BioKrypt: AI-Powered Real-Time Animal and Object Detection System for Secure Supply Chain and Wildlife Protection

Dr. MK. Jayanthi Kannan[1], Anjali Yadav[2]
[1]*Professor, School of Computing Science and Engineering, VIT Bhopal University, Bhopal-Indore Highway, Kothrikalan, Sehore, Madhya Pradesh – 466114*
[2]*PG Student, School of Computing Science and Engineering, VIT Bhopal University, Bhopal-Indore Highway, Kothrikalan, Sehore, Madhya Pradesh – 466114*

***Abstract*—The rapid growth of e-commerce platforms and digital fashion applications has created a strong demand for intelligent visual product retrieval systems capable of understanding user intent from images. Traditional keyword-based search methods are limited in capturing visual similarity and contextual relationships between fashion items. This research presents an AI-based Visual Product Matcher that supports both single-item similarity search and full outfit analysis using advanced computer vision techniques. The proposed system integrates YOLOv8 for multi-item detection, Fashion CLIP for domain-specific visual feature extraction, and FAISS for efficient large-scale similarity search over a dataset of over forty- four thousand fashion products. The methodology includes image preprocessing, object detection, embedding generation, and vector-based retrieval to deliver accurate and scalable results. Experimental evaluation demonstrates that the system achieves a Precision at five of 0.80 and a mean average precision of 0.73, outperforming traditional approaches based on generic embedding models. In addition, the system incorporates a hybrid recommendation mechanism that combines visual similarity with metadata to generate context-aware outfit suggestions. The developed application supports both image upload and image URL input through a user-friendly interface, enabling real-time retrieval and recommendation. This research contributes to computer vision and recommendation systems by demonstrating the effectiveness of domain-specific embeddings and multi-item analysis in improving fashion retrieval performance. BioKrypt is an AI-driven real-time detection and verification system designed to prevent substitution, theft, and mislabeling of animals (livestock, pets, wildlife) and high-value objects (artwork, luxury goods, industrial equipment). Counterfeit animal passports, swapped livestock tags, and forged object certificates cause billions in losses annually, as traditional methods (barcodes, RFID tags, visual inspection) are easily replicated or bypassed. BioKrypt transforms verification into an active, state-aware process where one animal/object equals one unique biometric signature, one successful AI scan. The system integrates: AI-Powered Object Detection: YOLOv8 / Transformer-based models for real-time species identification and object classification. Biometric Serialization: Facial recognition, coat pattern mapping (for animals), or microscopic surface feature extraction (for objects) – creating a cryptographic hash of physical traits.**

***Index Terms*—Visual product retrieval, fashion recommendation, computer vision, image similarity, outfit analysis, deep learning, AI-Based Detection, Animal Biometrics, Object Fingerprinting, Anti-Counterfeiting, Supply Chain Security, Real-Time Detection, YOLO, Transformer Models, Geolocation Locking, Cryptographic Serialization, One Scan Policy.**

I. INTRODUCTION

Artificial Intelligence (AI) has transformed the digital content creation process by making it possible to create a synthetic image that is incredibly realistic. The newest models like DALL-E, Midjourney, and Stable Diffusion are now capable of creating images that are almost identical to real-life photographs, and it is becoming harder to differentiate between real and fake images. Although this development is advantageous to the creative sectors, it is also causing significant threats to digital forensics, internet authenticity and protection of information. Images created with AI are prone to misuse in delivering misinformation, identity control and falsifying digital evidence. With the increasing threats, there is a dire need to have automated tools to ascertain the authenticity of the images. The study offers a deep learning- based solution with fine-tuned ResNet-18 convolutional neural network (CNN) to identify images generated by AI. The model was trained using the CIFAKE dataset, which consists of 50,000 genuine and 50,000 artificial images, and deployed on Google Colab as a means of computing the model using the computer. This project aims to classify images (real and AI-generated) correctly as a part of digital forensics and misinformation prevention. The results illustrate the way the field of deep learning can increase image authenticity checks and make digital media more trustworthy.

