Sigma J Eng Nat Sci, Vol. 44, No. 2, pp. 1526−1539, April, 2026

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2024.2048

Review Article

# Systematic review and analysis on recent techniques of photo sketch matching

Vishakha SEHDEV VERMA[1,*], Nidhi MALIK[1], Anshul BHATIA[2]

[1]*Department of Computer Science and Engineering, The NorthCap University, Gurugram, Haryana, 122017, India*
[2]*University School of Automation and Robotics, Guru Gobind Singh Indraprastha University, 110075, India*

## ARTICLE INFO

*Article history*
Received: 03 June 2025
Revised: 30 July 2025
Accepted: 10 October 2025

**Keywords:**
Deep Learning; Feature Extraction; Matching Algorithm; Photo Comparison; Sketch to Photo

## ABSTRACT

The technique of aligning photographs with an artist's drawing is referred to as photo sketch matching. Matching the exact photo from the database is a big challenge, and the best technique must be selected. The focus of the study is to do an analysis of the techniques used in photo-sketch matching. The different techniques analysed in the study include Gabor Shape, Generative Adversarial Networks (GAN), Smart Switching Slime Mold Method, and Lightweight Vision Transformer. The review indicates that no single technique universally outperforms others across all conditions; rather, the choice of method depends heavily on the application context (e.g., forensic sketches, artist-drawn sketches, or automatically generated sketches), suggesting that combining the strengths of classical and deep learning methods, solving domain adaptation, and ensuring scalability for real-world deployment. The novelty of this review lies in the comprehensive synthesis and reorganization of existing research on photo–sketch matching. The work goes beyond by providing a critical comparison of techniques, identifying their strengths, limitations, and applicability, such as in forensics. The results show that in order to prevent overfitting, a large dataset is required for training. The intricacy of the issue requires an extensive set of forensic drawings. The accuracy is affected by multiple factors and can be improved by using hybrid techniques or deep learning approaches.

**Cite this article as:** Sehdev Verma V, Malik N, Bhatia A. Systematic review and analysis on recent techniques of photo sketch matching. Sigma J Eng Nat Sci 2026;44(2):1526−1539.

## INTRODUCTION

Automation is important in modern, rapidly evolving, and technologically advanced societies. As human labor demands become increasingly intense, machines need to be trained to collaborate with people [1]. This would result in enhanced capabilities, quicker tasks, and greater productivity. Image processing is a key method for enhancing or refining the image [2]. Tools for machine learning have been created to reduce the level of effort needed to handle images. The automation of creating, synthesizing, and recognizing face sketches from photographs has been a prominent field of research in computer vision and image processing. In this survey, 178 publications were reviewed; 49 were chosen for deeper analysis, and research questions

***Corresponding author.**
*E-mail address: vs4821@gmail.com

*This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic*

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

(RQs) were developed [3]. Furthermore, it details the criteria used for selection, which encompasses both inclusion and exclusion. The manuscript is prepared in sections. The primary subject of Section 2 is picture sketch matching. Section 3 outlines the methodology. Section 4 addresses the implementation and detailing of the various studies identified. Section 5 focuses on the reporting of the review, where all the research questions are answered, and in Section 6, the paper is concluded.

**Photo Sketch Matching**

The matching of photo sketches is a crucial part of criminal investigations. In video surveillance, face recognition helps identify a person for criminal investigations. A drawing is created with the assistance of a forensic artist after the witness provides a facial description of the offender [4]. Once the sketch is finished, it is compared to the agency's database. It is a difficult process because face matching data is collected from multiple sources and under different conditions. Verifying that the drawing and photo match is the aim of facial sketch recognition. Conventional uniform face recognition methods struggle with identifying faces from sketches because of the differences in modality between photographic images and sketch representations. This issue has been addressed using soft computing methods [5]. Previously, creating a face sketch was a tedious task where accuracy depended on the skill level of the artist.

Fortunately, with advancements in technology, face sketch recognition systems can now operate automatically [6]. Identifying a face through sketches begins with creating a hand-drawn sketch, which is then matched against images stored The image that best matches the sketch in the identification database can be used to determine the individual's identity [7]. More advanced preprocessing techniques cannot be applied to forensic sketches, as witnesses at the crime scene usually recall the outer features of the face rather than the inner details. Traits such as gender, age, and color are examples of superficial characteristics. Notable internal features are seldom documented [8]. The disparity in visual cues is what causes the modality gap. Facial features are extracted with the help of gabor filter. Features that are taken from coarse texture (facial shape) can greatly reduce the modality gap. Gabor filters attenuate fine texture first in the Gabor Shape framework. The Radon transform is then used to model the facial shape, which is represented by coarse texture. The experimental results using Gabor filter found to be superior thant the state-of-art approaches.

GAN is a very popular deep learning technique. A GAN learns from an existing dataset to produce new, realistic-looking data (such text, audio, or photos). For investigative purposes, Kokila R. offers a study on matching sketches to pictures [9]. The system's strong application focus contributed to its low level of complexity. Several sketches were compared to images of the face taken from different perspectives, yielding incredibly accurate results [10]. Because the technique was entirely reliant on the quality of the sketches utilized, low-quality sketches produced unsatisfactory results For the generative challenge, Sumit Gunjate suggested a network structure, and testing revealed that it outperformed existing arrangements. To improve the SMA's performance in particular domains, researchers suggested a number of modifications. The approch used the S2SMA to improve the accuacy [11].

A lightweight deep learning model called a Light Semantic Transformer was created to capture semantic (meaning-based) links between data pieces. It was inspired by the Transformer architecture. Another face sketch identification technique based on transfer learning was presented by Wan et al. They used the triple loss to learn discriminative features, created a three-channel CNN structure, and suggested a hard triple sample selection technique to speed up model convergence. To finish the target classification, ViT immediately stacked Transformer encoders that operated on picture patch sequences. On multiple image recognition benchmarks, ViT outperformed CNN networks [12]. Lin Cao introduced the LWVT-ResNet18 framework for the light semantic transformer network. The modules are trained from scratch which address the overfitting problem. The mdoel can directly the semantic features from photo and sketch which further improves the accuracy. FG-SBIR is a technique aims at retrieving a single picture from a gallery tailored to a specific category that corresponds with a query sketch. Descriptive tags, hierarchical co-attention mechanisms, self-supervised pre-training techniques, cross-category generalization methods, attention-based components featuring an advanced retrieval loss, and reinforcement learning approaches have all contributed to the gradual improvement of FG-SBIR. In general, they seek to learn a joint embedding space to minimize the cross-modal gap, usually by employing a triplet ranking loss, which seeks to distance a drawing from other images while bringing it closer to its matched photo [13]. CLIP allows for generalizability in downstream tasks without labels by representing text and images in the same embedding space. According to recent studies, CLIP's open-vocabulary generalization capability through an easy-to-use prompt-based design can improve SBIR's efficiency. The techniques are shown in Figure 1.

