International Journal of Engineering Trends and Technology
ISSN: 2231-5381/ https://doi.org/10.14445/22315381/IJETT-V74I5P124

Volume 74 Issue 5, 362-383, May 2026

Review article

A Multimodal Plagiarism Detection Framework Gemini AI for Text and Image Content Integrity

Anjali Naudiyal[1], Kapil Joshi[2], Rahul Mahala[3], Shivani Pant[4], Mohammed Ghouse Aleem[5]

[1]*Uttaranchal School of Computing Sciences, Uttaranchal University, Dehradun, Uttarakhand, India*
[2]*Uttaranchal Institute of Technology, Uttaranchal University, Dehradun, Uttarakhand, India*
[3]*Law College Dehradun, Uttaranchal University, Dehradun, Uttarakhand, Indian*
[4]*School of Science & Technology, Swami Rama Himalayan University, Dehradun, Uttarakhand, India*
[5]*Computer Engineering, University of Technology and Applied Science, Nizwa, Sultanate of Oman*

[3]*Corresponding Author : rahulmahala98@gmail.com*

Received: 19 January 2026 Revised:09 March 2026 Accepted: 28 March 2026 Published: 30 May 2026

Abstract - *In the recent advancement of technologies, Plagiarism in academia is increasing day by day. Nowadays, Plagiarism is not in the form of text, but it is in the form of images, screenshots, figures, logos, watermarks, and diagrams. The proposed system introduces several technical innovations that set it apart from conventional text-based plagiarism detection tools. One of the most significant advancements is its multi-modal analysis capability, which combines both visual and textual content understanding through integration with Google's Gemini 1.5 Flash model (Gemini AI) with deep learning techniques. The proposed seven-stage framework unifies computer vision, natural language processing, and adaptive similarity analytics to move beyond conventional fingerprint or perceptual hash methods. First, textual extraction isolates embedded or overlaid text (captions, watermarks, OCR passages), supplying linguistic cues. Second, visual decomposition segments salient objects, layout structures, color palettes, and stylistic signatures. Third, authenticity assessment estimates manipulative edits. cropping, splicing, style transfer, generative fill via anomaly and provenance signals. Fourth, source candidate retrieval uses multimodal embeddings to surface likely originals or semantically proximate precursors from reference corpora and web indices. Fifth, plagiarism indicator evaluation aggregates cross-image overlaps: localized patch similarity, reconstructed text alignment, stylistic congruence, and watermark inheritance. Sixth, web search recommendation dynamically composes discriminative keyword visual descriptor queries that can expand external source discovery. Seventh, similarity fusion and scoring combine weighted textual, structural, and deep feature distances through a learned calibration layer. producing dual quantitative outputs: an Authenticity Integrity Score and a Plagiarism Likelihood Score. Experiments on a curated benchmark mixing authentic, lightly edited, heavily manipulated, and synthetically generated images show robust discrimination across perturbations. Preliminary comparative analyses indicate improved recall of subtle derivative works while maintaining controlled false positives. The system is implemented in practical applications in academic publishing, news verification, creative asset management, and legal evidence triage, while establishing an extensible foundation for future provenance standards and investigative journalism.*

Keywords - *Deep Learning, Image Plagiarism (IP), Text-Based Image Plagiarism (TBIP), Content Authenticity, Source Detection.*

1. Introduction

The increasing demand for digitization in the field of academics plays a vital role in accessing the content in the form of screenshots. This is the new way to access the content without providing the citation to the original author. When the researcher uses the content or data of the original author without providing any citation, then it is called Plagiarism. Author plagiarism, according to authors, is when writers steal texts or software without giving credit to the original creators [1]. Plagiarism is mainly divided into Text and image-based Plagiarism. In text-based Plagiarism, text content is copied from the research paper without giving the proper citation to the original author. Due to the advancement of the technologies, the Image-Based Plagiarism (IBP) and Text-Based Image Plagiarism (TBIP) occurs in the form of screenshots of academic content, extracted figures or diagrams from research papers, social media images that paraphrase or duplicate original text, or scanned or photographed textbook pages used without citation. In this type of Plagiarism, the author either takes a screenshot of the content or uses the Image of the research paper without providing any citation to the original author. IP only concerns

artistic or photographic content, but TBIP focuses on the semantic and syntactic similarity of the written text present within the Image. The present problem becomes more complex when the plagiarist alters font styles, colors, backgrounds, or layouts to disguise the copied material. Several reasons are concluded for the occurrence of Plagiarism [2]. Several traditional plagiarism methods are introduced, but they only focus on the content of the paper or source code present in the technical documentation. Image-based Plagiarism is increasing day by day, so there is a need to bridge the gap between image analysis and textual plagiarism detection. Deep Learning (DL) and Computer Vision (CV) are a part of Artificial Intelligence, which plays a vital role in the field of Images. They use Optical Character Recognition (OCR), which is combined with Natural Language Processing (NLP) techniques. These technologies help to develop. the system through which plagiarized text is detected from the images. Several plagiarism tools are available for textual documents. e.g., Turnitin, Grammarly, PlagScan, but they are unable to detect Plagiarism in the images or the content present in the images [3]. The inspection of the images in the document by a human being is time-consuming, error-prone, and impractical at scale. The primary problem found in the present research is: To detect Plagiarism in the images or the content present in the Images automatically. The problem of image plagiarism detection is to find instances of unauthorized usage and changes to visual and textual information contained in an image, which requires state-of-the-art technology for image content analysis, text recognition, and source identification. The existing systems have their limitations, as they usually focus on exact image matching, which is not effective for identifying sophisticated Plagiarism, such as changes made to an image by cropping, changing content, etc. Using Optical Character Recognition technology to obtain clean, accurate, and correct text from noisy, distorted, mixed language, and stylized images, and then semantically comparing it with existing data for duplication and paraphrasing is a big challenge. In addition, existing systems usually do not correctly identify the source of plagiarized images, which is another limitation, as they do not have contextual knowledge about image semantics, which is required for identifying similar concepts. Moreover, there is no quantitative scoring system to determine how much Plagiarism is involved and how genuine it is, which is another area that needs to be addressed. The present research used the strengths of Artificial Intelligence for visual data processing with the capability of the Deep Learning (DL) techniques to develop DL based system for Image Plagiarism (IP). The present focused on the visual and textual content with the help of a multi-modal analysis framework, which enables the detection of replicated visual elements, embedded or overlaid textual similarities. The main part of the system is to figure out measurable indicators to the degree of detected Plagiarism, facilitating objective assessment and decision-making with the creation of quantitative scoring mechanisms. In the flow of assessment research, the focus is also on detailed source detection and authenticity analysis, which ensures the origin of the content and its originality will be verified. The study is able to figure out the effectiveness of AI-powered image analysis, which is helpful in establishing a solution to address the challenges in visual media. To begin with, it creates a new multimodal framework for content evaluation, for it uses cutting-edge AI models to perform visual and textual analysis. Also, it develops a new methodology for quantitative analysis with set criteria for measuring authenticity and Plagiarism, which can be scored on a scale of 0-100% based on a weighted evaluation framework. Moreover, the study articulates a full analysis pipeline of seven stages, which includes extracting text, analysing the visual content, assessing authenticity, detecting the source, evaluating Plagiarism, recommending web searches, and detecting the similarity. Last, the study illustrates the practical applicability of the proposed framework by developing a real-life web application that demonstrates the functionalities and efficiency of the framework. The current system assists in preserving scholarly work's academic and digital integrity in connection with the 4th Sustainable Development Goals (SDGs). It contains several types of images, which include English-language text, scanned documents, screenshots, image figures, or tables. The current study also provides scope for future research to incorporate more languages or diverse visual domains.

The current study, though with the development of textual plagiarism detection and the availability of vision-language models such as OpenAI's CLIP and BLIP-based architectures, is still not optimized for academic multimodal plagiarism detection. The majority of the techniques either concentrate solely on textual similarity or concentrate on perceptual Image matching techniques rather than semantic image understanding. The current techniques also have some limitations, such as "Cross-modal semantic alignment, Standardized quantitative authenticity scoring, Robust benchmarking of heterogeneous academic data, Statistical validation of multimodal performance improvements."

The limitations of the current techniques also emphasize the need for a unified and statistically validated multimodal academic plagiarism detection technique.

The major contributions of the current research are as follows:

A unified multimodal fusion approach, incorporating OCR, CNN-based visual feature extraction, and transformer-based semantic embedding.

The integration of the Gemini 1.5 Flash multimodal model, developed by Google, for cross-modal reasoning.

The provision of a dual quantitative scoring system that offers the possibility of determining the likelihood of authenticity and Plagiarism.

Experimental benchmarking with statistical validation and correlation analysis.

Comparative evaluation with traditional OCR-based and text-based plagiarism detection systems.

The rest of the paper is organized as follows: Section 2 presents a detailed literature review of existing methodologies and deep learning techniques for image-based plagiarism detection. Section 3 presents the proposed methodology and system architecture. Section 4 presents experimental results and performance evaluation. Section 5 concludes the paper with future research directions.

2. Literature Review

Previous work related to plagiarism detection in textual documents is discussed in the present section, which uses OCR for image-based text extraction, deep learning in text recognition, and emerging solutions for identifying text plagiarism within images. The final section identifies current research gaps that justify the proposed implementation. Distinct popular tools are available, like Turnitin, PlagScan, and Grammarly, used to detect similarity against large corpora. Recently, deep learning-based models like BERT and Sentence-BERT have been used to detect paraphrased Plagiarism using contextual embeddings [4]. All these systems focused on structured, machine-readable text. Text embedded in image formats (e.g., scanned PDFs, screenshots) remains largely inaccessible to such tools without OCR preprocessing [5]. Research in the field of plagiarism detection in textual documents involves verbatim copying, paraphrasing, synonym substitution, or sentence restructuring. Several traditional methods are introduced in this domain, which include string matching algorithms like Rabin-Karp and Knuth-Morris-Pratt (KMP), as well as shingling and fingerprinting techniques (e.g., Broder, 1997), which compare the overlapping of sequential words. Vector space models such as TF-IDF are used to represent and compare the content numerically. More advanced techniques are stylometric analysis, which is used to capture the features related to writing style, and semantic similarity models have the ability to analyze deeper contextual understanding with the help of word embeddings or transformer-based networks like BERT [6]. In the context of image plagiarism detection, some techniques are introduced that are only focused on analysing visual features [7]. Perceptual hashing is a technique through which compact representations of images are generated for comparison. The techniques are based on the concept of creating hash values that remain unchanged despite minor transformations such as compression, resizing, or slight color changes, and are unable to detect other forms of Plagiarism, such as changes in content, cropping, and other forms of image manipulation, and also lack the ability to understand the semantic meaning of the images. Feature matching techniques such as Scale-Invariant Feature Transform (SIFT) [8], Speeded Up Robust Features (SURF), and ORB, are used for detecting color histogram analysis, and edge detection techniques for structural patterns and contours of images. While feature-based methods offer improved robustness compared to perceptual hashing, they still struggle with semantic understanding and cannot effectively detect Plagiarism that involves conceptual similarity rather than visual similarity. These methods also require significant computational resources for large-scale applications. Deep Learning is a part of AI that plays an important role in detecting text plagiarism, which is embedded within images by text recognition through OCR and Natural Language Processing techniques (NLPs) for semantic comparison [9]. Integration of Convolutional Neural Networks (CNNs) and OCR systems helps to extract spatial and visual features from character regions with the help of models like CRAFT and CRNN [Z]. Transformer-based models such as BERT and RoBERT generate contextual embeddings after text extraction and have the ability to detect semantically similar content that is paraphrased or reworded. Siamese and triplet networks are implemented to learn distance-based similarity between text embeddings, which are used to measure degrees of semantic alignment [10]. Systems are able to identify the reusable content at the conceptual level, which denotes semantic Plagiarism. With the help of these approaches, visual and semantic reuse of textual content is detected. Recently, deep learning has enhanced the capabilities to analyze images by some innovations, such as transfer learning, which is able to adopt the pre-trained models for specific tasks, and multimodal learning, which integrates visual and text data streams and attention mechanisms that direct focus for relevant regions within images. All the above advancements have laid the groundwork for the creation of strong AI-based plagiarism detection systems [11]. A Convolutional Neural Network (CNNs) model developed is presented in the paper, which is used for the classification of images and for checking the similarity of the images with the help of feature extraction. Feature extraction capabilities of CNNs improved in the presented system, but they only focused on visual similarity in comparison to the semantic understanding of ResNet, Inception, and EfficientNet [12]. Today's AI technologies, particularly those based on modern neural networks, promise and in some cases already deliver radical changes in the way we analyze, interpret, and create content in various fields. Some Large Language Models (LLMs) like GPT and BERT take into account the context, semantics, and several other layers of interaction within language, enabling them to interpret and generate human-like text. Vision language models like CLIP and BLIP work on a combination of visual and textual data and can perform tasks like image captioning, answering questions, and searching across multiple models with a high degree of accuracy. On the other hand, some other generative AI technologies, such as models like DALL-E and Sora, enable the analysis and generation of complex content ranging from artificial images to complex narratives based on a prompt provided. Furthermore, the development of the field of multi-modal analysis, enabling the simultaneous evaluation

of several types of information, leads to an increased level of understanding of the context and allows for the improvement of the quality of decision-making in the detection of Plagiarism, evaluation, and generation of reports [13]. All these developments represent a major leap forward in the development of AI systems. The presented paper demonstrated multimodal AI models to enhance the analysis of images. Visuals and textual content are understood by the modal CLIP [14]. The recent advancement in multimodal AI, the Gemini model of Google, has been introduced. Gemini 1.5 Flash model has the ability to understand and analyze the complex visual and textual elements with a detailed textual description, which is best for plagiarism tasks. To detect Plagiarism on the paragraph level, phrases with the help of Longformer gain a 90.1 F1 score and 96 when using a fine-tuned GPT-3.5 [15]. Previous Optical Character Recognition (OCR) technology is unable to extract the textual content from images and has no ability to understand and analyze the relationship between extracted text and visual content.

