viola-Jones Algorithm: An early real-time detector using Haar-like features. It is fast but highly sensitive to lighting and pose variations.

4. Probabilistic Linear Discriminant Analysis (PLDA): Useful for verification with good detection rates, but performance drops under environmental changes.

Deep Convolutional Neural Networks (DCNNs): By Reducing false positives and improve recognition accuracy, making them ideal for surveillance. However, they may misclassify person with highly facial features.

Compared to traditional methods, CNN and DCNN-based Methods provide high accuracy and robustness, prefer them as the standard for modern systems.

6. EXISTING VS PROPOSED SYSTEM

Face recognition systems have progressed from traditional methods to advanced deep learning approaches. The key differences include:

1. Parameter Focus: Traditional systems mainly focus on illumination effects. however proposed systems handle multiple factors such as illumination, occlusion, and pose variations.

2. Input Data: Both methodology utilize continuous video streams. Proposed system applies more effective frame.

3. Preprocessing: frame are generate mainly for face detection. proposed methods use Python tools to extraction and improve detection accuracy.

4. Face Detection: Traditional approaches use Viola-Jones, while the proposed system applies Haar Cascades for improved detection.

5. Feature Extraction: Older systems rely on BRISK, while the proposed systems leverages CNNs for stronger features representation.

6. Classification: CNNs are used in both but larger and more diverse datasets improve accuracy in the proposed system.

7. Accuracy: Existing methods achieve moderate results while proposed systems consistently deliver higher accuracy.

7. IMPLEMENTATION

1. RCNN: Uses selective search to generate 2000

region proposals, which are classified using CNNs and SVMs. Accurate but computationally expensive.

2. **Fast RCNN** :Improves efficiency by running CNNs once per image to create a feature map, reducing redundancy. Faster but still reliant on selective search.
3. **Faster RCNN** : Eliminates selective search by introducing a Region Proposal Network, allowing near real-time performance with higher accuracy.
4. **YOLO (You Only Look Once)**: Treats detection as a regression problem, dividing images into grids and predicting bounding boxes and class probabilities directly. Extremely fast but less effective with small or clustered objects.

8. WORKFLOW

The general workflow for video-based face recognition involves:

1. **Video Input**: Capturing a live feed or recorded video.
2. **Frame Extraction**: Breaking the video into individual frames.
3. **Preprocessing**: Standardizing frames by converting to grayscale, resizing, and noise reduction.
4. **Face Detection**: Using algorithms like Haar Cascades or CNNs to locate faces.
5. **Feature Extraction**: Transforming facial details into numerical embeddings.
6. **Classification**: Matching embeddings against stored templates to confirm identity.
7. **Output**: Displaying recognized faces with bounding boxes and labels.

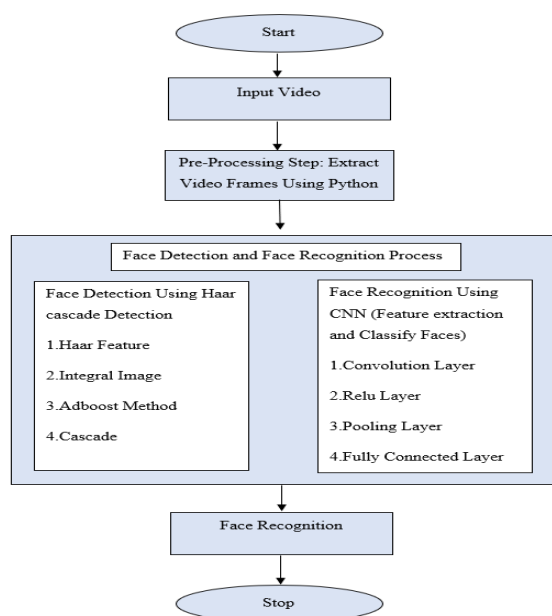


Figure 2. Work Flow

9. ISSUE AND CHALLENGE

The Facial Recognition System has come a long way. Its usage is crucial in quite a few applications, for instance - photo retrieval, surveillance, authentication/access control systems etc. but, it is yet to completely overcome the challenges which have constantly played with its quality of delivery. Listed below are the challenges which limit the potential of a Facial Recognition System to go that extra mile.