Diffusion models are emerging as a promising approach in photo–sketch matching due to their ability to generate high-quality and detailed images through iterative denoising. They provide stable optimization and capture fine-grained structural features essential for preserving facial identity across modalities. Recent advancements in diffusion-based methods have significantly enhanced photo–sketch matching, particularly in 2024 and 2025. These models have demonstrated improved fidelity and realism in generating face sketches from photos, addressing challenges such as the modality gap and data scarcity [14]. In 2024, jain et al. Introduced clip4sketch, a controllable latent diffusion model that utilizes text and image embeddings to generate diverse and realistic sketches. This approach aids

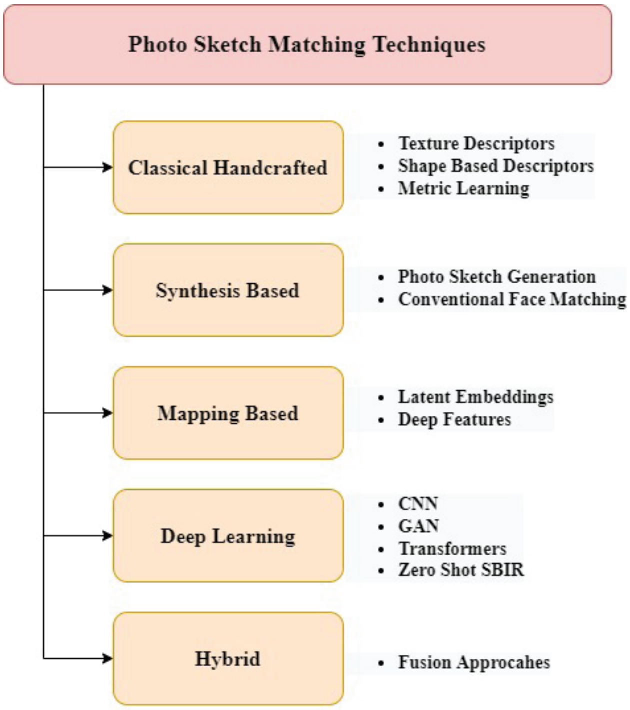


**Figure 1.** Techniques.

in augmenting datasets for sketch-to-mugshot matching, thereby enhancing the performance of face recognition systems. Additionally, a novel architecture for face sketch-to-photo synthesis was proposed, employing denoising diffusion probabilistic models (ddpms). The method simplifies the transformation process into sequential denoising steps, incorporating a pretrained coarse generator to encode sketch information, and a detail diffusion branch to refine the generated images. In 2025, wu et al. Developed one-shot face sketch synthesis in the wild, a method that employs generative diffusion priors and instruction tuning to convert face photos into realistic sketches using a single example. This approach addresses data scarcity by leveraging a new benchmark dataset, os-sketch, which includes diverse sketches and photos. These developments underscore the growing potential of diffusion models in bridging the modality gap in photo–sketch matching, offering more robust and realistic solutions for face recognition tasks.

The paper acknowledges the importance of ethical, legal, and practical limitations in forensic applications of photo–sketch matching. Ethical concerns arise from demographic and dataset biases, which can lead to unequal identification outcomes or misidentification of certain groups, raising issues of fairness and accountability. Legal considerations involve privacy, consent, and the admissibility of synthesized or generated images in judicial proceedings, which may vary across jurisdictions. Practical challenges include variability in sketch quality, differences in artist skill, inconsistencies between sketches and actual photographs, and environmental or contextual factors such as lighting and pose. Additionally, limitations in dataset diversity and the lack of standardized evaluation protocols can hinder the reliability and generalization of matching systems. These combined factors highlight the need for careful consideration when deploying photo–sketch matching technologies in real-world forensic scenarios.

## MATERIALS AND METHODS

This section describes the methods that were utilized. Figure 2 illustrates the review methodology applied in this study. In the first phase, following the identification of search databases and preliminary studies, research questions were formulated. In the second step, articles were selected after organizing redundant and unnecessary research [15]. At

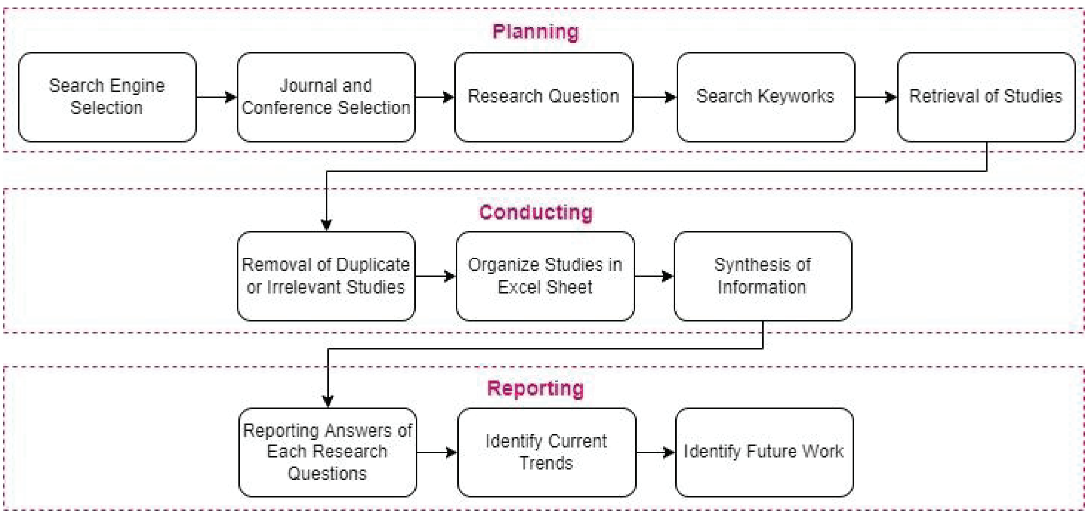


**Figure 2.** Review methodology.

this stage, synthesis was also conducted using the data collected from each of the shortlisted studies. In the final phase, responses to all of the research questions, as well as current trends and challenges in each topic, were provided.

The process of planning includes the following steps:

**Selection of Search Database**

The search process comprises two stages. In the initial phase, keywords like "Photo sketch matching" were entered to generate results. During this phase, resources such as Google Scholar, Scopus, Science Direct, Springer, and IEEE Explore were identified. The second stage involved locating articles in relevant journals and conferences.

**Identification of Important Journals and Conferences**

Prior to their use in research, the quality of journals is evaluated based on several criteria [16]. These criteria encompass journal citations and citation counts. Table 1 provides a summary of the study topics related to photo-sketch matching.

**Formulation of Research Questions**

After completion of the initial phase, the potential directions have been formulated for the subsequent phase. Properly defining study objectives is essential, as it will help to stay focused and work through the data. The goal of this work is to examine the existing methods and tools used for photo sketch matching [17]. From an initial pool of 178 articles, only those that specifically addressed are selected. This chapter's research aims to answer the identified research questions.

**Search of Keywords**

The aim of the review is to gather all relevant papers and ultimately present a comprehensive overview. During the initial search, various keywords like face sketch recognition, image sketch matching, and sketch-to-photo matching were examined [18]. After an in-depth analysis of the material, new keywords were incorporated to enrich the study.