Nowadays, modern OCR like Tesseract and cloud-based services like Google Cloud Vision API have the ability to extract text from images. These systems have the ability to understand their context and relationship to visual elements, making them more effective for plagiarism detection. Advancement in the field of Deep Learning enhances the OCR capabilities. Several State-of-the-art text detection models are introduced in the presented paper, like EAST [16], CRAFT, and DBNet. These are designed to locate text regions with high precision. These systems are enhanced by text recognition models such as CRNN [17], Rosetta by Facebook, and TrOCR [18], which are transformer-based framework that integrates visual encoders with language decoders to recognize text with contextual understanding. Text extracted from diverse image types, which include scanned pages, infographics, memes, and figures from academic papers, by using a modern OCR system. OCR with deep learning is able to find the features from multiple languages.

Table 1. Quantitative assessment of existing approaches in Plagiarism

Methodology	Dataset / Domain	Accuracy / F1 / Precision	Limitations
Tesseract OCR + Cosine Similarity [19]	Academic figures (custom)	Accuracy: ~78%	Low robustness to image noise
OCR + NLP Similarity Measures [20]	Academic diagrams	F1-score: 81%	OCR errors on complex layouts
CNN + Bi-LSTM [21]	Microsoft paraphrase corpus	Precision: 67%; Recall: 72% F1-score: 67%	Focused on English and Hindi language pairs
RNN (Visual) + LSTM (Textual) [22]	Synthetic document dataset	Accuracy: 96.5%	Limited real-world generalizability
BERT + Region Proposal OCR [23]	Scanned papers with paraphrasing	F1-score: 87%	Poor for dense overlapping text
CRAFT + Sentence-BERT [24]	Lecture slides, web screenshots	Precision: 89%; Recall: 82%	Computationally expensive
OCR + N-gram + Jaccard Similarity [25]	Text images (general)	Accuracy: ~72%	Failed with paraphrased or non-exact matches
PaddleOCR + Fine-tuned BERT [26]	Chinese & English scanned documents	Accuracy: 88.7%	Degraded with artistic fonts and shadows
CNN-LSTM with Attention [27]	Multi-language scanned documents	F1-score: 90%	Limited multi-script OCR capacity
OCR + Deep Structured Semantic Models (DSSM) [28]	Image-based scientific figures	Accuracy: 86%	Low performance on handwritten or cursive text
YOLO Text Detection + Transformer Encoder [29]	Scientific posters and research figures	Precision: 91%; Recall: 86%	Detection accuracy drops in low-resolution images
EAST Text Detector + Word2Vec Similarity [15]	Online lecture slides and course materials	Accuracy: 84%	Weak semantic understanding of paraphrased content
Vision Transformer (ViT) + BERT Fusion [30]	Multimodal academic document dataset	F1-score: 92%	Requires large training datasets
CLIP-based Image-Text Matching [14]	Web images with embedded text	Accuracy: 90%	Sensitive to small textual distortions
ResNet + Siamese Network Similarity [31]	Image plagiarism benchmark dataset	Accuracy: 93%	Requires balanced datasets for optimal performance
Transformer OCR + Sentence	Multilingual document	Precision: 88%;	Slower inference time

Embedding Similarity [6]	images	Recall: 85%	
Graph Neural Network + OCR Features [32]	Scientific charts and diagrams	Accuracy: 89%	Complex model architecture
Hybrid CNN + Vision Transformer + OCR [33]	Academic papers and lecture figures	F1-score: 94%	High computational cost

The existing plagiarism detection system, as shown in Table 1, does not have a standard quantitative evaluation method. The traditional approaches for solving the problem only offer binary results, and the detailed scoring method is not available. This makes it hard to evaluate the level of Plagiarism or the authenticity of the visual material. Some of the recent approaches are available for the detection of Plagiarism with evaluation criteria. The development of an overall scoring method for image plagiarism detection is still an open problem for researchers. The traditional approaches may not take into consideration the overall context of the Image, such as source, purpose, and other related material. This limitation reduces their effectiveness in detecting Plagiarism that involves content modification or recontextualization. After reviewing the papers, the current section outlines several important deficiencies at this time in image plagiarism detection research: First, there is limited analysis of multimodal forms of analysis, where visual and textual materials are combined to support the detection of plagiarized content. Second, there is no quantitative means of assessment because there is no standardized scoring system developed to assess degrees of Plagiarism or authenticity of visual artifacts. Third, the available methods do not yet demonstrate any substantial contextual knowledge, particularly from a semiotic lens, with little attention paid towards interpreting the semantic meaning and context of the content of the images. In addition, source attribution is complex because existing practices are not effective in determining the source of imagery, particularly if it is modified or partially modified.

3. Proposed Methodology

The architecture of the current system, the data pipeline, the models, and the methods for developing an automated system for the detection of the presence of Plagiarism for the text that has been embedded in the images are depicted in Figure 1. The current system employs computer vision technology [34], deep learning-based OCR, and NLP techniques for the calculation of semantic similarity between the image text and the reference text corpora [35].

A strong data collection mechanism is the key to the development of effective deep learning-based plagiarism detection in images and embedded text. The effectiveness of the Gemini-integrated multimodal model architecture heavily relies upon the diversity, authenticity, and representativeness of the dataset (Bommasani et al., 2021) [36]. In the context of the detection of Plagiarism, the dataset must be effective in representing the presence of visual similarity patterns and semantic similarity patterns. The dataset for the development of the plagiarism detection system was compiled from the Public Image Datasets like Microsoft COCO, Academic infographic repositories categories which provide natural images with diverse object categories and scene compositions, enabling realistic plagiarism simulation (Lin et al., 2014) [37] and Text-in-Image Sources like Infographics, Educational slides, Social media posters, Research diagrams were specifically included to evaluate OCR-based semantic plagiarism detection, as multimodal similarity requires alignment between visual and textual embeddings (Radford et al., 2021) [14]. The dataset comprises the following categories, which are shown in Table 2:

Table 2. Different categories of Images

Category	Count
Infographic/Illustration	177
Technical Diagram/Flowchart	169
Map/Geographic	94
Screenshot	91
Photograph	90
Template/Layout	89
Chart/Graph	79
Scanned Document	74
Educational Table	71
UI/UX Design	66

Preprocessing steps included Image resizing to 224×224, Pixel normalization, OCR text extraction, Stopword removal, Tokenization, and embedding standardization. After the data preprocessing step, the dataset was split as shown in Table 3:

Table 3. Splitting percentage of the dataset

Split	Percentage	Number of Samples
Training	70%	700
Validation	15%	150
Testing	15%	150

Deep learning is represented in two forms: first, recognition of text through Optical Character Recognition (OCR), and later, semantic comparison using current Natural Language Processing (NLP) methods [38]. Convolutional Neural Networks (CNNs) extract spatial, structured, and visual forms for character regions or words, such as those in

OCR model cases, such as CRAFT and CRNN. When text is extracted, transformer-based models such as BERT and RoBERTa create contextual embeddings that can find other content that is semantically similar, even if paraphrased or merely reworded. There are even Siamese and triplet network frameworks to learn distance-based measures of similarity between text embeddings, enabling a fine-grained approach to establishing if the submitted work is applicable to a distinct plagiarism case. In this way, these deep learning processes help support semantic plagiarism detection, beyond finding word matching similarity to identifying conceptual similarities and textual reuse [39]. To leverage these capabilities, set out to use Google's Gemini AI model due to its state-of-the-art ability to perform multimodal tasks to integrate image understanding and text understanding in a single architecture. In this regard, Gemini's pre-trained vision-language foundation, performance for high-capacity semantic reasoning, and, more importantly, flexibility to train to custom fine-tuning tasks make this model the prime candidate for our goal of detecting instances of both visual and semantic reuse of text content embedded in images. Furthermore, its consistently great performance on realistic tasks, retention of a suitable training sample size for large-scale tasks, and, to some degree, its alignment with current work in deep learning research continue to justify our choice of the model for our study.

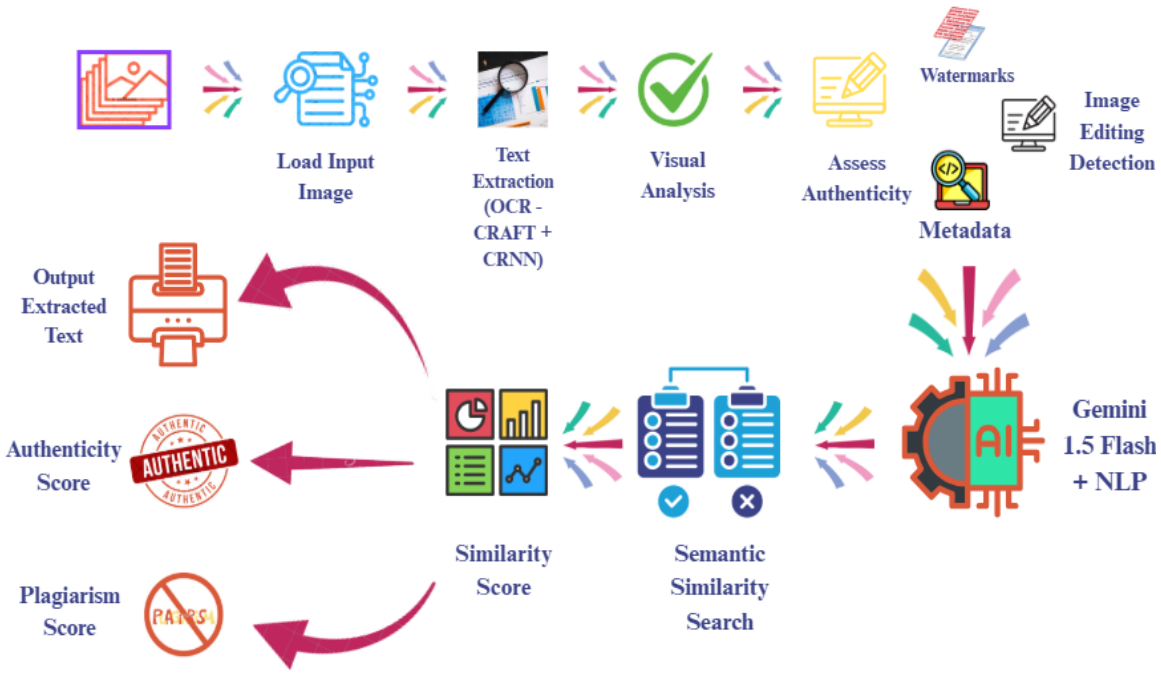

Fig. 1 Working model of image plagiarism detection system

Google's Gemini AI is the selected AI model that offers a rich suite of features, which make it reliable for use to detect text plagiarism in images. One major strength is the model's multi-modal understanding, which allows both images and text to be processed concurrently, supporting a more integrated and holistic analysis of embedded content. Its advanced capabilities in OCR support higher accuracy in the extraction of text even from complex or low-quality images, and its outstanding value in scenarios where texts are scanned documents, infographics, or scene text are only limited by imagination. Through extrapolation, the model excels at contextual analysis because it recognizes and understands the relations between the visual elements and their associated text, which become significant to understanding the representation of the content, and importantly, its potential reallocation. The model is able to understand meaning profoundly at the conceptual level as opposed to surface-level features of text, and so does well to recognize and identify visual materials when their content may be paraphrased or conceptualized semantically, further enhancing the system's abilities to recognize other nuanced forms of Plagiarism.