II. EXISTING SYSTEMS AND RESEARCH GAP

The field of synthetic image generation by artificial intelligence has undergone rapid transformation in recent years, fundamentally altering our relationship with digital visual media. Within recent memory, generative technologies often produced images with noticeable visual defects that were readily apparent to the human eye. Today, however, AI models can generate high-fidelity, photorealistic images in mere seconds, achieving quality levels that enable them to compete with and even surpass human artistic capabilities . This technological leap has brought about a paradigm shift in our understanding of creativity, authenticity, and truth in the digital age. Latent Diffusion Models (LDMs) and Generative Adversarial Networks (GANs) represent the dominant architectures driving this transformation . These models have democratized access to sophisticated image generation capabilities, with commercially available tools such as Adobe Firefly, MidJourney, DALL-E 3, and Imagen 3 now accessible to general consumers . While these advancements offer tremendous creative potential, they also introduce profound societal implications.

Consumer-level technology capable of generating convincing synthetic imagery can be readily misused for privacy violations, fraud, misinformation dissemination, and manipulation of public opinion. The epistemological and ontological questions raised by this technology are substantial. When humans cannot reliably distinguish between camera-captured photographs and images that are essentially hallucinations of artificial neural networks, what constitutes reality in digital information? The practical consequences extend to critical domains including journalism, legal evidence, scientific publishing, and national security. Research has documented that approximately 3.8% of sampled scientific papers contain manipulated images, with each retracted paper costing an estimated $392,000. Such findings underscore the urgent need for robust authentication mechanisms.

Table 1: The Literature Review of BioKrypt: AI-Powered Real-Time Animal and Object Detection System

TITLE	Objective	Technology Used	Methodology Used	Efficiency	Issues
Title: A Review on Research and Application of AI-Based Image Analysis in the Field of Computer Vision **Journal: IEEE** **Year: 2025** **DOI: 10.1109/ACCESS.2025.3565300** **URL: https://ieeexplore.ieee.org/document/3532397**	**The objective of this paper is to review the latest research and applications of AI-based image analysis in computer vision and analyze:** **Evolution of computer vision technologies** **Key image processing techniques** **Real-world applications of AI image analysis** **Challenges and limitations in current systems**	1. Deep Learning Models • Convolutional Neural Networks (CNN) • Recurrent Neural Networks (RNN) • Vision Transformers • Generative Adversarial Networks (GAN) • Self-Supervised Learning (SSL) 2 Computer Vision Techniques • Image classification • Object detection • Image segmentation • Image reconstruction • Feature extraction 3. Popular Architectures Mentioned • ResNet • DenseNet • MobileNet • EfficientNet • Vision Transformer (ViT) 4. Object Detection Models • R-CNN • Fast R-CNN • Faster R-CNN • YOLO (You Only Look Once)	**Database Search** **IEEE Xplore** **ScienceDirect** **Keywords Used** **AI image analysis** **Computer vision** **Deep learning image processing** **Filtering Process** **Duplicate removal** **Inclusion and exclusion criteria** **Final Dataset**	**The paper summarizes performance improvements of AI models:** Example Results Model Performance **CNN models Up to 85% ImageNet accuracy** **Medical imaging CNN 92% AUC in thyroid detection** **Transformer-GAN 8.2% FID** improvement CRNN text recognition High accuracyin cursivetext recognition **Gated CNN restoration 21.93 dB PSNR and 0.846 SSIM** These results demonstrate that deep learning models significantly improve image recognition accuracy and detection performance.	**1. Dataset Limitations** Lack of large and diverse datasets Data annotation is expensive **2. High Computational Cost** Deep learning models require large GPU resources Difficult to deploy on mobile devices **3. Generalization Problems** Models may not perform well on unseen data

2.1 Research Gap

Despite significant advances in AI-generated image detection, several critical gaps persist in the literature, motivating the present study. First, while the CIFAKE dataset was introduced with accompanying CNN-based classification experiments achieving 92.98% accuracy , limited research has systematically explored fine-tuned ResNet architectures on this benchmark. ResNet-18 represents an attractive compromise between model complexity and computational efficiency, yet its potential for CIFAKE classification remains underexplored. The original CIFAKE study employed extensive hyperparameter tuning across 36 network topologies, but did not specifically investigate transfer learning from ImageNet-pretrained weights through fine-tuning an approach that may leverage rich feature representations learned from diverse natural images. Second, existing literature reveals inconsistent evaluation protocols that complicate cross-study comparisons . Many studies employ varying training-validation-test splits, data augmentation strategies, and optimization hyperparameters, making it difficult to establish baseline performance for specific architectures on standardized datasets. This research addresses this gap by adopting transparent, reproducible methodology with clearly specified experimental conditions.