**Retrieval of Studies**

The pertinent digital libraries were utilized to find all significant research. This method involved examining the reference sections of each study to identify additional relevant papers or publications, which were subsequently gathered and arranged for analysis [19].

## ANALYSIS

The second phase of the analysis commences after the completion of the first stage. At this point, all research articles undergo a detailed review [20]. Research questions (RQs) form the foundation for specific criteria, which are then applied to each paper according to the standards outlined previously [21]. These articles have been utilized to assess the questions highlighted in the earlier phase. The limitations of each paper are identified, as this will help steer the research in the right direction.

**Selection Criteria**

The required documents are manually selected. Initially, the total number of papers that were carefully reviewed is reduced when the specified section of a document is evaluated [22]. In the second criterion, the entire group of papers selected under the first criterion is analyzed [23]. Two separate sets of selection criteria have been applied.

Inclusion Criteria

(i) Papers of reputed journals like Springer, IEEE, etc.
(ii) Papers that are focused on the topic of photo sketch matching.
(iii) Papers with recent publication.
(iv) Articles from SCOPUS and SCI publications.

Exclusion Criteria

(i) Articles of low-quality publication.
(ii) Articles only containing an abstract but not with full paper.
(iii) Articles without a dataset.
(iv) Unpublished articles without source information.

Based on the identified keywords, a total of 178 studies were initially gathered through various search engines. These studies were then analyzed according to specific inclusion and exclusion criteria. A total of 49 studies were selected, while those exhibiting low publication quality, lacking complete presentations, failing to cite the dataset, and those not belonging to either journal or conference categories were eliminated. In the subsequent stage, the remaining research was assessed against the quality evaluation guidelines to filter out irrelevant studies from the extensive pool of available data [24]. The metrics used for quality evaluation are presented in Table 2.

**Table 1.** List of research questions

| S.No. | Research Questions |
|---|---|
| RQ1 | What facial sketch mapping methods are commonly utilized? |
| RQ2 | What kinds of datasets have been utilized? |
| RQ3 | What challenges have image sketch recognition methods faced, and what future possibilities do these approaches hold? |
| RQ4 | Which different performance metrics are used in the study? |
| RQ5 | Which pre-processing methods have been employed thus far? |

**Table 2.** List of quality assessment measures

| S.No. | Quality assessment measures |
|---|---|
| QAM 1 | Research objectives must be mentioned in detail. |
| QAM 2 | It is important to identify the dataset that was utilized to implement the methods. |
| QAM 3 | It is essential to evaluate various methods to determine their effectiveness. |
| QAM 4 | A proper literature survey must be done. |
| QAM 5 | Future scope must be present for applied techniques. |

**Data Extraction**

A comprehensive database has been established that includes various fields such as the author's name, the article's title, the source of publication, the tools and techniques used, the dataset involved, and a concise summary of the paper's content [25]. To address all the research questions, an in-depth analysis of these articles has been performed. The chosen publications have been examined from multiple angles, assessing their methodologies and evaluating any connections among them [26]. The use of different visualization methods has been beneficial in analyzing technologies, tools, processes, and performance assessments. Multiple forms of visualization, such as bar graphs, pie charts, and column charts, are utilized for the analysis.

**Distribution of Papers**

The articles are organized in sections as per publication year and the source [27]. The papers can be divided into three categories according to their sources: journal articles, conference papers, and other formats such as technical reports, book chapters, or various study materials.

**Year of Publication**

A novel approach to identifying criminals has come to light: comparing photo sketches. The techniques previously used aimed to minimize the differences in modalities, while later research focused on the direct comparison of photographs and illustrations. Diverse approaches have been applied across different datasets. Figure 3 illustrates the breakdown of publications by source, Figure 4 shows the yearly distribution of studies, and Figure 5 outlines the process of searching for and selecting documents.