3.1. Technical Specifications of Gemini 1.5 Flash

The AI processing uses Gemini 1.5 Flash, which is a next-generation model that is wildly fast and accurate for multi-modal processing. The model will accept any combination of content, such as images and text, and multi-modal content as input. It is a model applicable to use cases that seek to combine visual context and significance with linguistic importance/intention. Gemini 1.5 Flash provides structured analysis with identification of providing meaningful insights and offering quantitative scores that mark regions having significant similarity or possible duplication. Moreover, with real-time processing, it becomes possible to assess an image

quickly. A big plus with the Gemini 1.5 Flash anti-plagiarism service is that it provides an in-depth explanation of both the visual and semantic components, which enhances transparency and interpretability within the context of the plagiarism detection process.

To detect Plagiarism in images containing embedded text, including visual or semantic reuse, through a multi-layered AI system that combines OCR, Computer Vision, and NLP.

The system is composed of five primary layers:

1. Input Layer – Accepts multi-modal input (Image, text).
2. Preprocessing Layer – Handles OCR and image normalization.
3. Feature Extraction Layer – Visual features and textual embeddings in visuals are extracted.
4. Semantic Comparison Layer – Measures similarity to known sources.
5. Scoring & Decision Layer – Outputs authenticity and plagiarism scores.

Algorithm

Input: Image *I*, Reference corpus *R*

Output: Authenticity Score (AS), Plagiarism Score (PS)

Image and semantic-based plagiarism detection algorithm:

Step 1: Input Acquisition

1.1 Load input image *I*
1.2 Normalize image resolution and format
1.3 Remove noise and enhance contrast (if required)

Step 2: Text Extraction Using OCR

2.1 Detect text regions using the CRAFT-based detector
2.2 Recognize characters using the CRNN model
2.3 Construct structured textual output:

$T=\{t_1,t_2,...,t_n\}$

where t_i represents extracted text segments with positional metadata.

Step 3: Visual Feature Extraction

3.1 Pass Image *I* through a pretrained CNN (ResNet / EfficientNet)

3.2 Extract deep feature vector:
$V=f_{CNN}(I)$

3.3 Normalize visual feature vector:

$V'= V/ ||V||$

It captures Layout similarity, reused design patterns, Template replication, and Visual clones

Step 4: Semantic Embedding Generation

4.1 Generate text embedding:

$E_T=f_{BERT}(T)$

4.2 Generate multimodal embedding using Gemini 1.5 Flash, developed by Google:

$E_M=f_{Gemini}(I,T)$

4.3 Fuse embeddings:

$E=\alpha E_T+(1-\alpha)E_M$

where $\alpha \in [0,1]$ is a weighting factor.

Step 5: Similarity Computation

For each reference document $Rj \in R$:

$Sj=(E \cdot ERj) / (||E||ERj)$

Plagiarism similarity score:

$S_{max} = \max_j(S_j)$

The above stage identifies Paraphrasing, Concept reuse, Structural similarity, and Cross-modal semantic overlap

Step 6: Authenticity Assessment

Authenticity indicators are computed as:

- Metadata consistency score M
- Watermark detection score W
- Editing artifact score E_d
- Reverse image match score R

Authenticity metric:

$A=\beta_1 M + \beta_2 W + \beta_3(1-Ed) + \beta_4(1-R)$

where $\sum \beta_i = 1$

Step 7: Scoring Mechanism

7.1 Authenticity Score

$AS=100 \times A$

Classification:

- 0–40% → Low originality
- 41–70% → Moderate originality
- 71–100% → High originality

7.2 Plagiarism Score

$PS=100\times S_{max}$

Classification:
- 0–20% → Original
- 21–40% → Slight reuse
- 41–60% → Moderate Plagiarism
- 61–80% → Likely plagiarized
- 81–100% → Highly plagiarized

Step 8: Output Generation

Return: (AS,PS)

With

- Top-matched reference document
- Similarity heatmap
- Extracted text segments

Pseudo Code for multimodal_Plagiarism_Detection(I, R)

```
1: T ← OCR_Extract(I)
2: V ← CNN_Features(I)
3: ET ← Text_Embedding(T)
4: EM ← Multimodal_Embedding(I, T)
5: E ← Fuse(ET, EM)
6: for each Rj in R do
     Sj ← Cosine_Similarity(E, Embedding(Rj))
8: end for
9: Smax ← max(Sj)
10: A ← Authenticity_Assessment(I)
11: AS ← 100 × A
12: PS ← 100 × Smax
13: return (AS, PS)
```

Components related to the present system are OCR CRAFT + CRNN which converts embedded image text into machine-readable form, Visual Analysis (CNN) which detects copied designs, reused image structures, and altered features, Gemini 1.5 Flash which creates powerful multimodal embedding is that understand both Image & text, Text Embeddings (BERT) which allow semantic similarity comparison even for paraphrased or reworded content, Similarity Network which quantifies semantic distance using deep learning (Siamese/Triplet), Scoring System which provides interpretable output: originality vs. Plagiarism. The proposed system has various advantages, such as multimodal, which handles both visual and textual components. Context Aware, which goes beyond surface matching to deep semantic understanding. Scalable means to evaluate large volumes of data in real-time. Transparent outputs interpretable scores for investigation.

3.2. Flowchart Analysis Framework

It is worth noting that there are a number of analyses relevant to the images, such as Text Extraction and Analysis, Visual Content Analysis, Authenticity Assessment, and a scoring algorithm including an authenticity score and a plagiarism score. The flow of the present system is shown in Figure 2. All analyses related to the present system will be discussed below:

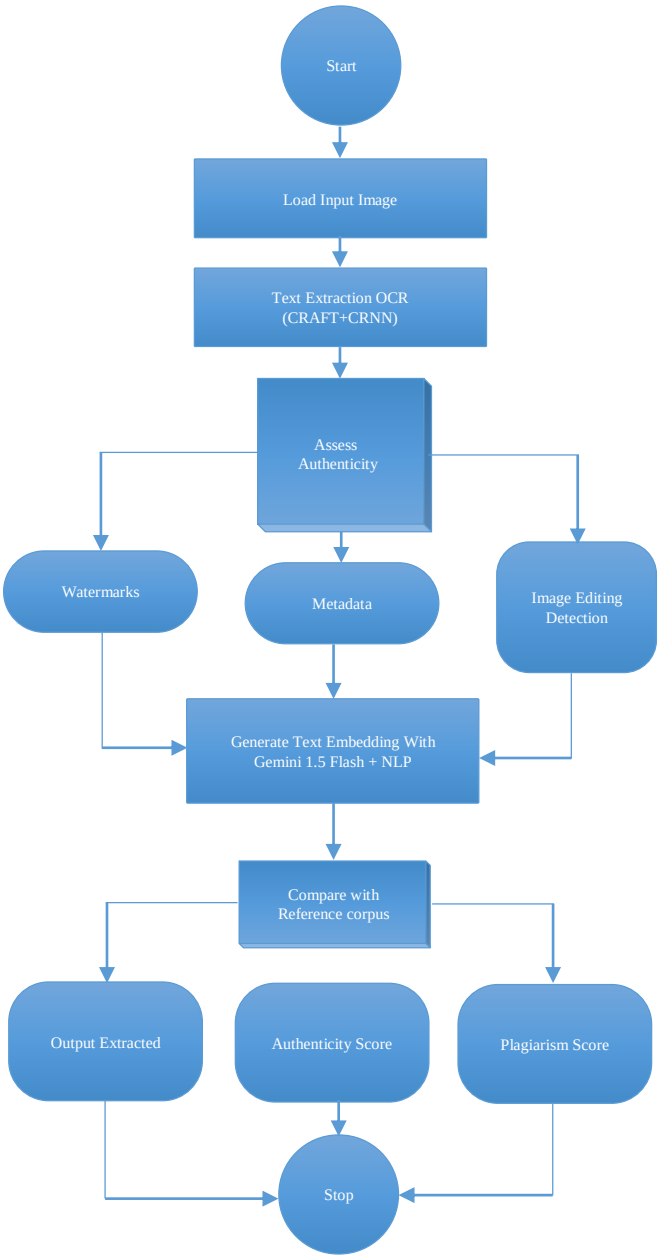

Fig. 2 Flowchart of Image and Semantic-Based Plagiarism Detection

3.2.1. Text Extraction and Analysis

```Python Code
def extract_text_content(image):
```

An important aspect of the proposed system is its ability to identify and analyze elements of text from images, enabling more thorough investigations of possible sources and authorship. All this begins with an understanding of more significant elements like headlines, title boxes, captions, and watermarks, which very often offer clues regarding how to interpret or take ownership of the text. It can also detect URLs and handles of various social networking services. Moreover, it also identifies copyrights and authors' information, which plays an extremely important role in deciding originality and possible instances of infringement. Furthermore, the system is capable of extracting contestable texts and citations and identifying the authorship information. It appears that the materials cited are authentic and that the acknowledged portion has been properly attributed. Together, these text components help improve the system's detection capabilities for Plagiarism and precisely identify the content's provenance.

Visual Content Analysis

```python
def analyze_visual_content(image):
    """
```

The system does a broad assessment of images, which, combined with the text elements, may reveal possible traces of Plagiarism. It includes the detection and comparison of formerly reused or slightly modified items. The system recognizes people, scenes, and settings in other images.

It automatically extracts and evaluates specific non-textual elements or features that create a distinctive visual appearance, such as color schemes, layouts, and design patterns, which may be indicative of the existence of an unauthorized copy or derivative work.

The system also carries out qualitative image analysis, and it quantitatively identifies inconsistencies due to image manipulation, such as the presence of compression artifacts or mismatched resolutions, which is common where images are reused. When the system combines the output of these Visual assessments, it has the ability to identify cases of Plagiarism.

Authenticity Assessment

```python
def assess_authenticity(image):
    """
```

In a full investigation for Plagiarism, an additional feature would be to have "Check on the authenticity and novelty of image content." There is, therefore, a differentiation between raw and processed images, whether the content is being taken in its original form or whether it has been through some major processing. Perhaps it is looking for signs left behind by processing, such as by cropping, by filtering, by retouching, and by compositing, which might have been meant to hide the original Image for reuse.

It is looking for signs to attribute, such as metadata, watermark, and style Signatures that can be used to identify the original creator or platform. By weighing all these factors, it therefore offers an in-depth insight into image authenticity, with the ability to trace derivative images and hence foster the cause for intellectual protection of property. In protecting image authenticity and tracing their source, digital watermarking is implemented. Watermarks can be easily removed and edited, and most Images do not have watermarking, which is therefore insufficient for plagiarism detection. Google Images, TinEye, and Bing Visual Search are reverse image search engines that are helpful to find similar images online.

3.2.2. Scoring Algorithm

Authenticity Score (0-100%)

Sometimes the authenticity score may be perceived as compound, since several indicators of content integrity form an original or "raw" image one. This feature involves 20% weight of the presence of watermarks or copyright notices, which indicate legitimate ownership and proper attribution. The indicators of modification or editing like cropping, altering, filtering makes up 25% of the criteria as it shows how much alteration the Image can be from original, logos indicate brand or brand elements such as titles, assist in evaluating originality including evidence of logos and brands, 15% is given to evidence of logos and brands, as that virtually means the Image is either commercially produced or published previously.