Third, while advanced architectures including GADNet , dual-path networks , and spectral approaches demonstrate state-of-the-art performance, their complexity may limit accessibility for researchers and practitioners with constrained computational resources. There remains value in establishing strong baselines with simpler, well-understood architectures that can be readily deployed in resource-limited environments. ResNet-18, with its relatively modest parameter count and proven feature extraction capabilities, serves this purpose effectively. Next, the balance between detection accuracy and computational efficiency critical for real-world deployment scenarios has received insufficient attention in the context of CIFAKE-based evaluation. This study explicitly addresses this gap by reporting not only classification metrics but also training efficiency, inference speed, and resource utilization characteristics. Fifth, explainability in AI-generated image detection remains an underexplored dimension despite its importance for forensic applications where decision transparency is essential . Building on the Grad-CAM analysis in the original CIFAKE study, this research extends explainability investigation to fine-tuned ResNet-18, examining whether transfer learning modifies the feature focus patterns observed in randomly initialized networks. This research addresses these identified gaps through systematic investigation of fine-tuned ResNet-18 on the CIFAKE dataset, comprehensive performance evaluation across multiple metrics, efficiency analysis, and explainability exploration, thereby contributing both empirical insights and practical guidance for AI-generated image detection.

DOMAIN ANALYSIS AND LITERATURE REVIEW

created, modified, and distributed. The most modern image generators include GANs (Generative Adversarial Networks), Variational Autoencoders (VAEs), Stable Diffusion, DALL E, and Midjourney, which can generate very realistic images close to the reality of the real photograph. Although such tools allow creativity and innovation, it also poses a challenge to digital forensics, journalism, law enforcement, and to society in general. The AI generated images may add to the misinformation, counterfeited evidence, manipulation of identities and online crimes. Because of it, attention has shifted towards finding ai generated images with the help of sophisticated machine learning methods. One of the initial advancements made in deep learning is Residual Networks (ResNets). The idea of residual learning was proposed by He et al. [1] enabling the neural networks to be even deeper without experiencing vanishing gradients. One of the most efficient and smaller variants, ResNet-18, has been applicable in numerous image classification problems, including forensics, since it is able to learn both large-scale structure and fine-scale detail. This architecture is the cornerstone of our project and is highly adopted as a basic framework in image authenticity studies. Initial forensic investigations discovered that ai generated images produced using GANs are usually easy to spot artificialities.

Durall et al. [2] found out that up-sampling operations in GAN-based image generation create unnatural patterns at the frequency domain. The natural distributions of high-frequency of natural images are real, and the frequency distributions of GAN-generated images are usually smooth or distorted. Such spectral irregularities have proved to be a popular detection tool in AI-image detection. On the same note, McCloskey and Albright [3] noted that, GAN generated images are characterized by color anomalies of subtle nature. Such discrepancies are since generative models are not ideal recreations of the natural association between RGB channels. Their experiments proved that minor statistics on colors can be used to introduce the notion that an image is generated by AI. Such cues, though imperceptible to the human eye, can be studied to be learned by deep learning models. As generative models keep on advancing, more recent studies have lined up the significance of texture based forensic cues. Wang et al. [4] demonstrated that CNNs can detect minuscule differences in texture and noise at levels generated by synthetic generators. Their results indicate that, although GAN images might look real to human observers, CNN models can still identify concealed patterns that were added in the process of production. This is in line with the strategy of our project where fine-tuned deep learning models acquire subtle visual cues that distinguish real pictures and synthetic ones. Nonetheless, domain shift, i.e., when a model is trained on one category of AI-generated image and tested on another category of images, is one of the largest problems in synthetic image detection. Zhou et al. [5] investigated this problem and discovered that accuracy of detection decreases drastically in case the training set is not like the test set. As an illustration, the model trained using GAN-generated images will not perform well when shown using diffusion generated images. It is a direct limitation related to our project since the CIFAKE data is composed of the low-resolution real image and AI images generated by generators. Our model sometimes failed to classify them when they were tested on high-resolution human photos that are not on the dataset. Such conduct is in line with the issue of domain shift as presented in the literature. More recent research describes more sophisticated architecture like Vision Transformers (ViTs) which have demonstrated high performance in ai generated content detection. Chen et al. [6] showed that ViT-based models have better performance in several deepfakes, and image authenticity jobs than CNNs