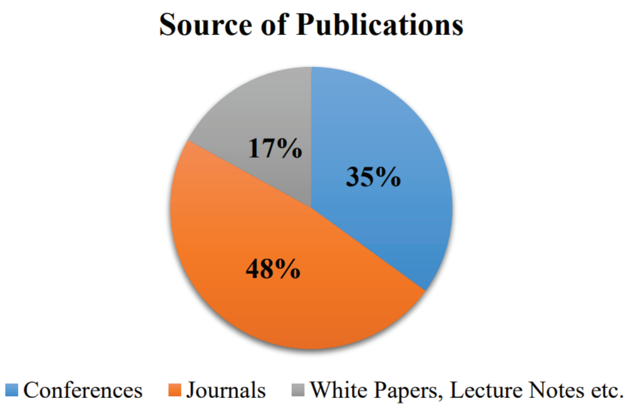


**Figure 3.** Publication sources.

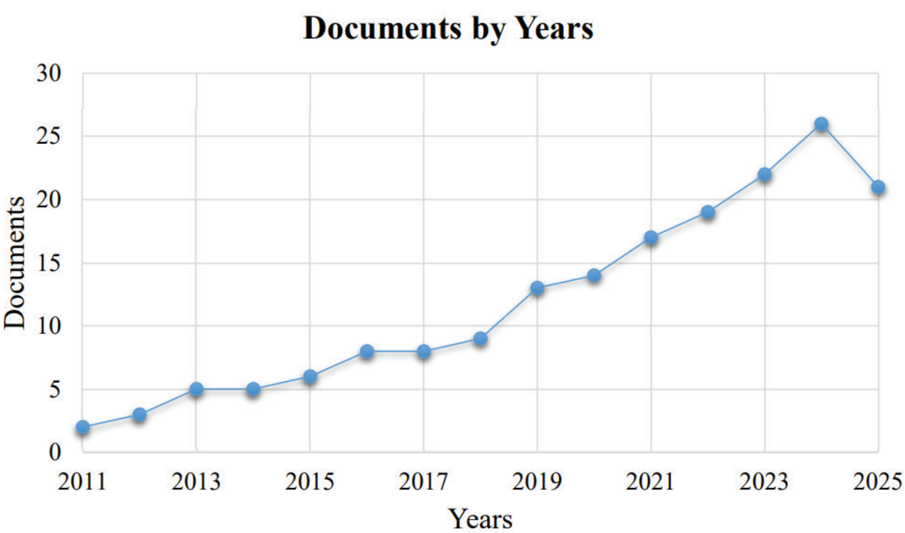


**Figure 4.** Documents by year.

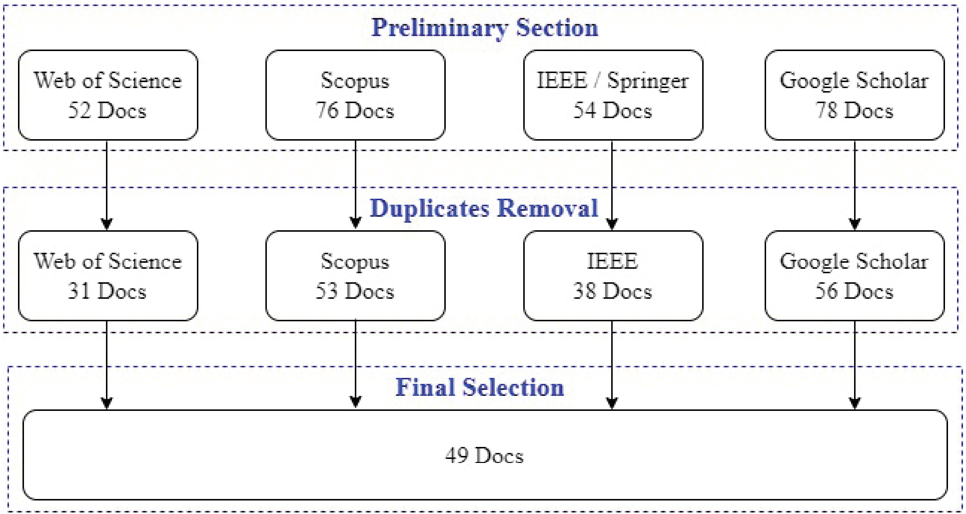


**Figure 5.** Selection process.

## RESULTS AND DISCUSSION

In order to find the answers to the Research Questions (RQ), articles from different conferences and journals have been collected and analyzed.

### RQ1. What Facial Sketch Mapping Methods Are Commonly Utilized?

Photo-sketch matching has been accomplished using different techniques. Figure 6 depicts several approaches to match a picture and sketch.

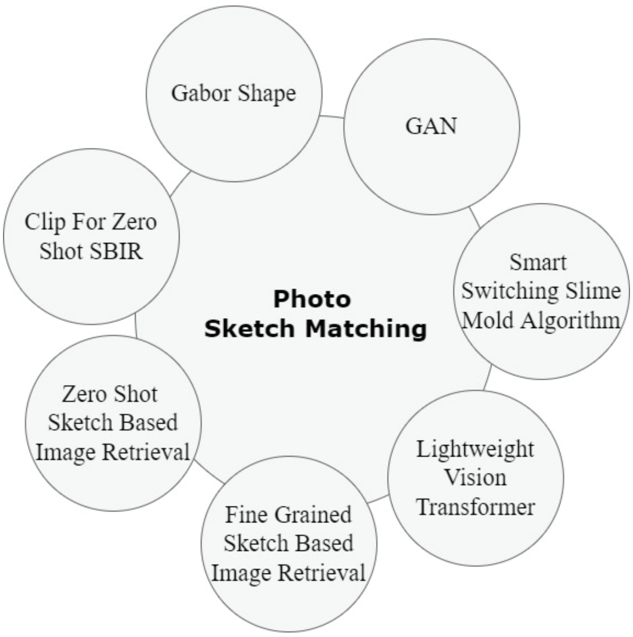


**Figure 6.** Different techniques for photo-sketch matching.

### Gabor Shape

It functions in levels: Initially, Gabor filters are employed to soften the fine texture. In the subsequent phase, the visual signals associated with face shape are represented through a random transform as a face descriptor [28]. Gabor filters are commonly used for face feature extraction due to their remarkable ability to capture visual information, including spatial localization, frequency, and orientation [29]. This approach has been utilized in research to evaluate its effectiveness on the CUFSF and CUFS datasets.

### GAN

In generative adversarial networks (GANs), two neural networks engage in competition using deep learning techniques to enhance prediction accuracy. A GAN consists of a generator and a discriminator. The discriminator features six convolutional layers, each equipped with different filters, and the final convolutional layer converts the output into a single dimension. Although it relies solely on the loss from the discriminator during training, it is associated with two distinct loss functions, and the failure of the generator is not taken into account. The discriminator is responsible for distinguishing between genuine and fake data produced by the generator. If a legitimate instance is mistakenly recognized as fake, or if a fake instance is incorrectly classified as real, the discriminator faces a penalty. The generator comprises encoder and decoder layers. The architecture utilized is U-Net. Research has been performed on various types of GANs, such as conditional, unconditional, cyclic, and style GANs. Figure 7 depicts the functioning of a GAN.

### Smart Switching Slime Mold Algorithm

This is tackled using three distinct types of procedures. The first involves a synthesis-oriented approach. The second is a projection-focused method, utilizing soft

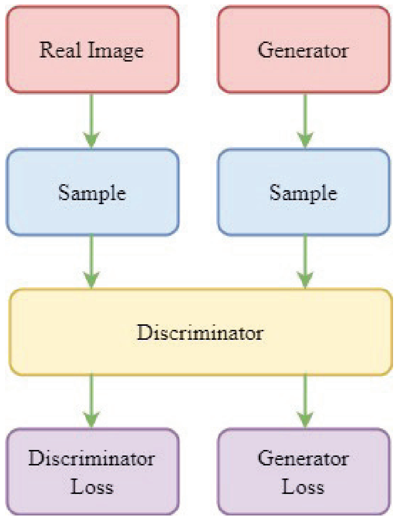


**Figure 7.** Working of GAN.

computing techniques to minimize discrepancies. The third category is based on optimization, utilizing various optimization methods for fine-tuning the parameters.

**Lightweight Vision Transformer**

It is a technique for categorizing images that utilizes a structure similar to that of a transformer to classify specific regions within an image. After pre-training on extensive datasets, the Vision Transformer (ViT) is fine-tuned for a smaller number of classes. Typically, classification is performed by adding an additional learnable classification token to the sequence. To capture the semantic features that enhance the precision of sketch face recognition, Convolutional Neural Networks are integrated with restricted LWVT.

**Fine-Grained Sketch-Based Image Retrieval**

The goal is to identify specific image using human's freehand drawing. This topic has garnered significant research attention in recent years because of its commercial significance and the challenges it poses visually. A major obstacle lies in the gap between the realms of drawings and photographs. Images are perspective representations of real-world objects depicted as dense pixels, whereas sketches are well-known line drawings that are interpreted in a personal and abstract manner.

**Zero-Shot Sketch-Based Image Retrieval**

The primary issue lies in the disparity of information between drawings and photographs; sketches primarily illustrate the shape of an object, resulting in less informative content compared to photographs [30]. Another challenge is that individuals typically create sketches with varying levels of abstraction, creating significant differences within the same class of sketches [31]. ZS-SBIR can be understood as the process of generating supplementary information that is lacking in the sketch to create images that can be compared. The research mainly concentrated on a transformer-based cross-modal network, which extracts local patches.

**Clip For Zero Shot SBIR**

CLIP is known as Contrastive Language Image Pre-training. It conducts analysis, image processing, and action recognition using video content. It has a remarkable capability to comprehend unmatched multi-modal data and is responsible for providing a semantic space for category transfer. In this regard, the challenges are tackled through the implementation of regularization words and methods [32]. To ensure that the global parameter operates consistently across all of them, the former is necessary to standardize the distance between sketches and photos. In the latter case, after the sketches and photos are split into patches, a random permutation of these patches occurs.