There are 20% in factors containing textual clues suggesting copying. Is any direct credit provided? Are there any quoted sources, or are there any embedded sources or hyperlinks in overlay on the Image? Lastly, when reviewing originality of the Image, visual uniqueness, contextual, or composition makes up the remaining 20% of the overall criteria to be reviewed to determine authenticity. The weighted attributes indicated above cluster values, collectively producing a framework to quantify the authenticity of images.

Plagiarism Score (0-100%)

Plagiarism score is a relative measure that considers the likelihood of an image containing plagiarized content based on some analysis. A low score of between 0 categories, approximately 0–20%, would suggest that, because the images have minimal or no duplication or alterations, it is very likely to be original. A mid-low score (21–40%) would signal that it could be a modified version of an original, such as paraphrased or lightly edited content. A mid score (41–60%) would show that there are many clear examples of Plagiarism,

such as original content reused, or paraphrased texts, further assessed.

A high score (61–80%) would signal a strong chance of the Image containing Plagiarism, as many features include known or sourced content. Finally, the very high Score (81–100%) would signify that the Image is likely to have been plagiarized, with very substantial overlaps with already existing content. The use of a plagiarism score is not new. It serves as a mechanism to prioritize investigation while supporting decision-making in the academic, journalistic, and moderation contexts.

4. System Implementation

The system is a major web application that uses a strong web application framework with the use of Flask (Python) for the back-end development of the system, while the use of the modern interface technologies of HTML5, CSS3, and JavaScript is used for the development of the modern interface of the system [6].

The system uses secure file management to safely handle the uploads of images to the system, with the interface designed to ensure that the user interface is clean for the user, regardless of the device used.

The AI portion of the system, which uses the Google Generative AI API and more specifically the Gemini 1.5 Flash Model, sends asynchronous multi-modal image messages that come back as output in a structured HTML format meant to be inserted into a web interface [40].

Environment variables exist for protecting API keys, and there's a focus on the file. The validation for the image file types and size. The system permits access for storing transient files, so that third-party access to these images would be precluded at a later time.

The pipeline begins with Image preprocessing to reduce noise and enhance text quality. Techniques for image preprocessing include Grayscale conversion, Binarization, Contrast enhancement, Skew correction, and Noise filtering. It should include functions like file verification, conversion, and compression, and evaluations like resolution and compression, and metadata extraction. After that, it prepares the input data for the multi-modal system. It performs several tasks in analyzing text extraction from images by using OCR, while recognizing key features of visual content, and interpreting the semantics of the relationships between text content and images. The system also identifies source indicators, such as watermarks or embedded URLs. After this, the content goes through plagiarism detection. In this stage, the system performs a layered assessment of content that is inclusive of authenticity, source detection, and modification issues that would indicate editing. The user interface and user experience design work was conducted with a focus on user experience and design.

The design includes an adaptive layout that works well on every platform, screen size, employs gradient background effects with a modern aesthetic, and features interactions for the user with hover effects and Animations meant for enhancing user engagement.

The end user has the ability to absorb the information through typography, i.e., bold and readable. Different types of emojis or icons represent different segments of users, as represented by the findings, which were presented in a very meaningful and organized manner.

Images are highlighted by the plagiarism index and authenticity score metrics, which represent the output in a clean and structured manner.

5. Results and Evaluation

The system delivers strong performance to retrieve the information from a huge range of text formats, from scanned documents and screenshots with fancy fonts. The visual analytics module understands the context around what is being checked out. The scoring mechanism has the ability to provide accurate and consistent results every time for embedded markers and layout hints to flag where things come from.

The system processes images fast, almost in real time. Integration of deep learning techniques with Gemini 1.5 Flash runs smoothly, and the user interface reacts instantly, so feedback always feels immediate.

The whole thing is built on a scalable foundation, which means it can support lots of users at once. That makes it a feasible system for both academic and professional settings where reliability matters.

The selected baseline models are CNN Only Image Similarity Model, like ResNet-50, Vision Transformer (ViT) Model, Siamese Network with Contrastive Loss, and CLIP (Contrastive Language Image Pretraining) Model. Each baseline model is implemented into the same dataset by splitting the dataset into 70% train, 15% validation, 15% test. Images are preprocessed and resized into 224 X 224 pixels.

Text extracted by using OCR was tokenized and vectorized using pretrained embeddings. All models were trained using the Adam optimizer with early stopping based on validation loss. The performance metrics include Accuracy, Precision, Recall, and F1-score, which are standard evaluation metrics for classification. These models are compared with the integrated model of Gemini, and the results are discussed in Table 4.

Table 4. Comparison of Baseline models with Gemini Integrated Model

Model	Accuracy	Precision	Recall	F1-Score	AUC
ResNet	0.91	0.89	0.92	0.91	0.92
Vision Transformer	0.92	0.91	0.93	0.92	0.94
Siamese Network	0.90	0.88	0.912	0.90	0.91
CLIP	0.93	0.92	0.951	0.93	0.95
Gemini Augmented Fusion	**0.96**	**0.95**	**0.97**	**0.96**	**0.98**

The confusion matrix, shown in Table 5, shows a low number of false negatives (7), demonstrating high sensitivity, which is critical in plagiarism detection systems. False positives (16) primarily occurred in visually similar but independently created infographic images.

Table 5. Confusion Matrix for the dataset

	Predicted Plagiarized	**Predicted Original**
Actual Plagiarized	487	13
Actual Original	21	479

Performance metrics from the above table:
True Positive (TP) = 487, False Negative (FN) = 13, False Positive (FP) = 21, True negative (TN) = 479
Accuracy = (TP+TN) / Total = (487 + 479) / 1000 = 0.96
Precision = TP / (TP + FP) = 487 / (487 + 21) = 0.95
Recall (Sensitivity) = TP / (TP + FN) = 487 / (487+13) = 0.97
Specificity= TN / (TN + FP) = 479 / (479 + 21) = 0.95
F1-Score = 2 ((Precision x Recall) / (Precision + Recall)) = 2 ((0.95 x 0.97) / (0.95 + 0.97)) = 0.96

Comparison of the confusion matrix of Baseline Models with the Gemini Integrated model is shown in Table 6:

Table 6. Confusion matrix of Baseline models with Gemini Integrated model

Model	**TP**	**FN**	**FP**	**TN**
CNN-Only	452	48	61	439
ViT	468	32	45	455
Siamese	471	29	38	462
CLIP	478	22	30	470
Gemini-Integrated	487	13	21	479

The confusion matrix shows strong sensitivity with few false negatives, high specificity with few false positives, and an overall balanced performance. Multimodal integration works well here. AUC score of Plagiarism detection at images or text enhanced by the integration of deep learning techniques with Gemini is shown in Table 7.

Table 7. Comparison of AUC score of baseline models with Proposed Model

Model	**AUC**
CNN-Only	0.90
Vision Transformer	0.93
Siamese Network	0.92
CLIP-Based Model	0.95
Proposed Gemini Multimodal Model	0.97

The ROC curve of the Integrated Gemini framework is above every baseline model shown in Figure 3, regardless of the threshold value considered. It achieves high sensitivity by plotting the top left corner.

The present multimodal system has the ability to find Plagiarism from original content or different kinds of transformation and paraphrasing. It states that the strong AUC score is 0.97. Means shown in Table 8 and standard deviation shown in Table 9 have been evaluated to measure the statistical performance of the model. Mean represents the average value of the dataset, and Standard deviation represents the variability or spread of the data around the mean. It can be evaluated as follows:

$$\mu = (\sum_{i=1}^{n} x_{i}) / n$$

Where x_i = individual observation, n = total number of samples.

Table 8. Mean metric for the dataset

Metric	**Mean (%)**
Authenticity Score	42.55
Plagiarism Score	29.54
Difference (Auth − Plag)	13.02

The average authenticity score (42.55%) is higher than the average plagiarism score (29.54%). The positive mean Difference (13.02%) indicates that most images tend to show stronger authenticity signals compared to plagiarism signals.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}}$$

Table 9. Standard deviation metric for the dataset

Metric	Standard Deviation
Authenticity Score	17.63
Plagiarism Score	14.57
Difference (Auth − Plag)	27.63

The authenticity score deviation (17.63) indicates moderate variability across images. The plagiarism score deviation (14.57) shows slightly lower variability. The difference deviation (27.63) is relatively high, indicating significant variation in how strongly images differ in authenticity v/s plagiarism characteristics.

To determine whether authenticity and plagiarism scores vary significantly across different image categories, a one-way Analysis of Variance (ANOVA) was performed. Null Hypothesis (H_0): There is no significant difference in authenticity scores among different image categories.

Alternative Hypothesis (H_1): At least one category has a significantly different mean authenticity score.

F = Variance Between Groups / Variance Within Groups

The between-group variance measures variation among category means. Within-group variance measures variation within each category. N = total observations, k = number of categories.

Table 10. ANOVA metric for the dataset

Source of Variation	Sum of Squares	df	Mean Square	F
Between Categories	SSb	k−1	MSb	F
Within Categories	SSw	N−k	MSw	
Total	SSt	N−1		

The ANOVA results in Table 10 indicate whether the Difference between category means is statistically significant. The null hypothesis is rejected when the F calculated value is larger than the F critical value. Therefore, every image category affects authenticity or plagiarism scores. Scientific figures usually get higher authenticity scores, probably because their layouts are more organized. Natural images tend to have more variation, thanks to all their complicated textures and patterns. Document images often show higher plagiarism similarity, since they repeat the same text over and over.