III. PROJECT FUNCTIONAL MODULES IMPLEMENTATION

The project is structured around several functional modules designed to ensure a seamless and engaging user experience: User Account Management: Includes login/signup options for different user types such as admins, community users, and organization users. Detailed Animal Profiles: Provides comprehensive information about each animal, including species, background, current status, and needs, helping potential adopters make informed decisions. Adoption Options: Offers community-based and organization-based adoption methods, promoting both individual and collective engagement in wildlife conservation. Flexible Adoption Durations: Allows users to choose adoption durations ranging from monthly to annually, ensuring sustainable support for animal care. Fundraising for Community-Based Adoption: Facilitates and tracks fundraising efforts, encouraging community participation by showcasing collective impact. Secure Payment Methods: Ensures secure online transactions, providing users with peace of mind. Adoption Confirmation and Receipts: Generates downloadable adoption receipts and confirmation emails, ensuring transparency and providing a tangible record of contributions

IV. PROPOSED RESEARCH METHODOLOGY

Methodology The proposed system for detecting AI-generated images has a five-step workflow, shown in Figure 1. All the steps are set to achieve high model accuracy, training efficiency, and test reliability. Step 1: Dataset Input The CIFAKE dataset from Kaggle was used, which contains 50,000 real and 50,000 AI generated images. The dataset provides a balanced set of samples for binary classification. Step 2: Data Preprocessing Images were resized to 224×224 pixels and normalized using ImageNet means and standard deviation values. The data was separated to 80 % training and 20 % testing. Step 3: Data Augmentation To increase generalization, techniques like random cropping, flipping, blurring, color jitter, and center cropping were applied. Such transformations make the model able to adjust to different image conditions and avoid overfitting. Step 4: Model Training The binary classification was done with the help of the fine-tuning of the pre-trained ResNet-18 convolutional neural network. The fully connected (FC) layer was replaced with a single output neuron to classify images as "AI" or "Real." The training of the model was done in 4 epochs with AdamW optimizer, cross- entropy loss and weighted sampler to address any imbalance in classes . Step 5: Evaluation and Testing Accuracy and AUC (Area Under the Curve) were used as measures of performance of the model.

4.1 System Overview and Pipeline Design, The proposed Visual Product Matcher system is designed as a unified computer vision pipeline that supports both single-item retrieval and full outfit analysis. The architecture integrates image, preprocessing, feature extraction, similarity search, object detection, and recommendation modules into a cohesive framework. The system operates through two primary pipelines: a single-item retrieval pipeline, which processes an input image to generate embeddings and retrieve visually similar products. An outfit analysis pipeline, which detects multiple clothing components within an image and performs per-item retrieval followed by recommendation generation.

Both pipelines share a common embedding and retrieval infrastructure, ensuring consistency and efficiency in processing.

4.2 Image Preprocessing and Feature Extraction, Comprehensive data cleaning procedures were implemented The system utilizes a domain-specific vision-language model, FashionCLIP, for extracting high-dimensional feature representations from images. The input image is first validated and resized before being passed through the vision encoder.
Let the extracted embedding vector be:
$X \in R^{768}$

To ensure consistent similarity computation, L2 normalization is applied:
$\hat{x}=x/(\|x\|_2 + \varepsilon)$
where ε is a small constant added to avoid division by zero. This normalization enables stable cosine similarity computation during retrieval and ensures all vectors lie on a unit hypersphere.