1. **Illumination** :Slight change in lighting condition can major impact on its result. when environmental illumination fluctuates, capturing the exact same subject using same sensor, pose and facial expression can still produce drastically different results.
2. **Background** : The placement of the subject also serves as a significant contributor to the limitations. A facial recognition system accuracy drops when moving from indoor to outdoors environments due to sudden shifts in lighting and background conditions. Natural human variations such as aging and changing facial expressions, specifically decrease the system's overall reliability.
3. **Pose** : Facial Recognition models are highly susceptible to pose variations. The changes in head or camera angle significantly alter face appearance and generate inter class variations that make automated face recognition across pose a tough nut to crack.
4. **Occlusion** : The Facial hair such as breads and moustaches along with accessories like mask, hats, eyewear, significantly interferes with the accuracy of facial recognition systems. These occlusions increase the visual variable of a subject making it highly challenging for the system to perform reliably in real world unconstrained environments.
5. **Expressions** : Another critical factor is the variability of a subject's face driven by underlying emotional changes. The frequent occurrence of both macro and micro expressions alters facial geometry, thereby hindering the efficiency of recognition models.
6. **Complexity** : Deep learning models often incur high computational costs due to their extensive architectural depth. Consequently deploying these complex system in real time applications or resource constrained embedded device presents a significant technical challenge. To overcome these limitations an optimal recognition framework must balance computational efficiency with robust performance ensuring accurate detection

across varying poses, lighting conditions and facial occlusions.

Motivation

Biometric technology makes data system safer by using people's unique behaviour and physical attributes to identify them. these attributes can be Physgical like face, fingerprints, iris pattens, voice or behavioral like signature. To validate, most system need user to do something like scan fingerprint or take a picture of face. A video is a sequence of frames. A video can be captured

by a person who is professional videographer or by some nonprofessional person. When the video is recorded by some nonprofessional there are many chances that the video can be categorized as unconstrained video that can have many problems related to quality of video due to motion of camera while recording, poor or low light ambience, background variation, occlusions, etc. Due to above all reasons, it makes an interest of study and to do research on all such videos and to recognize face from that.

10. PRE-PROCESSING

The video is uploaded, then sequence of frames is generated from the video. These frames are off different illuminance conditions, hence to detect the face with this background is difficult. So, the RGB frames generated are converted into gray scale images for better detection purposes.

face detection : Viola & Jones

The Haar Cascade algorithm is a widely adopted machine learning technique for localizing objects within images and video streams [34]. Formulated by Viola and Jones, the method relies on training a predictive cascade function using a vast corpus of positive and negative image samples. Its operational framework systematically executes four distinct steps: selecting Haar features, computing integral images, executing AdaBoost training, and deploying a cascade of classifiers.

The algorithm has four sequential phases:

1. Haar Feature Selection
2. Creating Integral Images
3. Adaboost Training
4. Cascading Classifiers

1. Haar Feature Selection

The initial step to collect the Haar Features. This process examines adjacent rectangular blocks inside a detection frame, calculates the total pixel intensity for each block, and determines the final feature value by subtracting these sums.

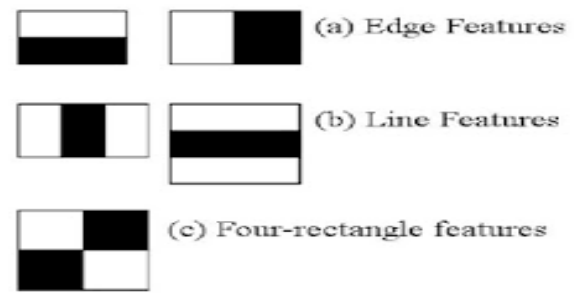


Figure 3. Haar Code Processing [17]

Figure types of Feature Selection



Figure 4. Face Read [17]

2. Integral Image

Integral Images are used to speed up processing. even though this is efficient, many of the features were calculated does not help with classification. the first feature based on brightness difference, where eye region has lower intensity values than nose and cheek region. The second features is based on contrast between the darker eye region and brighter bridge of the nose. but when the windows position on different parts of face like cheeks, they become less meaningful for detection.

11. CONCLUSION

Deep learning has fundamentally revolutionized video-based face recognition, effectively overcoming many limitations of traditional methods. Haar Cascades method provides an effective means of face detection, while CNNs helps to enhance recognition by learning complex features. This hybrid approach offers robust performance across varying lighting conditions and poses.

The analysis confirm that deep learning-based models elevate both recognition accuracy and adaptability. However, challenges remain in reducing computational costs and handling extreme occlusions. Future research should focus on expanding training data, optimizing models for embedded devices, and combining multimodal biometrics to improve overall reliability.

Table 1. Result

Method	Object Class	Dataset	Accuracy
CNN	Person Face (User)	Self-Created	82%
Haar Cascade	Person Face (User)	Self-Created	84%

12. FUTURE SCOPE

In future Face Reorganization Technology

implemented using real-time solution. For Security Purpose large Store or Business hub use contact less system facial recognition system. In Smart cities, facial recognition software attached with IOT for manage Smart Home, School-Organization in-out time(attendance) and apply widely in security. In health care Facial Recognition used for patient emotion and monitoring of patient health condition.