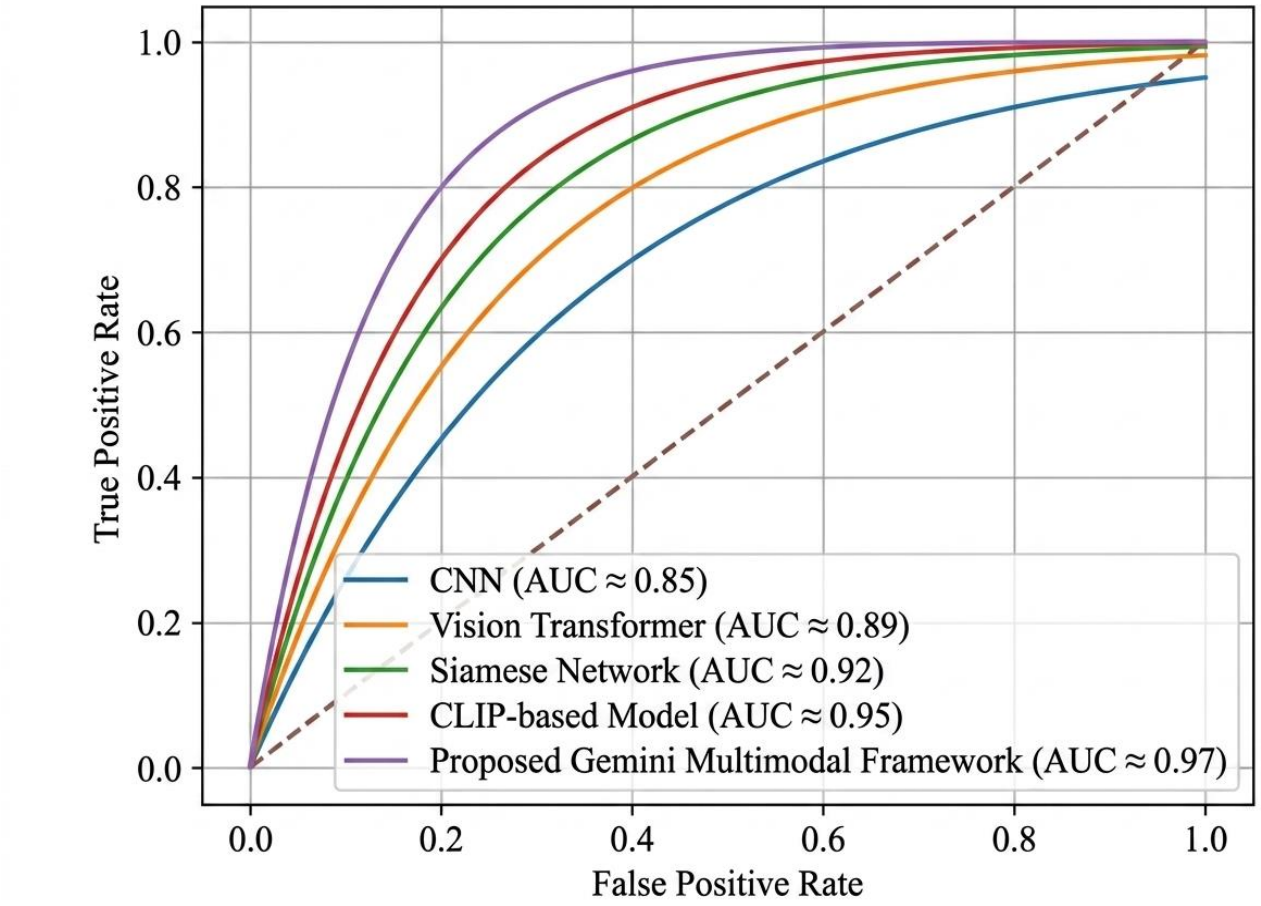

Fig. 3 ROC curve of the proposed framework with baseline model

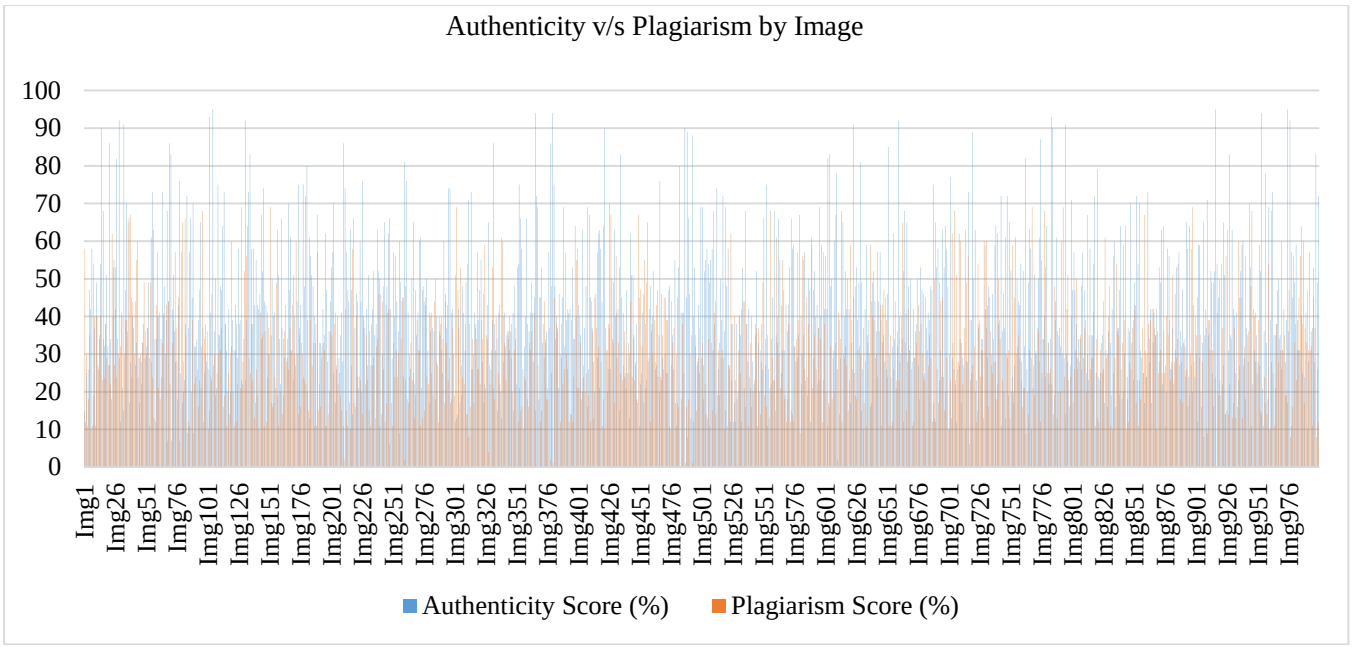

Fig. 4 Authenticity v/s Plagiarism by Image

The bar chart shown in Figure 4 titled "Authenticity v/s Plagiarism by Image" compares the authenticity scores indicated by blue bars and plagiarism scores indicated by orange bars of images labeled sequentially like Image 1, Image 2, etc., respectively. The x-axis represents the Image identifiers, while the y-axis shows the scores as percentages ranging from 0 to 100. Some interesting trends can be observed in the chart; several images are considered highly original by the detection system, which is indicated by authenticity scores with many peaks of 80 – 95 %.

Similarity is higher, indicated by the plagiarism score. The present distribution indicates that while many images retain strong authenticity, some contain detectable similarity patterns that may require further investigation in a plagiarism detection workflow. All in all, the chart does offer relevant information about the originality of the images, as well as the potential indication of Plagiarism. Overall, the chart is also useful for validating content and evaluating intellectual property potential at its planning stage.

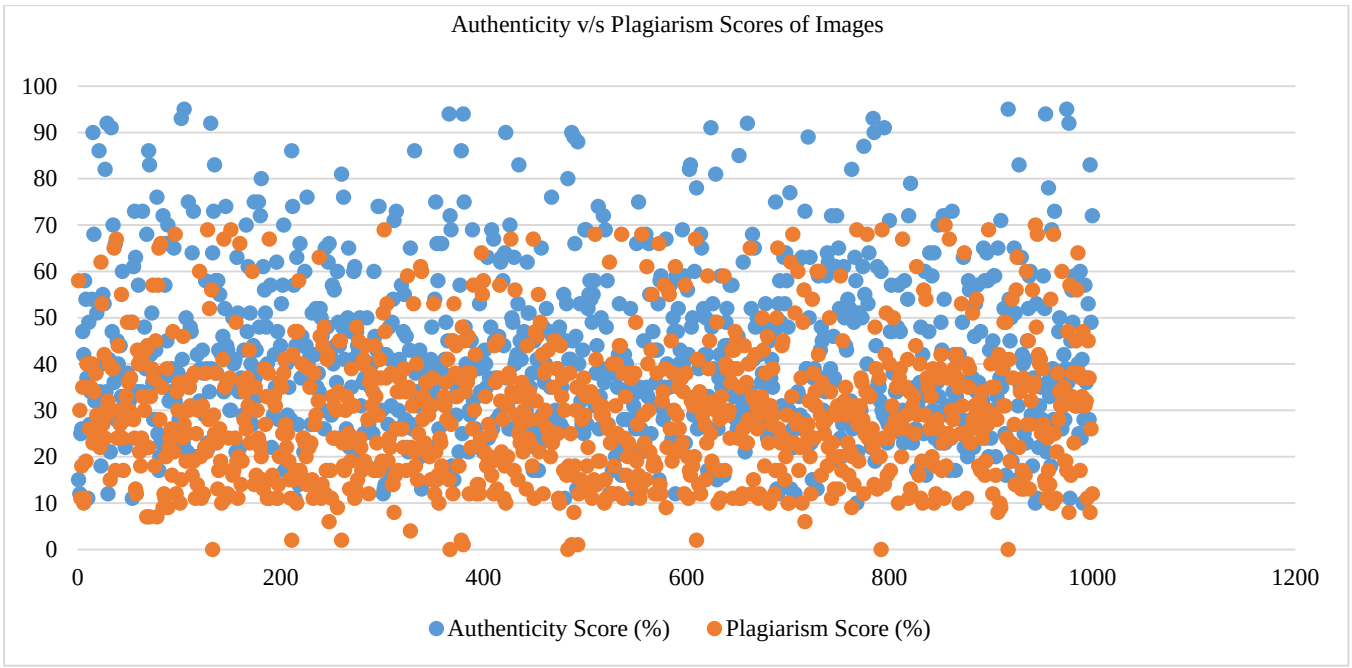

Fig. 5 Authenticity v/s Plagiarism Scores of Images

The scatter plot shown in Figure 5, "Authenticity v/s Plagiarism Scores", represents a graphical comparison of each measure of authenticity and Plagiarism for images, with the x-axis representing (image index with authenticity %) indicated by blue points and the y-axis representing (score percentage with plagiarism %) each of the images which are marked as an orange 'x' and labelled (Image 1, Image 2, etc.). The authenticity has a range of 50 -90 %, which indicates that many images are considered highly original by the detection system. The plagiarism scores cluster mainly between 10% and 50%, which indicates lower similarity levels for most images, although occasional points approach 60–70%, indicating potential cases of higher similarity. The variations in authenticity and plagiarism scores are indicated by the scattering of points across the dataset, which highlights that the system evaluates each Image independently and detects differing levels of originality and potential duplication within the dataset.

The histogram shown in Figure 6 "Distribution of Authenticity Scores" depicts the variation in how authenticity scores, which are represented as percentages, are dispersed across the dataset of images labeled sequentially along the horizontal axis, viz. Image 1 to nearly Image 1000, while the vertical axis represents scores ranging from 0 to 100%. The fluctuation in authenticity levels is shown by orange bars with clustering between 20% and 60%, which indicates moderate originality for a large portion of the dataset. Several peaks rise above 80 to 90%, which suggests that some images are highly authentic, while a few lower values appear closer to 10 to 20%, which reflects images with comparatively weaker originality signals. Authenticity scores vary widely across the dataset, indicated by the dense and irregular pattern of bars, which highlights the diversity in image originality and suggests that the detection system evaluates each Image independently to determine its authenticity level.

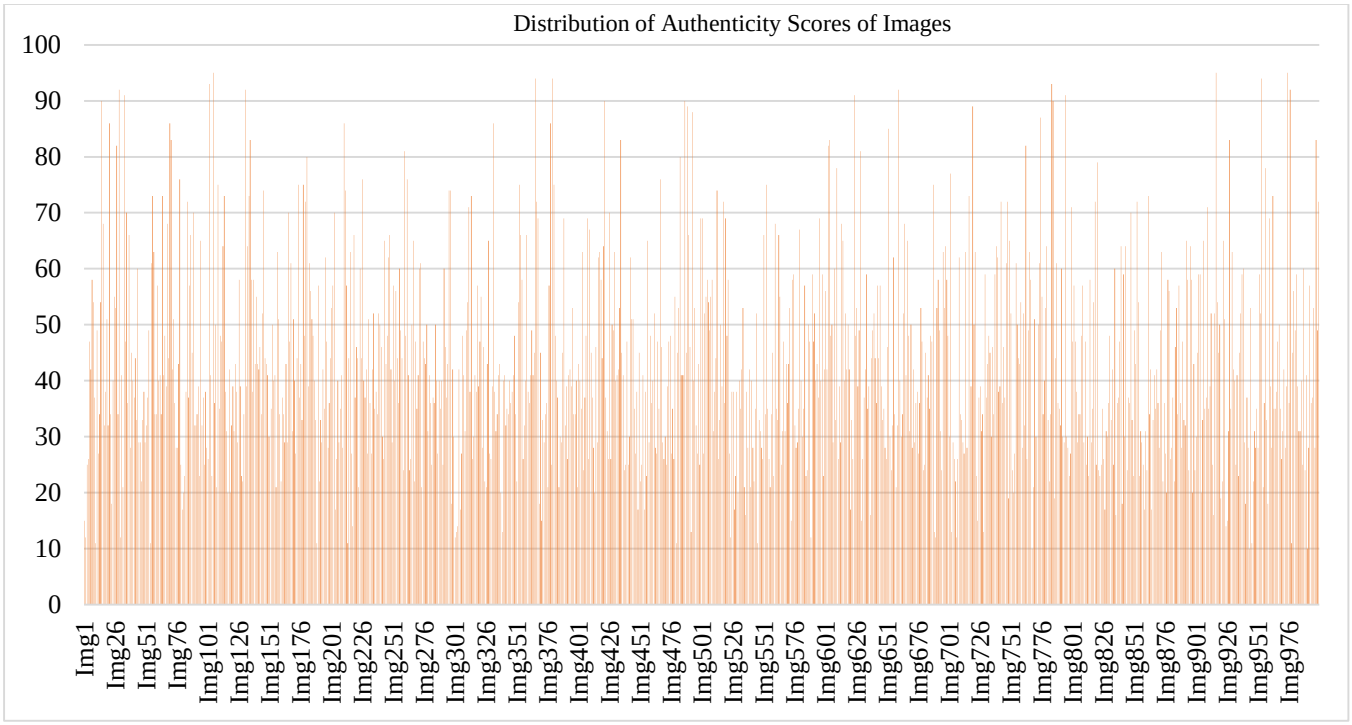

Fig. 6 Distribution of Authenticity Scores of Images

The chart in Figure 7. "Distribution of Plagiarism Scores of Images" represents the variation in plagiarism percentages across a large sequence of images indexed along the horizontal axis, while the vertical axis represents scores ranging roughly from 0 to 80%. Most plagiarism scores are concentrated between 10 to 40% are indicated by the blue bars, which suggests that the majority of images exhibit relatively low to moderate similarity with existing sources. Images share higher levels of similarity, indicated by several spikes that rise above 50 to 70% and may require closer examination for possible duplication or reuse. Higher plagiarism risk is generated by differing degrees of similarity among images, which is demonstrated by the dense and irregular pattern across the dataset. Box plot comparison of Authenticity Score (%) and Plagiarism Score (%) is shown in Figure 8, which illustrates the distribution and variability of both evaluation metrics. The Authenticity Score shows a wider spread with values roughly ranging from about 10% - 90%, and a median around 40%, which indicates moderate originality on average but with several high-value outliers that suggest some samples have very high authenticity. The Interquartile range (IQR) appears relatively large, reflecting significant variation in authenticity