**Analysis**

Overall, traditional Gabor features offer interpretability but fail on complex, noisy sketches, whereas GANs achieve superior sketch-to-photo synthesis at the cost of data hunger and training instability. Optimization strategies like SSSMA are effective in hybrid pipelines but lack standalone strength. Lightweight vision transformers strike a balance between accuracy and efficiency, making them suitable for real-time use. While both GAN-based methods and Vision Transformer (ViT)-based approaches achieve competitive results in image generation and recognition tasks, review studies indicate that their data requirements and computational costs are indeed comparable. GANs typically demand large-scale datasets and significant computational resources for stable training due to issues like mode collapse and adversarial optimization. Similarly, Vision Transformers require extensive training data and high computational budgets, particularly because of their quadratic self-attention mechanism. Comparative analyses in the literature suggest that, although the architectural challenges differ, both families of models exhibit similar statistical demands in terms of dataset scale and computational complexity, making them broadly equivalent in these aspects. FG-SBIR excels in fine-grained, instance-level retrieval but requires highly detailed sketches. Zero-Shot SBIR relaxes the need for paired data, though with reduced accuracy, and CLIP extends this to large-scale zero-shot retrieval with strong generalization but limited forensic specialization. Together, these techniques demonstrate complementary strengths, highlighting the need for hybrid, data-efficient, and generalizable approaches in future photo–sketch matching research.

**RQ2. What Kinds of Datasets Have Been Utilized?**

The comprehensive analysis utilizes several datasets. Table 3 presents datasets along with sketching method. Starting with the artist who produces the sketch, the face sketch recognition system evaluates the sketch against various images saved in the database. The objective is to identify the image that most closely resembles the drawing.

**Table 3.** Dataset and method of sketch

| Dataset | | Type of sketches | Method of sketch |
|---|---|---|---|
| CUFS | CUHK | Viewed | Hand-drawn sketch |
| | AR | Viewed | Hand-drawn sketch |
| | XM2VTS | Viewed | Hand-drawn sketch |
| CUFSF | | Viewed | Hand-drawn sketch |
| UOM-SGFSV2 | | Forensic | Software-generated |
| E-PRIP | | Viewed | Software-generated |
| IIIT-D | | Semi-forensic | Hand-drawn sketch |
| QMUL-SHOE-V2 | | Viewed | Not known |
| QMUL-CHAIR-V2 | | Viewed | Not known |
| SKETCHY (EXTENDED) | | Viewed | Not known |
| TU BERLIN | | Viewed | Not known |
| QUICK DRAW (EXTENDED) | | Viewed | Not known |
| COLORFERET | | Semi-forensic | Not known |
| MULTIPLE ENCOUNTER | | Viewed | Not known |
| WILD-A (LFW-A) | | Viewed | Not known |
| FLICKR-FACES-HQ | | Viewed | Not known |

Various styles of sketching are available. These drawings are more precise and contain the most information about the subject. In the semi-forensic method, the artist is asked to create a drawing from memory after getting a brief chance to view a photograph of the subject. This technique, which is also employed in academic research, is considered more applicable to real-world situations than sketches made from observation since it factors in memory. The artist creates forensic illustrations based on these forensic sketches.

There are limited collections of images available for study, as most of these real-life examples originate from law enforcement departments. The greatest challenge lies in recognizing a brief glimpse of a face, often occurring under stress and influenced by the design of the sketch [33]. The IIIT-D database contains all three types of sketches analyzed in the SLR. It has 238 pairs of pictures created by a talented sketch artist, utilizing photographs sourced from various locations. Much of the research has been conducted on CUFS and CUHK. It features 295 images from XM2VTS, 123 from AR, and 188 from CUHK, totaling 606 images. These sketches were derived from artworks that were observed by an artist. Another relevant database is the EPRIP database. The AR Face dataset was used to create the images and composite drawings. In total, there are 123 pairs available. Four different groupings have been established for composite sketches drawn with various subjects.

### Analysis

For face sketch matching, CUFS and CUFSF remain the most commonly used controlled datasets, though they lack real forensic variability. IIIT-D and UOM-SGFSV2 introduce semi-forensic sketches and greater diversity, while E-PRIP provides more realistic forensic-style data, making them stronger candidates for practical evaluation. In contrast, general SBIR datasets like QMUL-SHOE-V2, QMUL-CHAIR-V2, and SKETCHY (Extended) focus on fine-grained object retrieval across categories, enabling advances in zero-shot and cross-domain matching but offering limited relevance to facial sketches. Together, these datasets have advanced the field, yet their constraints in scale, diversity, and forensic authenticity emphasize the urgent need for large, varied, and real-world benchmarks to build robust and deployable systems. The paper acknowledges the importance of ethical and fair data representation. Dataset bias, including demographic imbalance and limited diversity, can reduce the robustness of models and hinder their applicability across varied populations. Such imbalances often result in unequal performance, where certain groups are represented accurately while others are misclassified or overlooked. Addressing these challenges is essential to ensure fairness, reliability, and broader acceptance of face sketch matching systems.

### RQ3. What Challenges Have Image Sketch Recognition Methods Faced, and What Future Possibilities Do These Approaches Hold?

Different techniques have been employed to predict accuracy across various types of datasets, and these results have been compared with other methods to evaluate their effectiveness. To enhance the identification rate, some methods have incorporated optimization strategies. A significant challenge in photo-sketch matching is the modality gap, which has been extensively researched to reduce its impact. Another primary challenge in photo-sketch matching is the lack of paired training samples for training the deep learning network [34]. Additionally, overfitting has been identified as a problem. When comparing the experimental results with earlier techniques, it was found that these approaches could benefit from being applied to larger datasets and more intricate models. Some methods require addressing hardware and software constraints to yield more precise outcomes in the future [35]. GAN and sketch-based image retrieval techniques are the most frequently utilized methods that have consistently demonstrated success in systematic reviews of research. Upcoming studies will explore alternative pre-training techniques that may enhance the convergence of the sketch-to-photo translation method, which currently relies on limited data. The possibility of generating photographs from intricate drawings derived from eyewitness testimony is another area that warrants exploration; in this scenario, the sketches could vary greatly from the actual photographs.

**Analysis**

Photo–sketch recognition faces persistent challenges such as limited and controlled datasets, loss of detail in hand-drawn sketches, and the large modality gap between photos and sketches. Deep learning methods like GANs and transformers perform well but require extensive data, high computation, and risk overfitting, while handcrafted and optimization-based methods struggle with scalability. Future progress depends on developing larger, diverse, and forensic-oriented datasets that capture real-world variability. Hybrid approaches combining interpretability with deep learning robustness, along with lightweight models for real-time use, are promising directions. Emerging paradigms like zero-shot and CLIP-based retrieval can reduce dataset dependency and enhance generalization, but they need tailoring for forensic applications. Overall, the focus must shift toward scalable, explainable, and deployable systems for real-world law enforcement and security use.

**RQ4: Which Different Performance Metrics Are Used in the Study?**

Various studies utilize indicators to evaluate the effectiveness of the techniques. These studies use a range of datasets, including CUFSF, CUHK, AR, XM2VTS, UOM-SGFSV2, QMUL-SHOE-V2, among others, to measure performance [36]. The use of these databases has been extensive in the research. Some of these databases are publicly available, whereas others are proprietary.