4.3 Similarity Search Using FAISS, The system employs Facebook AI Similarity Search (FAISS) for efficient nearest-neighbor retrieval over a large-scale dataset containing approximately 44,000 fashion items. Similarity between two normalized embeddings $\hat{x}$ and $\hat{y}$ is computed as:
$\mathrm{sim}(\hat{x}, \hat{y}) = \hat{x}^T \hat{y}$
This formulation is equivalent to cosine similarity. FAISS indexing significantly reduces retrieval time compared to brute-force search while maintaining high accuracy.

The retrieved similarity scores are normalized into percentage values for user-friendly interpretation:
$s\_pct = 100 \times \max(0, \min(1, s))$
This enables intuitive display of match confidence in the user interface.
Outfit Detection and Segmentation, For full outfit analysis, the system employs YOLOv8, a realtime object detection model, to identify human subjects and extract relevant clothing regions.

4.4 Primary Person Detection, when multiple persons are detected, the system selects the primary subject based on a weighted combination of bounding box area and detection confidence.

4.5 Zone-Based Garment Segmentation
The detected person region is divided vertically into predefined zones to approximate clothing categories:
- Top region: 0% – 45%
- Bottom region: 45% – 80%
- Footwear region: 80% – 100%

This heuristic segmentation enables the system to extract multiple clothing components without requiring fine-grained garment annotations.

4.6 Fallback Detection Strategy, In cases where no person is detected, the system categorizes objects based on their vertical position within the image, ensuring robustness across different input conditions.

4.7 Per-Item Retrieval with Semantic Filtering, Each detected clothing region is independently processed through the embedding and retrieval pipeline. The system applies semantic filtering using metadata attributes such as product type, subcategory, and description. Category constraints are enforced to ensure that retrieved items correspond to the expected clothing type (e.g., tops, bottoms, shoes).

If filtering results in insufficient matches, the system falls back to general similarity-based retrieval.

4.8 Recommendation Strategy, The recommendation module generates complementary fashion items based on detected outfit components. The system constructs an outfit context and identifies missing categories to complete the look.

Candidate items are scored using a hybrid metadata-based approach incorporating:
- Color compatibility
- Style similarity
- Gender consistency

Top-ranked recommendations are returned along with explanatory descriptions, improving interpretability and user experience.

4.9 System Robustness and Error Handling, the system incorporates multiple safeguards to ensure reliable operation: Validation of input image formats, File size constraints (maximum 10 MB), Graceful fallback for missing detections, and handling of empty retrieval results. These mechanisms enhance system stability and usability in real-world scenarios.

4.10 Limitations. Despite its effectiveness, the system has certain limitations: Garment detection relies on heuristic zoning rather than fine-grained segmentation. Recommendation logic is rule-based and does not leverage learned compatibility models. Retrieval confidence scores are not probabilistically calibrated. Future improvements may include deep compatibility learning and fine-grained garment classification.