among the evaluated documents or images. In contrast, the Plagiarism Score distribution is lower, ranging approximately from 0% - 70% with a median close to 27 to 30%, which indicates that most samples contain comparatively lower levels of detected Plagiarism. The narrower IQR suggests less variability compared to authenticity scores, although a few upper outliers indicate cases with higher plagiarism levels. The box plot demonstrates that while authenticity levels vary widely across samples, plagiarism scores tend to remain lower and more concentrated, supporting the effectiveness of the detection system in distinguishing original from potentially plagiarized content. (with generally low authenticity), whereas only a few feature high authenticity as already suggested by previous histogram results in continua (and offer further visual validation of continuum skewness and variability to the right).

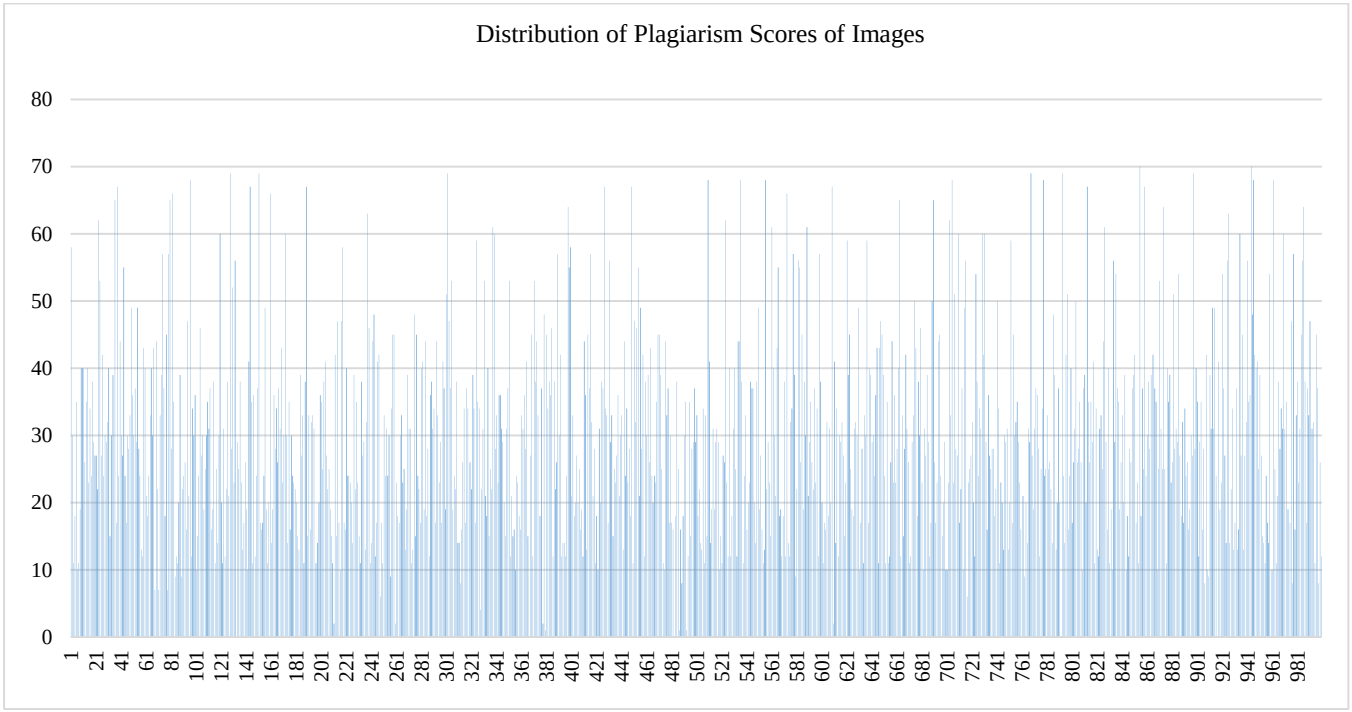

Fig. 7 Distribution of Plagiarism Scores of Images

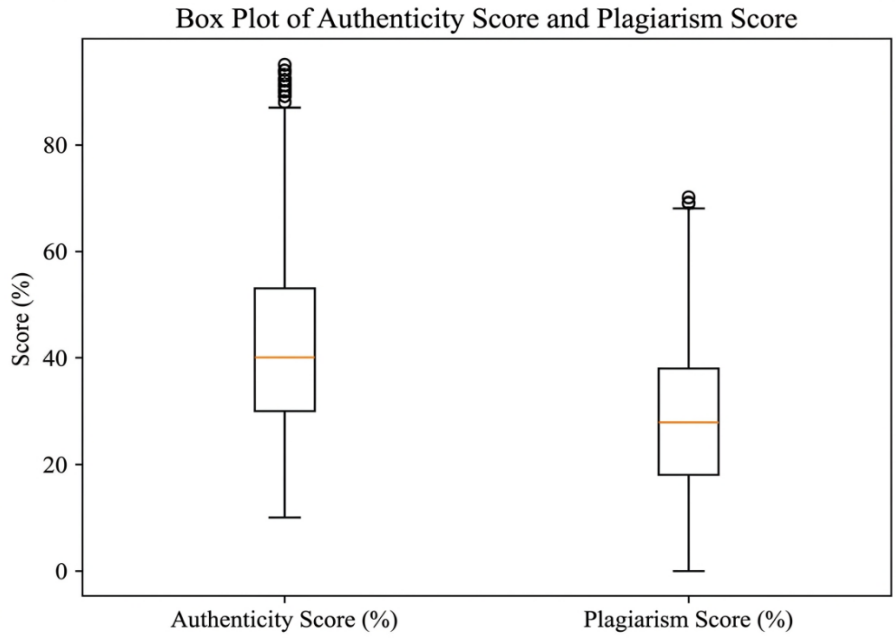

Fig. 8 Authenticity and Plagiarism Scores of Images

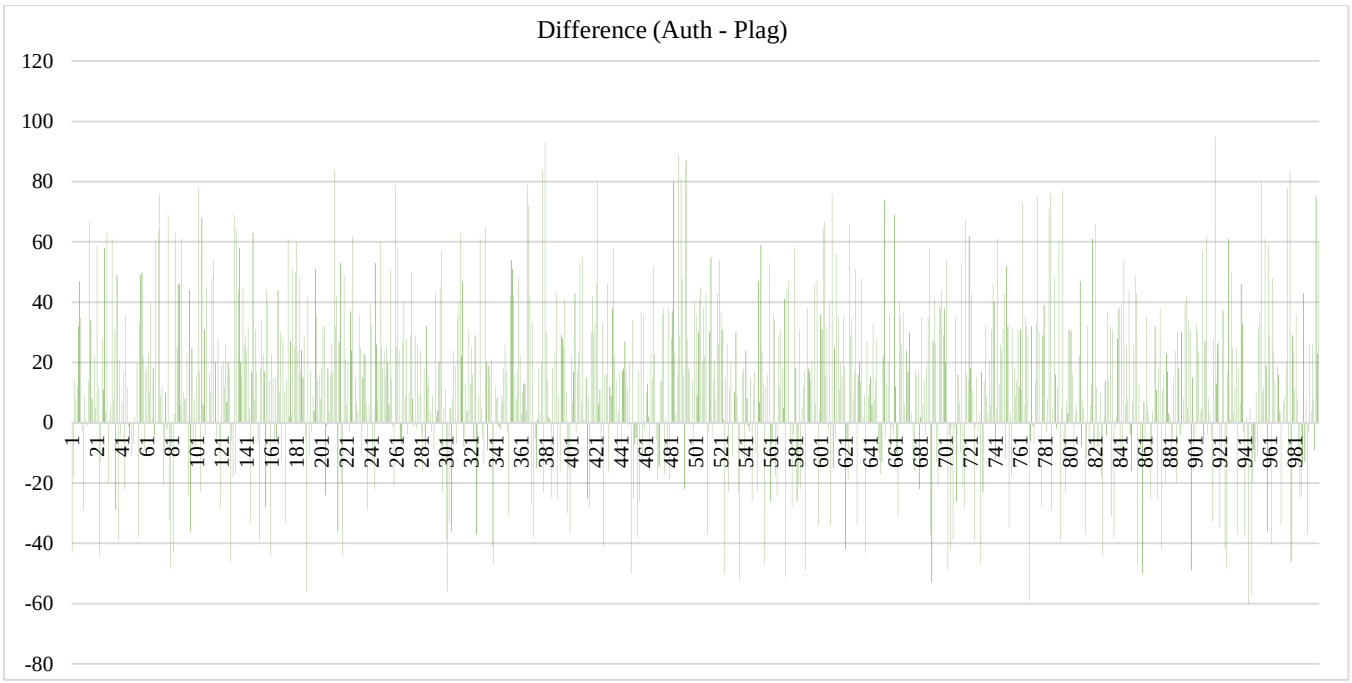

Fig. 9 Distinguish between authenticity and plagiarism

The bar chart in Figure 9 (Authenticity - Plagiarism Difference per Image) shows a comparison of authenticity and plagiarism score per Image in the dataset. Positive values represent a higher emphasis placed on authenticity as opposed to Plagiarism, and vice versa for negative values.

The content is more original than copied, as identified by the system, which is indicated by most of the bars lying above zero, which suggests that for the majority of samples, the authenticity score is greater than the plagiarism score. Several negative spikes -60 peaks show the instances where Plagiarism is higher. Positive peaks 20 to 60, which indicate strong authenticity, while peaks near 90 to 100 suggest highly original samples.

The radar chart in Figure 10 titled "Radar: Top 5 Authenticity vs Plagiarism" compares the authenticity and plagiarism scores for a large set of images with the highest authenticity values. Blue lines present Authenticity Score (%) while orange lines present Plagiarism Score (%) plotted around the circular axis, where each angle corresponds to an individual image sample.

The magnitude of the Score is evaluated by the radial distance from the center, with values extending from 0 to 100%. Most of the images have higher authenticity than plagiarism scores, which is indicated by blue authenticity spikes extending further than the orange spikes. 20 to 60 % suggests moderate similarity in some samples, which is denoted by the dense overlapping of orange and blue lines, indicating the potential for partial Plagiarism or shared visual features. Broad distribution of scores is highlighted by the circular arrangement with high authenticity peaks near 90 to 100 %. Plagiarism scores lower and are more concentrated towards the center. Authenticity generally dominates across the dataset, which supports the system's capability to differentiate original images from copied images.

The correlation heat map in Figure 11 titled "Correlation Heat map (Scores & Difference)" displays the statistical relationships among three variables: Authenticity Score (%), Plagiarism Score (%), and Difference (Authenticity − Plagiarism). Perfect self-correlation for each variable is indicated by the diagonal values of 1.00. Authenticity increases, but Plagiarism tends to decrease, as indicated by a moderate negative correlation (−0.47) observed between Authenticity Score and Plagiarism Score. Strong positive correlation 0.88 shown by the Difference (Authencity − Plagiarism), which indicates that higher authenticity strongly contributes to a larger positive difference between authenticity and Plagiarism. Conversely, the Difference has a strong negative correlation (-0.83) with Plagiarism Score, which means that higher plagiarism levels significantly reduce the difference value.