The following performance metrics are utilized:

**Mean Recognition Rate**

This refers to the accurately recognized images. The aim of the examination is to achieve consistency between the drawing and the photo. The formula used to determine this is (Accurately recognized images / Total images) * 100.

**Cumulative Matching Characteristic (CMC)**

It is a metric employed to assess the precision of identification and recognition algorithms across various rankings to gauge their effectiveness [37]. The comparative arrangement of match scores from every biometric sample is typically employed to evaluate identification performance in biometric systems. CMC utilizes values ranging from one to k to measure the effectiveness of identification on the x-axis. The y-axis represents identification rates corresponding to every rank.

**Fitness Value Analysis**

This measure is also utilized to compare the success of a specific algorithm against others. This comparison is performed by averaging the results of all algorithms. In the context of image sketch matching, the algorithm that achieves the highest value is considered the most effective.

**Inception Score**

It assesses the quality of the generated photographs. The score varies from 0 to 1, with 0 indicating the lowest quality and 1 indicating the highest. The Inception score primarily evaluates two main aspects: quality and variety. Quality refers to the image's overall standard, while variety pertains to the diversity of the images; the subject should be identifiable. This indicates the necessity for a broad range of images. The analysis metrics are illustrated in Table 4.

**Analysis**

Different studies in photo–sketch recognition have adopted varied performance metrics, each highlighting different aspects of model effectiveness. Mean Recognition Rate (MRR) remains the most widely used, offering a straightforward measure of overall matching accuracy but often overlooking rank-level insights. Cumulative Matching Characteristic (CMC) addresses this by

**Table 4.** Performance measures

| S. No. | Name | Definition |
|---|---|---|
| 1 | Average Recognition Rate | This signifies the count of images that correctly align with the training photos |
| 2 | Cumulative Matching Feature | This represents standard evaluation metrics utilized for techniques in re-identification |
| 3 | Fitness Value | It is used to determine how closely a specific solution matches the optimal answer to the problem |
| 4 | Inception Score | This is applied to evaluate the caliber of images produced by GANs |

evaluating performance across multiple ranks, making it more informative in retrieval-based tasks. Fitness Value is mainly employed in optimization-driven approaches, capturing how well parameters or features are tuned, though it lacks universal comparability across methods. Inception Score (IS), widely applied to generative models, assesses the quality and diversity of synthesized sketches or photos, yet does not always align with human visual judgment. Together, these metrics provide complementary evaluations, enabling a more comprehensive understanding of model performance across recognition, retrieval, optimization, and generation tasks.

**RQ5: Which Pre-Processing Methods Have Been Employed thus Far?**

So far, numerous pre-processing methods have been utilized for images. Preparing an image through preprocessing is essential as it enhances its features and readies it for further processing.

**Face Alignment and Cropping**

They are widely used preprocessing steps in photo–sketch recognition to standardize facial geometry. Alignment corrects pose variations by mapping key landmarks (eyes, nose, mouth) to a consistent reference, while cropping removes background clutter and focuses solely on the facial region. These steps reduce intra-class variation, enhance feature extraction, and improve recognition accuracy. By ensuring uniformity across samples, they make subsequent learning and matching more robust and reliable.

**Grayscale Conversion**

It is a fundamental preprocessing step in photo–sketch recognition that reduces images to intensity values, removing color information while retaining structural and textural details. Since sketches inherently lack color, this transformation minimizes the modality gap between photos and sketches. It also reduces computational complexity and memory requirements, allowing models to focus on shape, edges, and contrast, which are more relevant for recognition. By emphasizing structural features, grayscale conversion improves consistency and matching accuracy across modalities.

**Noise Reduction and Edge Enhancement**

They both are complementary preprocessing techniques in photo–sketch recognition. Noise reduction smooths unwanted distortions and artifacts, while edge enhancement emphasizes contours and structural details critical for distinguishing facial features. Together, they balance clarity and sharpness, ensuring that important sketch strokes or photo edges are preserved without distortion. This combination improves feature extraction, enhances robustness against poor sketch quality, and ultimately strengthens recognition accuracy.

**Histogram Equalization**

It is a contrast enhancement technique widely used in photo–sketch recognition to improve the visibility of facial features. By redistributing pixel intensity values, it enhances edge clarity and highlights key structural details that are often faint in sketches. This normalization reduces the impact of lighting variations across photos, ensuring more consistent inputs for recognition. As a result, models trained on equalized images achieve improved feature extraction and higher matching accuracy.

**Scaling and Normalization**

They are essential preprocessing steps to ensure uniformity across photo–sketch datasets. Scaling resizes images to a consistent dimension, eliminating variability caused by different resolutions, while normalization adjusts pixel intensity ranges for balanced feature representation. These steps reduce biases in model training, stabilize learning, and make comparisons across samples more reliable. By standardizing inputs, scaling and normalization enhance the robustness and accuracy of recognition systems.

**Analysis**

Pre-processing remains a vital step in improving the effectiveness of photo–sketch recognition systems. Face alignment and cropping standardize pose and scale, reducing variability and ensuring that only the facial region is analyzed. Grayscale conversion minimizes the modality gap by discarding color while preserving essential structural details, and histogram equalization enhances contrast to make features more visible under varying illumination. Noise reduction and edge enhancement further refine sketches and photos by suppressing distortions while highlighting critical contours. Scaling and normalization provide uniform input dimensions and balanced intensity ranges, enabling more reliable feature extraction and model training. Collectively, these techniques contribute to improved matching accuracy, though their effectiveness remains strongly dependent on the quality of both input images and preprocessing itself.

**Outcomes**

The comparison of techniques is shown in Table 5.

Euclidean distance is a widely used metric that quantifies similarity by computing the straight-line distance between feature vectors in a multidimensional space. Rank-k accuracy measures the likelihood that the correct match is retrieved within the top k results, with rank-1 accuracy denoting the strictest case where the correct match must appear first. Mean Average Precision (mAP) aggregates precision across varying recall levels to provide a comprehensive measure of retrieval effectiveness. Similarly, top-k accuracy evaluates the presence of the correct match within the top k predictions, while Precision quantifies the ratio of relevant items among the top k retrieved results, highlighting early retrieval performance.