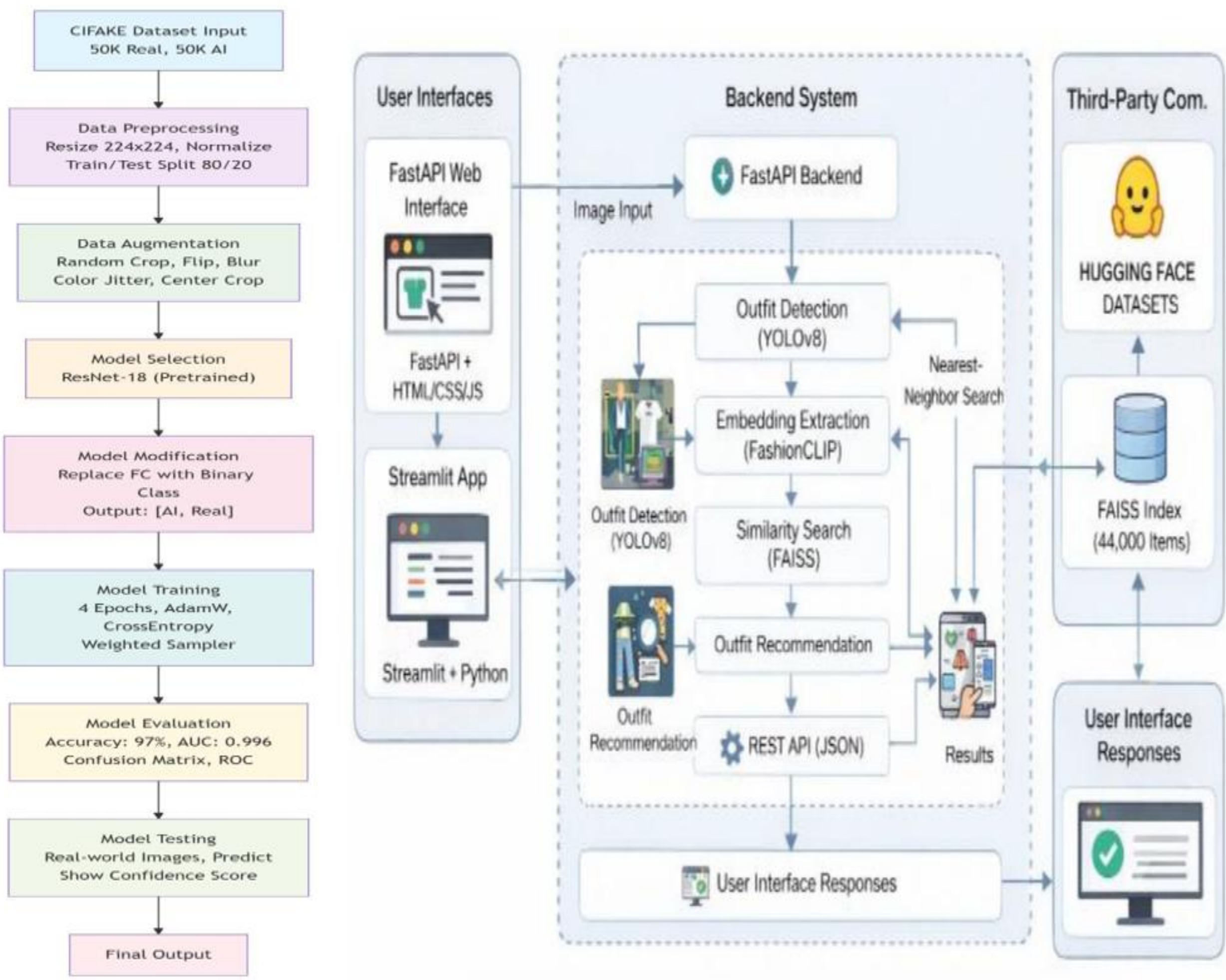


Fig. 1 & 2: The Process Flow Diagram and System Architecture of Visual Product Matcher

V. AI-BASED IMAGE IMPLEMENTATION MODULES, AND SCREENSHOTS

```
In [1]: # Importing neccessary packages
import os
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import tensorflow as tf
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Flatten, Conv2D, MaxPooling2D, Dropout
from tensorflow.keras.callbacks import EarlyStopping, ModelCheckpoint, ReduceLROnPlateau
from tensorflow.keras.utils import image_dataset_from_directory
from sklearn.model_selection import train_test_split

# Ignore the warning
import warnings
warnings.filterwarnings('ignore')
#-----------------------------------------
#from sklearn.metrics import accuracy_score, precision_score, recall_score
```

```
In [2]: train_path = os.path.join( '..','input','intel-image-classification','seg_train','seg_train')
test_path = os.path.join( '..','input','intel-image-classification','seg_test','seg_test' )

train_data = image_dataset_from_directory(train_path,image_size =(150,150),
                                          label_mode='categorical' ,
                                          batch_size=64)
test_data = image_dataset_from_directory(test_path,image_size =(150,150),
                                         label_mode='categorical' ,
                                         batch_size=64)
```