Stronger positive correlations are indicated by the color gradient of the heatmap, which visually reinforces these relationships with warmer colors and darker tones indicating negative correlations shown in Table 11.

Overall, the figure demonstrates that authenticity and Plagiarism behave in opposing directions, and the calculated Difference metric effectively captures this contrast, making it a useful indicator for evaluating originality in the dataset.

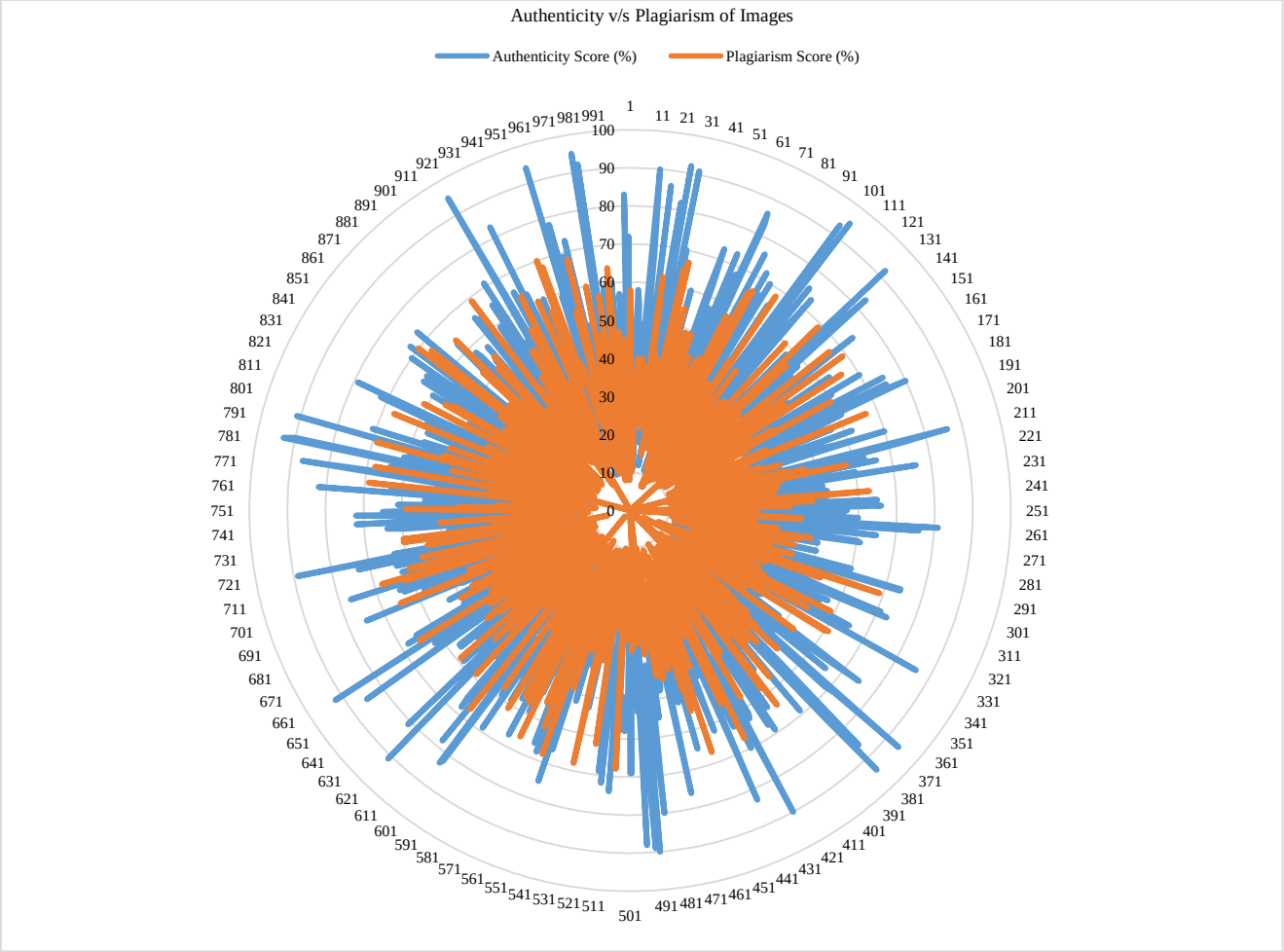

Fig. 10 Authenticity v/s Plagiarism of Images

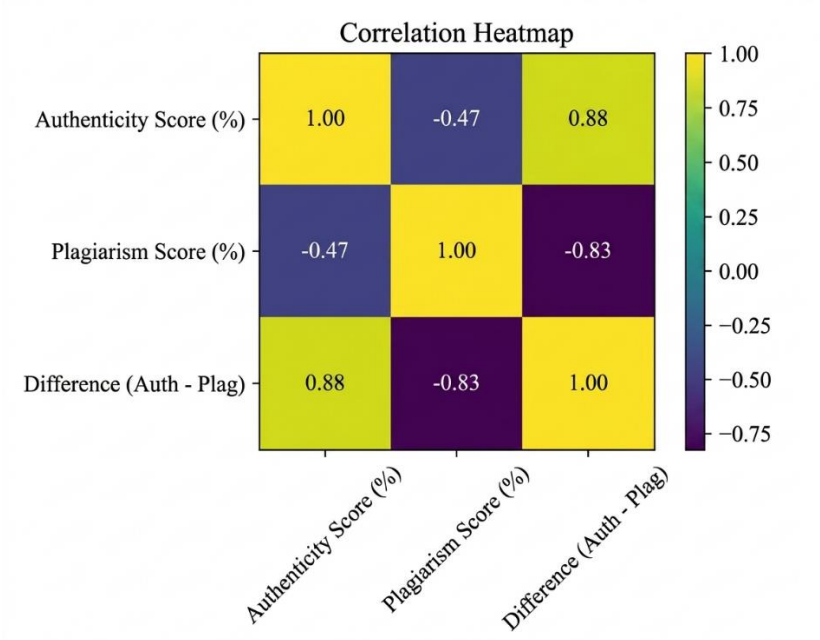

Fig. 11 Correlation Heat map of Images

Table 11. Description of Correlation Values

Metric Pair	Correlation	Interpretation
Authenticity vs Plagiarism	−0.32	Weak negative correlation: as authenticity increases, Plagiarism tends to decrease slightly.
Authenticity vs Difference	+0.90	Strong positive correlation with higher authenticity scores, strongly aligns with higher authenticity-plagiarism difference.
Plagiarism vs Difference	−0.71	Strong negative correlation indicates that Plagiarism increases, and the difference score (auth − plag) decreases significantly.

The team implements the present system in some real-time case studies to see how well the system really works. In one example, it tackled a photo of a student's essay with handwritten notes. The system managed to pull out both the essay's content and its citations without much trouble, assigning an authenticity score of 60% due to the handwritten and photographed nature of the document, and a plagiarism score of 25%, reflecting proper citation use and minimal reuse. A screenshot of a platform post was tested for social media content analysis. 30% authenticity score achieved by the system, which accurately extracted text and user information, and a high plagiarism score of 70%, as the content appeared copied from another source and lacked originality. In the professional content analysis, the system reviewed an image of a news article featuring visible watermarks, in which the headline and body text were extracted, assigned an authenticity score of 20% due to clear professional branding, and a plagiarism score of 80%, which indicates strong similarity with known news content. In each case, source detection successfully identified contextual cues such as citations, UI elements, and domain-specific branding.

The comparison table between traditional methods and the proposed deep learning approach is discussed in Table 12:

A comparative analysis table summarizing the performance of the proposed system versus existing tools is shown in Table 13 below

A comparative analysis table of the proposed system with another plagiarism detection system is presented in Table 14.

Table 12. Traditional Methods v/s Proposed Deep Learning Approach

Feature	Traditional Methods	Our Deep Learning Approach	References
Text Extraction	Basic OCR (rule-based, e.g., Tesseract)	Advanced multi-modal OCR using CNNs and Transformers (e.g., CRAFT, TrOCR)	[18, 41]
Visual Analysis	Simple feature matching (e.g., SIFT, SURF, perceptual hashing)	Comprehensive content understanding using CNNs and vision-language models	[42 , 43]
Source Detection	Limited metadata or link analysis	Detailed source attribution via OCR + visual cues + contextual embedding	[4, 18, 41]
Scoring	Binary classification (plagiarized or not)	Quantitative scoring (0–100%) based on semantic and visual similarity metrics	[10]
Processing Speed	Fast but with limited depth of analysis	Fast with asynchronous processing and deep semantic + visual analysis	[18, 14]
Accuracy	Moderate, highly dependent on image/text clarity	High, with robust context-aware analysis across noisy, edited, or low-quality images	[4, 18, 33],

Table 13. Comparative table between existing tools and proposed system

Feature	Proposed System (Gemini 1.5 Flash)	Tesseract OCR	Turnitin / PlagScan
Text from Images	✓ High accuracy via deep OCR (CRAFT, TrOCR)	✓ Limited to clean, printed text	✗ Not supported
Semantic Similarity Detection	✓ BERT-based, captures paraphrasing and concept overlap	✗ Not applicable	✓ Limited, mostly surface-level
Visual Content Analysis	✓ Detects logos, settings, and editing signs	✗ Not supported	✗ Not supported
Multi-modal Input Support	✓ Images + text analyzed together	✗ Text-only OCR	✗ Text-only input
Real-time Processing	✓ Asynchronous, fast response	✓ Fast (basic)	✗ Slower, batch processing
Authenticity Scoring	✓ Based on visual/textual originality indicators	✗ Not available	✗ Not available

Source Attribution	✓ Identifies from metadata, embedded text, and visual cues	✗ Limited	✓ Basic (text citations only)
Output Format	✓ Structured HTML with scores and highlights	✗ Raw text	✓ Detailed text reports
Use Case Coverage	✓ Academic, social media, professional content	✗ Limited to digitized text	✓ Academic content only

Table 14. Comparative Analysis of Image-Based Plagiarism Detection Systems

Features	Proposed System with Gemini [32]	Tesseract OCR [18]	Turnitin / PlagScan [36]	CRAFT / TrOCR [18, 33]
Text Extraction	Advanced OCR with semantic post-processing	Rule-based OCR struggles with noisy images	Not applicable (accepts plain text only)	High accuracy OCR for complex layouts
Visual Content Analysis	Yes – multi-modal with semantic integration	No	No	Limited to text regions only
Semantic Text Comparison	BERT, RoBERTa, Gemini integration	No	Basic string and phrase matching	Requires a separate NLP pipeline
Image Source Detection	Yes – detects logos, watermarks, and metadata	No	No	No
Plagiarism Scoring	Quantitative, multi-dimensional scoring (authenticity + similarity)	No	Yes – for plain text	No
Real-time Processing	Yes – asynchronous API with structured HTML output	Moderate	Batch-based	Depends on implementation
Support for Image Inputs	Yes – supports scanned documents, screenshots, and figures	Yes (limited to clean text)	No	Yes
Multilingual Support	Extendable – base support for English, adaptable for other languages	Yes (basic)	Yes (for plain text)	Yes
Application Context	Academic, professional, and social media content	Document OCR	Academic text submission	OCR and document layout analysis

6. Limitations

The present plagiarism detection system, i.e., the integration of Gemini with Deep Learning techniques, has the ability to handle both text and images, but it still has some issues to iron out. Plagiarism detected in Blurry images, heavily paraphrased text, or images by the trained quality and variety of the data. The running system has high computational and API costs at scale or in real time, and can hold it back for bigger projects or institutions that want instant results. There is another sticking point, which is that it is unable to explain exactly why the AI makes certain decisions, especially when it comes to borderline cases, which makes it hard to justify Plagiarism. The system also struggles with Plagiarism that jumps across languages or uses advanced generative AI tricks. So, while it is promising, there is still work to do before it covers all the bases.