**Table 5.** Comparison of techniques

| S.No. | Technique | Features | Purpose | Dataset Used | Metrics |
|---|---|---|---|---|---|
| 1 | Gabor Shape | Handcrafted descriptors (edges, orientations, multi-scale filters) | Basic photo–sketch matching using low-level shape cues | CUHK Face Sketch, CUFSF, small object datasets | Accuracy evaluated with Euclidean/$\chi^2$ distance and Rank-k accuracy. |
| 2 | GAN | Learned cross-modal representations + synthesized photo/sketch | Reduce domain gap by generating photo-like or sketch-like data | CUHK Face Sketch, CUFSF, LFW-Sketch, IIIT-D | Rank-1 accuracy is used for accuracy measurement |
| 3 | Smart Switching Slime Mold Algorithm | Optimized feature selection / weight adjustment | Acts as an optimizer for hybrid features in sketch–photo pipelines | CUHK Face Sketch, CUFSF, LFW-Sketch, IIIT-D | Accuracy measured with Rank-1 and mAP. |
| 4 | Lightweight Vision Transformer | Learned embeddings via self-attention (compact transformer) | Efficient cross-modal embedding for sketch ↔ photo | CUHK Face Sketch, LFW-Sketch, QMUL FG-SBIR | Accuracy evaluated using Rank-1 Top-k accuracy, and mAP. |
| 5 | Fine Grained SBIR | Deep features + metric learning (triplet/contrastive) | Instance-level retrieval (find the exact person/object, not just category) | QMUL FG-SBIR (Shoes, Chairs, Handbags), CUHK | Accuracy measured using Rank-1, and mAP. |
| 6 | Zero-Shot SBIR | Semantic embeddings (attributes, word vectors, unseen class generalization) | Retrieve sketches/photos from unseen categories | Sketchy | Accuracy Evaluated with Rank-1 and Precision@k. |
| 7 | CLIP for Zero-Shot SBIR | Pretrained joint vision–language embeddings. Takes sketch as image input | Zero-shot cross-domain retrieval using CLIP space + prompts | Sketchy, domain-adapted LFW/CUHK | Accuracy measured using Rank-1, and mAP. |

In addition to improvements in methods and accuracy, fairness and reliability are still big problems in photo-sketch matching. The quality and variety of the training datasets have a big effect on how well most models work. A lack of diversity in the sample might cause bias, which can make results differ by age, gender, or ethnicity [38,39]. Deep learning and hybrid methods work well on benchmark datasets, but they don't always work as well on sketches with exaggerated or forensic-level aberrations [40]. Generative techniques like GANs or diffusion models, while effective in minimizing domain gaps, may produce artifacts or illusory details that undermine authenticity [41]. These difficulties pose challenges for practical applications, especially in forensic and security settings, where even modest inaccuracies can lead to substantial repercussions. In addition to technical progress, future research must prioritize fairness-aware training, comprehensive error analysis across diverse populations, and transparent evaluation processes. These kinds of activities will assist make sure that photo-sketch matching gets better not only in terms of accuracy but also in terms of how it is used in a responsible and ethical way.

The review indicates that no single technique universally outperforms others across all conditions; rather, the choice of method depends heavily on the application context (e.g., forensic sketches, artist-drawn sketches, or automatically generated sketches), suggesting that combining the strengths of classical and deep learning methods, acknowledging domain adaptation, and ensuring scalability for real-world deployment.

The novelty of this review lies in the comprehensive synthesis and reorganization of existing research on photo–sketch matching. The work goes beyond by providing a critical comparison of techniques, identifying their strengths, limitations, and applicability, such as in forensics.

The results show that in order to prevent overfitting, a large dataset is required for training. The intricacy of the issue requires an extensive set of forensic drawings. The accuracy is affected by multiple factors and can be improved by using hybrid techniques or deep learning approaches.

## CONCLUSION

The systematic assessment of the work is done in the paper. The primary research questions to be explored are identifying the type of dataset used, examining the various techniques available for matching photo sketches, and discussing the challenges and potential applications of these methods. The review includes only conference papers and journal articles that propose research into photo matching.

The research summarizes all the key questions related to different methods of photo sketch matching using multiple datasets which are available. Though this topic has been explored by many researchers but still there is a gap in developing scalable and hybrid approaches that combine interpretability with deep learning, while addressing domain adaptation, robustness, and deployment in practical forensic scenarios. The process of transforming through deep learning is quicker and more efficient compared to earlier techniques. To avoid overfitting, it required a large training dataset, but sketch images sometimes missed facial characteristics, which could introduce significant noise. Future work needs to focus on building larger and more diverse benchmark datasets and developing hybrid models that combine classical interpretability with deep learning robustness. This is necessary to handle variations in sketch styles and real-world conditions, and can be achieved through domain adaptation, data augmentation, and scalable architectures designed for practical deployment.

Future research in photo–sketch matching prioritizes self-supervised learning models as promising paradigms to address data scarcity and modality gaps. Self-supervised learning can derive transferable, modality-invariant representations from large-scale unlabelled datasets, thereby enhancing cross-domain generalization. On the other hand, Diffusion models are also emerging as a promising approach in photo–sketch matching due to their ability to generate high-quality and detailed images through iterative denoising. At present, several researchers are developing algorithms aimed at matching illustrations and images. In the future, it may be possible to create an algorithm that integrates deep learning and artificial intelligence to match photos and sketches with greater accuracy by leveraging vast datasets. A feature extraction method can be employed to identify unique traits in drawings and photographs.

## AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

## DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

## CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## ETHICS

There are no ethical issues with the publication of this manuscript.

## STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

## REFERENCES


[1] Malathy N, Navin G, Sriram S, Bala Subramabiyan S. A secure framework for multimedia transmission in medical images using DNA cryptography. Sigma J Eng Nat Sci 2025:222–233. [CrossRef]

[2] Li CX, Zhang D, Hu Z, Wu XJ. Modality fused class-proxy with knowledge distillation for zero-shot sketch-based image retrieval. IEEE Trans Circuits Syst Video Technol 2025;35:6158–6169. [CrossRef]

[3] Guo M, Xiong M, Huang J, Hu X, Peng T. Face photo-to-sketch portraits transformation via generation pipeline. Vis Comput 2025;41:1183–1196. [CrossRef]

[4] Bian H, Lv B, Guo Y, Zhang B, Du K. Sketch face recognition method based on local-global adapter. IEEE Access 2025;13:59244–59253. [CrossRef]

[5] Kumar D, Pandey RC, Mishra AK. A review of image features extraction techniques and their applications in image forensic. Multimed Tools Appl 2024;83:87801–87902. [CrossRef]

[6] Khattar P, Mishra S, Tanwar R, Verma A, Bhatia S. Decoding information: A dual modality approach for sign language recognition. In: Proceedings of the 2024 International Conference on Computing, Sciences and Communications (ICCSC); 2024. p. 1–5. [CrossRef]

[7] Gupta S, Adhikari S, Zaid MS, Verma A, Bhatia S. Advancements in facial image processing for human age and gender identification. In: Proceedings of the 2024 International Conference on Computing, Sciences and Communications (ICCSC); 2024. p. 1–5. [CrossRef]

[8] Sharma A, Bhatia S, Verma A. Weather monitoring and cloudburst prediction based on machine learning algorithms: An initiative towards disaster management. In: Advances in Disaster Management; 2024. p. 589–603. [CrossRef]

[9] Zhang X, Yang C, Li F, Li Y, Fu B, Wang S. Multi-scale feature extraction and aggregation network for electroencephalography classification in face photo-to-sketch recognition task. IEEE Trans Biomed Eng 2025;73:709–719. [CrossRef]

[10] Raj G, Verma A, Dalal P, Shukla AK, Garg P. Performance comparison of several LPWAN technologies for energy constrained IOT network. Int J Intell Syst Appl Eng 2023;11:150–158.