REFERENCES

1. S. Meenakshi, M. Siva Jothi, and D. Murugan, "Face recognition using deep neural network across variations in pose and illumination," *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 8, no. 154, pp. 289–292, Jun. 2019.
2. K. G. S. Keerthi and C. R. Usha, "Human face recognition in video using convolutional neural network (CNN)," *International Research Journal of Engineering and Technology (IRJET)*, vol. 6, no. 5, pp. 4726–4732, May 2019.
3. A. V. Aswathy and S. L. Das, "Video based face recognition and tagging using deep learning," *International Journal of Engineering Research & Technology (IJERT)*, vol. 8, no. 7, pp. 78–81, Jul. 2019, doi: 10.17577/IJERTV8IS070043.
4. A. G. Abhila and S. H. Sreeletha, "A deep learning method for identifying disguised faces using AlexNet and multiclass SVM," *International Research Journal of Engineering and Technology (IRJET)*, vol. 5, no. 7, pp. 1239–1244, Jul. 2018.
5. N. Zhuang, Q. Zhang, C. Pan, B. Ni, Y. Xu, X. Yang, and W. Zhang, "Recognition oriented facial image quality assessment via deep convolutional neural network," *Neurocomputing*, vol. 358, pp. 109–118, Sep. 2019, doi: 10.1016/j.neucom.2019.04.057.
6. U. Zafar, M. Ghafoor, T. Zia, G. Ahmed, A. Latif, K. R. Malik, and A. M. Sharif, "Face recognition with Bayesian convolutional networks for robust surveillance systems," *EURASIP Journal on Image and Video Processing*, vol. 2019, no. 10, pp. 1–15, Jan. 2019, doi: 10.1186/s13640-019-0406-y.
7. M. Coşkun, A. Uçar, Ö. Yildirim, and Y. Demir, "Face recognition based on convolutional neural network," in *Proc. 2017 International Conference on Modern Electrical and Energy Systems (MEES)*, Kremenchuk, Ukraine, Nov. 2017, pp. 376–379, doi: 10.1109/MEES.2017.8248937.
8. C. T. Ferraz and J. H. Saito, "A comprehensive analysis of local binary convolutional neural network for fast face recognition in surveillance video," in *Proc. 24th Brazilian Symposium on Multimedia and the Web (WebMedia)*, Salvador, Brazil, Oct. 2018, pp. 265–268, doi: 10.1145/3243082.3267444.
9. R. Tapu, B. Mocanu, and T. Zaharia, "Face recognition in video streams for mobile assistive devices dedicated to visually impaired," in *Proc. 2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, Las Palmas de Gran Canaria, Spain, Nov. 2018, pp. 137–142, doi: 10.1109/SITIS.2018.00030.
10. K. Sai Krishna, G. Sreenivasa Raju, and P. Praveen Kumar, "A deep neural network for face recognition," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 8, no. 12S, pp. 420–423, Oct. 2019, doi: 10.35940/ijitee.L1105.10812S19.
11. B. Bhavana, V. Jagan Naveen, and K. Krishna Kishore, "Comprehensive analysis of machine learning algorithms for face detection," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 8, no. 10, pp. 2479–2482, Aug. 2019, doi: 10.35940/ijitee.J9542.0881019.
12. T. T. N. Thi and K. N. Trong, "An efficient face detection and recognition," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 7, no. 5, pp. 35–40, Feb. 2018.
13. Y. Zhou, H. Ni, F. Ren, and X. Kang, "Face and gender recognition system based on convolutional neural networks," in *Proc. 2019 IEEE International Conference on Mechatronics and Automation (ICMA)*, Tianjin, China, Aug. 2019, pp. 1091–1095, doi: 10.1109/ICMA.2019.8816192.
14. M. Wang, Z. Wang, and J. Li, "Deep convolutional neural network applies to face recognition in small and medium databases," in *Proc. 4th Int. Conf. on Systems and Informatics (ICSAI)*, Hangzhou, China, Nov. 2017, pp. 1368–1372, doi: 10.1109/ICSAI.2017.8248499.
15. X. Xie and K. M. Lam, "An efficient illumination normalization method for face recognition," *Pattern Recognition*, Elsevier, vol. 42, no. 12, pp. 3222–3232, 2009.

16. Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu, "Facial expression recognition using deep neural networks," *PLOS ONE*, vol. 9, no. 9, e99212, 2014.
17. J. Willberger, "Haar cascade explained," [Online]. Available: <http://www.willberger.org/cascade-haar-explained/>
18. S. Chatterjee, M. Choudhary, R. D. Sarvaiya and M. D. F. Abdulla, "Machine learning techniques for identifying DDOS attacks in cloud computing background," *Parul University International Conference on Engineering and Technology 2025 (PiCET 2025)*, Hybrid Conference, Vadodara, India, 2025, pp. 1564–1571, doi: 10.1049/icp.2025.1668.
19. Chatterjee, S., Patil, R.V., Choudhary, M., Sarvaiya, R.D. (2026). Optimizing DDoS Detection Using Machine Learning Approach. In: Saraswat, M., Rajan, A., Chakravorty, A. (eds) *Congress on Smart Computing Technologies. CSCT 2024. Smart Innovation, Systems and Technologies*, vol 122. Springer, Singapore. https://doi.org/10.1007/978-981-96-6250-0_45
20. Choudhary, M., Bhawsar, J., Sarvaiya, R., Patil, R., & Shah, V. (2025, April 24). Intelligent shopping cart systems enhance the checkout experience and enable realtime inventory tracking in retail environments. *Journal of Information Systems Engineering and Management*, 10(39s). <https://doi.org/10.52783/jisem.v10i39s.7287>
21. Choudhary, M., Solanki, P. S., Gamit, V., & Joshi, M. (2024). Machine learning classifier used to diagnosis of liver disorders. In *2024 Parul International Conference on Engineering and Technology (PICET)* (pp. 1–6). IEEE. <https://doi.org/10.1109/PICET60765.2024.10716100>
22. Prabakaran, P., Choudhary, M., Kumar, K., Loganathan, G. B., Salih, I. H., Kumari, K., & Karthick, L. (2024). Integrating Mechanical Systems With Biological Inspiration: Implementing Sensory Gating in Artificial Vision. In S. Padhi (Ed.), *Trends and Applications in Mechanical Engineering, Composite Materials and Smart Manufacturing* (pp. 193–206). IGI Global Scientific Publishing. <https://doi.org/10.4018/979-8-3693-1966-6.ch012>
23. Sudhagar, D., Saturi, S., Choudhary, M., Senthilkumaran, P., Howard, E., Yalawar, M. S., & Vidhya, R. G. (2024). Revolutionizing data transmission efficiency in IoT-enabled smart cities: A novel optimization-centric approach. *International Research Journal of Multidisciplinary Scope (IRJMS)*, 5(4), 592–602. <https://doi.org/10.47857/irjms.2024.v05i04.01113>
24. P. S. Solanki, Y. B. Adhyaru and M. Choudhary, "EHSM - Heartbeat Sensors & Machine Learning for Horse Health Monitoring," *2024 Parul International Conference on Engineering and Technology (PICET)*, Vadodara, India, 2024, pp. 1–6, doi: 10.1109/PICET60765.2024.10716092.
25. Offori, K., & Choudhary, M. (2024). Computer vision with DL/ML for crime data analysis. *Educational Administration: Theory and Practice*, 30(1), 4526–4532. <https://doi.org/10.53555/kuvey.v30i1.8226>
26. Choudhary, M., Gamit, V., Swaminarayan, P. R., Parmar, H., & Mehta, A. (2023). Identifying and enhancing concurrent coverage in wireless sensor networks through steering mechanism. *Shodhasamhita: Journal of Fundamental & Comparative Research*, 9(1, Book No. 5), 17–33. ISSN: 2277-7067.
27. Choudhary, M., Shah, N. K., Mehta, A. R., & Parmar, S. (2023). Emerging technologies in IoT security: Addressing current challenges and future directions. *Shodhasamhita: Journal of Fundamental & Comparative Research*, 9(1, Book No. 5), 34–41. ISSN: 2277-7067.
28. Gor, K., Patel, V., Dave, V., & Choudhary, M. (2023). Dynamic resource allocation strategies for efficient IoT applications in fog computing. *Shodhasamhita: Journal of Fundamental & Comparative Research*, 9(1, Book No. 6), 123–131. ISSN: 2277-7067.
29. Choudhary, M., & Shrimali, M. (2019). Intrusion detection on big data using machine learning: A review. *IJRAR - International Journal of Research and Analytical Reviews*, 6(2), 263–271. ISSN: e-2348-1269, p-2349-5138.
30. Choudhary, M., & Shrimali, M. (2019). A novel intrusion detection technique using boosting approach on big data. *Journal of Emerging Technologies and Innovative Research (JETIR)*, 6(6), 935–943. ISSN: 2349-5162.
31. Boahen, J. O., & Choudhary, M. (2024). Advancements in precision agriculture: Integrating computer vision for intelligent soil and crop monitoring in the era of artificial intelligence. *International Journal of Scientific Research in Engineering and Management (IJSREM)*, 8(3), 1–6. <https://doi.org/10.55041/IJSREM29725>