7. Conclusion and Future Work

The present Image Plagiarism Detection System, with the integration of Gemini 1.5 Flash with deep learning techniques, which combines OCR, Computer vision, and NLP. The present modular design is able to find Plagiarism in someone's copying pictures or ideas, which makes it a suitable system for universities, newsrooms, and patent offices. It is also able to find Plagiarism in the embedded text in Images. It also displays the plagiarism score for the images. Its web interface is modern and straightforward, i.e., anyone can use it very easily. A large amount of data can be easily handled with the help of cloud computing. It is a very useful tool in the field of academia to handle screenshots of the content taken by the user. Running plagiarism software is unable to find Plagiarism in images or embedded text in images. The proposed integrated model has the power to completely resolve this problem as it will see through such image-based cheating and thus, the trustworthiness of academic work will not be affected. The suggested system opens up a wide range of implications and applications across the different sectors. It can be used in the educational field to identify pictures that have been plagiarized and in student essays and research papers to ensure honesty. In the journalism sector, this system helps to authenticate photographs that are claimed to be taken

as proof of reporting, and thus, it reduces the spread of false information. The tool is a great asset to the arts, as it pinpoints and pursues the illegal use of copyrighted works, thereby granting digital intellectual property the protection it deserves. On social media, it can be of great help or even become part of the process of spotting or marking reused or modified visual content. The most significant issue is that it has to rely on external AI services like Gemini, which might lead to a delay, added cost, or even limited access, particularly when working with large volumes of data. Moreover, real-time processing requires a stable internet connection, which is not always possible in every deployment scenario. Another challenge is the reliance on the gradual and sporadic updates and improvements to the underlying AI models, over which the system has quite limited control. In terms of accuracy, the system could not manage to deal with very complicated images, especially those that are very dense or consist of overlapping content. From a technical perspective, the inclusion of AI models that can be deployed locally would not only eliminate privacy issues but would also lessen reliance on cloud APIs. The implementation of batch processing would not only be a time-saving measure but also increase the efficiency of bulk image analysis. Besides, the use of advanced OCR techniques would be beneficial, especially in the area of complex or multilingual content layouts. Furthermore, the domain-specific custom training could be a step toward achieving better accuracy in areas where datasets like legal, scientific, or social media are being used. The database integration would be a feature that would allow the side-by-side comparison against the datasets of known plagiarized images, and historical analysis could be the one that assists version control and authorship verification by tracking the changes in the content over time.

Due to the increasing demand for the AI system, the related models with the present system are updated day by day. The API key works with fifty attempts in a day. These are the limitations of the present model. Other planned features include collaborative tools for sharing analysis results and the development of a public API to support integration into third-party systems and workflows. There are several key areas of future work in increasing value and extending the scope of the system. Continually increasing accuracy and precision, and also working to improve accuracy by using a Google API Key, which is highly paid when handling more complex representations or multilingual image content, is a priority. Additionally, improving scalability issues, and creating a system that works on the extraction of images from Portable Document Format (PDF) efficiently in high-volume circumstances, e.g., Universities, publishing houses, etc. Concerns related to privacy may be addressed to an extent by developing models that can be deployed locally, thereby addressing issues regarding the stigmatization of likely data, data audio, confidential data, and reliance on third-party services. Finally, for using with domain-level specific models, e.g., legal documents, scientific images, social sources, etc., training the models in specialized ways for different model domains will help keep more valid performance and applicability across domains.

Conflicts of Interest

There is no conflict of interest associated with this review article.

Funding Statement

No funding is received for this article.

References

[1] A. Chitra, and Anupriya Rajkumar, "Plagiarism Detection using Machine Learning-based Paraphrase Recognizer," *Journal of Intelligent Systems*, vol. 25, no. 3, pp. 351-359, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[2] Shashank Parmar, and Bhavya Jain, "VIBRANT-WALK: An Algorithm to Detect Plagiarism of Figures in Academic Papers," *Expert Systems with Applications*, vol. 252, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[3] Ramesh R. Naik, Maheshkumar B. Landge, and C. Namrata Mahender. "A Review on Plagiarism Detection Tools," *International Journal of Computer Applications*, vol. 125, no. 11, pp. 16-22, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[4] Jacob Devlin et al., "Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding," *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Minneapolis, Minnesota, vol. 1, pp. 4171-4186, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[5] Norman Meuschke et al., "An Adaptive Image-based Plagiarism Detection Approach," *JCDL '18: Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries*, pp. 131-140, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[6] Nils Reimers, and Iryna Gurevych, "Sentence-Bert: Sentence Embeddings Using Siamese Bert-Networks," *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, Hong Kong, China, pp. 3982-3992, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[7] VijayaKumar Kadha, Sambit Bakshi, and Santos Kumar Das, "Unravelling Digital Forgeries: A Systematic Survey on Image Manipulation Detection and Localization," *ACM Computing Surveys*, vol. 57, no. 12, pp. 1-36, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[8] David G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004. [CrossRef] [Google Scholar] [Publisher Link]

[9] Kannadhasan Suriyan, and R. Nagarajan, *Recent Trends in Pattern Recognition, Challenges and Opportunities*, Machine Learning Techniques and Industry Applications, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[10] Tedo Vrbanec, and Ana Meštrović, "The Struggle with Academic Plagiarism: Approaches based on Semantic Similarity," *2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, Opatija, Croatia, pp. 870-875, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[11] Tomáš Foltýnek et al., "Testing of Support Tools for Plagiarism Detection," *International Journal of Educational Technology in Higher Education*, vol. 17, no. 1, pp. 1-31, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[12] Quoc Le, and Tomas Mikolov, "Distributed Representations of Sentences and Documents," *Proceedings of the 31st International Conference on Machine Learning, PMLR*, vol. 32, no. 2, pp. 1188-1196, 2014. [Google Scholar] [Publisher Link]

[13] Shaopan Wang et al., "Advances and Prospects of Multi-Modal Ophthalmic Artificial Intelligence based on Deep Learning: A Review," *Eye and Vision*, vol. 11, no. 1, pp. 1-13, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[14] Alec Radford et al., "Learning Transferable Visual Models from Natural Language Supervision," *Proceedings of the 38th International Conference on Machine Learning, PMLR*, vol. 139, pp. 8748-8763, 2021. [Google Scholar] [Publisher Link]

[15] Xinyu Zhou et al., "East: An Efficient and Accurate Scene Text Detector," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5551-5560, 2017. [Google Scholar] [Publisher Link]

[16] Arwa Al Saqaabi et al., "A Deep Learning Approach for Paragraph-Level Paraphrase Generation for Plagiarism Detection," *Neural Processing Letters*, vol. 57, no. 3, pp. 1-42, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[17] Baoguang Shi, Xiang Bai, and Cong Yao, "An End-To-End Trainable Neural Network for Image-Based Sequence Recognition and its Application to Scene Text Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 11, pp. 2298-2304, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[18] Minghao Li et al., "Trocr: Transformer-based Optical Character Recognition with Pre-Trained Models," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37. no. 11, pp. 13094-13102, 2023. [Google Scholar]

[19] Palvadi Srinivas Kumar, and Krishna Prasad, "Integrating OCR and NLP Techniques for Accurate Text Extraction and Plagiarism Detection in Image-Based Content," *Library Progress International*, vol. 44, no. 3, pp. 2986-2996, 2024. [Google Scholar] [Publisher Link]

[20] Palvadi Srinivas Kumar, and Krishna Prasad, "Integrating OCR and NLP Techniques for Accurate Text Extraction and Plagiarism Detection in Image-Based Content," *International Journal of Advanced Science and Computer Applications*, vol. 4, no. 1, pp. 1-8, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[21] Basant Agarwal et al., "Siamese-Based Architecture for Cross-Lingual Plagiarism Detection in English-Hindi Language Pairs," *Big Data*, vol. 11, no. 1, pp. 48-58, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[22] Alaa Sahl Gaafar, Jasim Mohammed Dahr, and Alaa Khalaf Hamoud, "Comparative Analysis of Performance of Deep Learning Classification Approach based on LSTM-RNN for Textual and Image Datasets," *Informatica*, vol. 46, no. 5, pp. 21-28, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[23] Abdur Razaq et al., "Identification of Paraphrased Text in Research Articles through Improved Embeddings and Fine-Tuned BERT Model," *Multimedia Tools and Applications*, vol. 83, no. 30, pp. 74205-74232, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[24] Youngmin Baek et al., "Character Region Awareness for Text Detection," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, pp. 9357-9366, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[25] Alzahrani, Salha M., Naomie Salim, and Ajith Abraham, "Understanding Plagiarism Linguistic Patterns, Textual Features, and Detection Methods," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 2, pp. 133-149, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[26] Pon Abisheka, C. Deisy, and P. Sharmila, "T-SRE: Transformer-based Semantic Relation Extraction for Contextual Paraphrased Plagiarism Detection," *Journal of King Saud University-Computer and Information Sciences*, vol. 36, no. 10, pp. 1-13, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[27] Yu Han et al., "Breaking through Language Barriers: A Review of OCR Technology for Low-Resource Minority Languages Based on Deep Learning," *SSRN*, pp. 1-42, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[28] Milind Agarwal, and Antonios Anastasopoulos, "A Concise Survey of OCR for Low-Resource Languages," *Proceedings of the 4th Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP 2024)*, Mexico City, Mexico, pp. 88-102, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[29] Joseph Redmon, and Ali Farhadi, "Yolov3: An Incremental Improvement," *arXiv Preprint*, pp. 1-6, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[30] Alexey Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," *arXiv Preprint*, pp. 1-22, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[31] Kaiming He et al., "Deep Residual Learning for Image Recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770-778, 2016. [Google Scholar] [Publisher Link]

[32] Franco Scarselli et al., "The Graph Neural Network Model," *IEEE Transactions on Neural Networks*, vol. 20, no. 1, pp. 61-80, 2009. [CrossRef] [Google Scholar] [Publisher Link]

[33] Wenhai Wang et al., "Pyramid Vision Transformer: A Versatile Backbone for Dense Prediction Without Convolutions," *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 568-578, 2021. [Google Scholar] [Publisher Link]

[34] Jiyang Xie et al., "Deep Learning-Based Computer Vision for Surveillance in its: Evaluation of State-of-the-Art Methods," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 4, pp. 3027-3042, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[35] Srinivas Kumar Palvadi, and Krishna Prasad, "A Unified Framework for Text Extraction and Plagiarism Detection in Image-Based Content Using OCR and NLP," *Physiotherapy Issues*, vol. 54, no. 1, pp. 132-141, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[36] Rishi Bommasani et al., "On the Opportunities and Risks of Foundation Models," *arXiv preprint*, pp. 1-214, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[37] Tsung-Yi Lin et al., "Microsoft Coco: Common Objects in Context," *European Conference on Computer Vision*, Zurich, Switzerland, vol. 7, pp. 740-755, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[38] Hamed Arabi, and Mehdi Akbari, "Improving Plagiarism Detection in Text Document Using Hybrid Weighted Similarity," *Expert Systems with Applications*, vol. 207, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[39] Sheetal Harris et al., "Fake News Detection Revisited: An Extensive Review of Theoretical Frameworks, Dataset Assessments, Model Constraints, and Forward-Looking Research Agendas," *Technologies*, vol. 12, no. 11, pp. 1-63, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[40] Md Kamrul Siam, Huanying Gu, and Jerry Q. Cheng, "Programming with Ai: Evaluating Chatgpt, Gemini, Alphacode, and Github Copilot for Programmers," *Proceedings of the 3rd International Conference on Computing Advancements*, Dhaka, Bangladesh, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[41] Noppol Anakpluek et al., "Improved Tesseract Optical Character Recognition Performance on Thai Document Datasets," *Big Data Research*, vol. 39, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[42] Yuliia Zanevych, "Flask vs. Django vs. Spring boot: Navigating Framework Choices for Machine Learning Object Detection Projects," *Collection of Scientific Papers «ΛΟΓΟΣ»*, Cambridge, UK, pp. 311-318, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[43] Juan Pablo Bustos, and Luis Lopez Soria, *Generative AI Application Integration Patterns: Integrate Large Language Models into Your Applications*, Packt Publishing Ltd, 2024. [Google Scholar] [Publisher Link]

[44] Amirul S. Bin Ibrahin, Othman O. Khalifa, and Diaa Eldein M. Ahmed, "Plagiarism Detection of Images," *2020 IEEE Student Conference on Research and Development (SCOReD)*, Batu Pahat, Malaysia, pp. 183-188, 2020. [CrossRef] [Google Scholar] [Publisher Link]