[11] Sezgin A. A complete study on and-product of soft sets. Sigma J Eng Nat Sci 2025;43:1–14. [CrossRef]

[12] Pattnaik I, Dev A, Mohapatra AK. CASK-Net fusion: Multi branch approach for cross-age sketch face recognition. Signal Process Image Commun 2025;138:117369. [CrossRef]

[13] Rodrigues AP, et al. Deep learning in forensic sketch analysis. In: Proceedings of the 2025 International Conference on Artificial Intelligence and Data Engineering (AIDE); 2025. p. 45–50. [CrossRef]
[14] Zhang T, Xie X, Du X, Xie H. Sketch-guided scene image generation with diffusion model. Comput Graph 2025;129:104226. [CrossRef]
[15] Kiral A. The effect of LRB stiffness changes with and without supplemental viscous dampers on seismic responses of an experimentally verified mdof building. Sigma J Eng Nat Sci 2025:301–315. [CrossRef]
[16] Setumin S, Aminudin MFC, Suandi SA. Canonical correlation analysis feature fusion with patch of interest: A dynamic local feature matching for face sketch image retrieval. IEEE Access 2020;8:137342–137355. [CrossRef]
[17] Deswal S, Verma A. Efficient routing protocol for IoT networks based on fog computing and routing protocol of low power lossy networks. Int J Internet Protoc Technol 2023;16:176–184. [CrossRef]
[18] Lu D, Chen Z, Liu C, Hu Y, Cai L, Wu QMJ. Few-shot facial sketch synthesis via progressive domain gap reduction. IEEE Trans Inf Forensics Secur 2025;20:7121–7136. [CrossRef]
[19] Verma A, Deswal S. FOG-RPL: Fog computing based routing protocol for IoT networks. Recent Adv Electr Electron Eng 2023;16. [CrossRef]
[20] Upreti K, Verma VS, Verma A, Vats P, Patil S, Kuwar V. Pandemic pulse: Unveiling insights with the global health tracker through AI and ML. In: Global Health Insights; 2024. p. 191–203. [CrossRef]
[21] Kapoor A, Kapoor S, Mishra K, Jain H, Upreti K, Verma A. Catalyzing security and efficiency: Blockchain's integration with IoT and cloud computing. In: Blockchain Integration; 2024. p. 457–467. [CrossRef]
[22] Pahuja S, Negi S, Verma A, Rathi P, Narang N, Chawla R. An authentication protocol for secure tag-reader communication. In: Proceedings of the 2012 IEEE Students' Conference on Electrical, Electronics and Computer Science; 2012. p. 1–4. [CrossRef]
[23] Upreti K, Verma A, Parashar J, Vats P, Verma A, Singh J. A comparative analysis of LSB & DCT based steganographic techniques: Confidentiality, contemporary state, and future challenges. In: Proceedings of the 2023 6th International Conference on Contemporary Computing and Informatics (IC3I); 2023. p. 1581–1588. [CrossRef]
[24] Yacoob S, Sujatha B, Anudeep T, Sharvi BE, Sai GVVS, Antony KC. Face sketch construction and recognition. In: Proceedings of the 2025 International Conference on Cognitive Computing in Engineering, Communications, Sciences and Biomedical Health Informatics (IC3ECSBHI); 2025. p. 1151–1155. [CrossRef]
[25] Verma A, Deswal S. Comparative study of routing protocols for IoT networks. Recent Pat Eng 2023;17. [CrossRef]
[26] Upreti K, Poonia RC, Verma A, Divakaran P, Mittal S, Vats P, et al. Blockchain computing: Unveiling the benefits, overcoming difficulties, and exploring applications in decentralized ledger infrastructure. In: Blockchain Computing; 2024. p. 397–410. [CrossRef]
[27] Prasad JS, Verma A. Optimum path routing algorithm using ant colony optimisation to solve travelling salesman problem in wireless networks. Int J Wirel Mob Comput 2017;13:131. [CrossRef]
[28] Pahuja S, Kumar R, Verma A, Negi S, Arora H. Priority based approach against congestion in sensor network. In: Proceedings of the 2011 Annual IEEE India Conference; 2011. p. 1–4. [CrossRef]
[29] Khaliluzzaman M. Comparative analysis on real-time hand gesture and sign language recognition using convexity defects and YOLOv3. Sigma J Eng Nat Sci 2024;42:99–115. [CrossRef]
[30] Prasad JS, Verma A. Performance enhancement by efficient ant colony routing algorithm based on swarm intelligence in wireless sensor networks. Int J Wirel Mob Comput 2017;12:232. [CrossRef]
[31] Verma A, Vashist PC. Enhanced clustering ant colony routing algorithm based on swarm intelligence in wireless sensor network. In: Proceedings of the 2015 International Conference on Advances in Computer Engineering and Applications; 2015. p. 150–154. [CrossRef]
[32] Guo M, Xiong M, Huang J, Hu X, Peng T. Face photo-to-sketch portraits transformation via generation pipeline. Vis Comput 2025;41:1183–1196. [CrossRef]
[33] Pahuja S, Negi S, Verma A, Rathi P, Narang N. A novel intrusion preventive routing scheme for data dissemination in sensor network. In: Proceedings of the 2012 IEEE Students' Conference on Electrical, Electronics and Computer Science; 2012. p. 1–6. [CrossRef]
[34] Zhu M, Chen J, Wei X, Wang N, Gao X. Fine-detailed facial sketch-to-photo synthesis with detail-enhanced codebook priors. IEEE Trans Circuits Syst Video Technol 2025;36:1075–1088. [CrossRef]
[35] Tang D, Jiang X, Zhang Y, Dai Y, Lin Y. Ipdm: Identity preserving diffusion model for face sketch and photo synthesis. Mach Vis Appl 2025;36:34. [CrossRef]
[36] Upreti K, Verma A, Mittal S, Vats P, Haque M, Ali S. A novel framework for harnessing AI for evidence-based policymaking in e-governance using smart contracts. In: AI for E-Governance; 2023. p. 231–240. [CrossRef]
[37] Haque M, Kumar VV, Singh P, Goyal AA, Upreti K, Verma A. A systematic meta-analysis of blockchain technology for educational sector and its advancements towards education 4.0. Educ Inf Technol 2023;28:13841–13867. [CrossRef]

[38] Buolamwini J. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In: Proceedings of the 1st Conference on Fairness, Accountability and Transparency; 2018.

[39] Drozdowski P, Rathgeb C, Dantcheva A, Damer N, Busch C. Demographic bias in biometrics: A survey on an emerging challenge. IEEE Trans Technol Soc 2020;1:89–103. [CrossRef]

[40] Zhang W, Wang X, Tang X. Coupled information-theoretic encoding for face photo-sketch recognition. In: Proceedings of the CVPR 2011; 2011. p. 513–520. [CrossRef]

[41] Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2019. p. 4396–4405. [CrossRef]