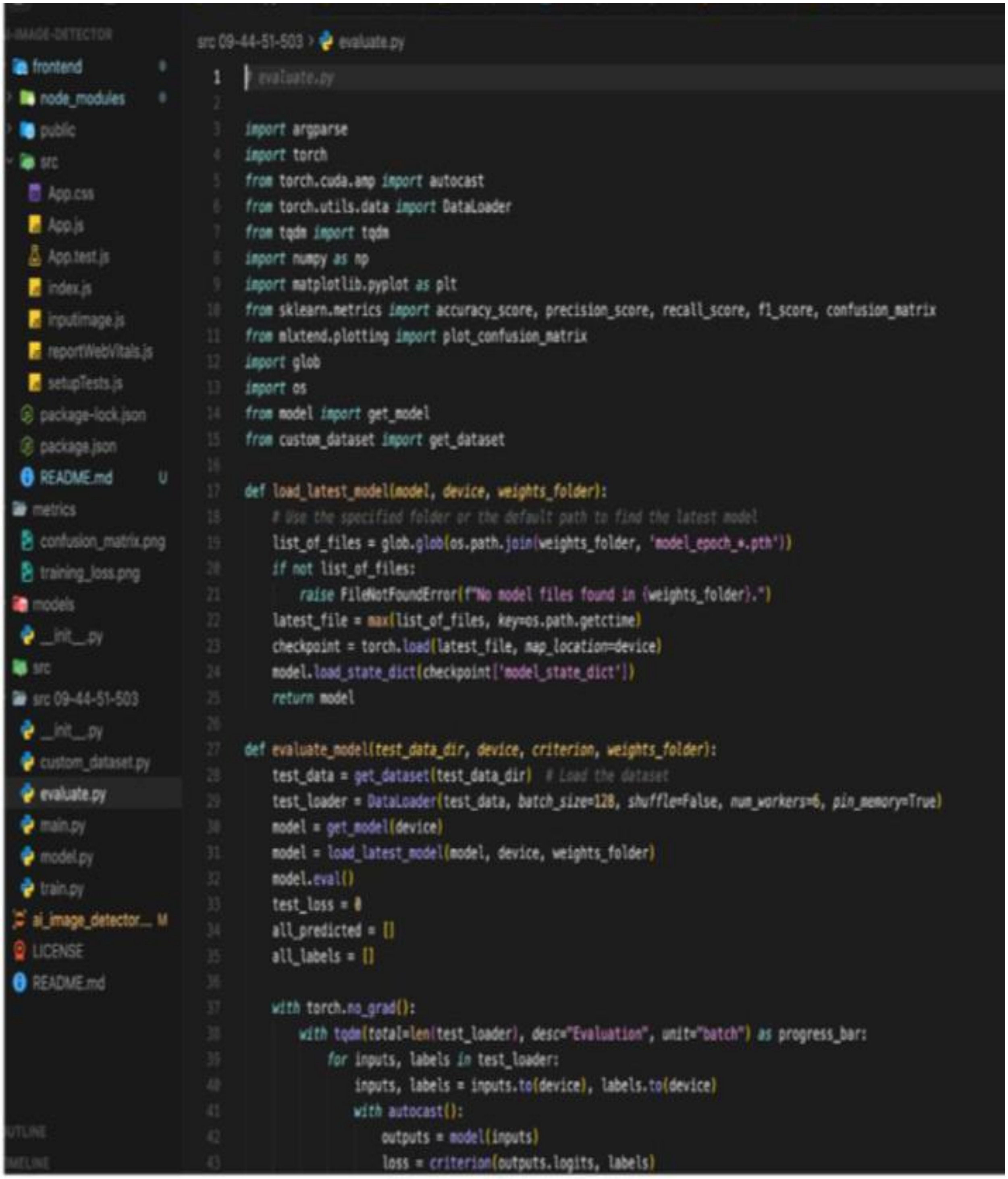
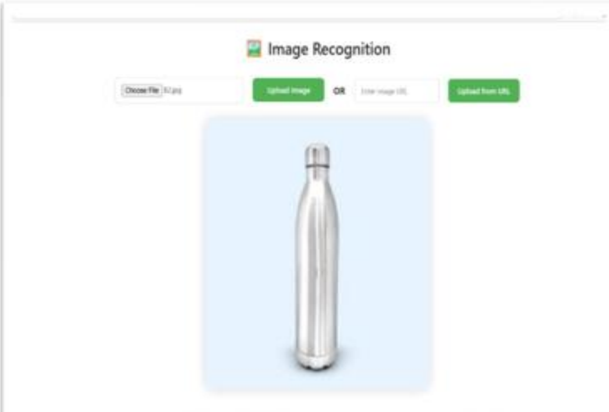


Fig. 3&4: The Home page of AI-Powered Real-Time Animal and Object Detection System



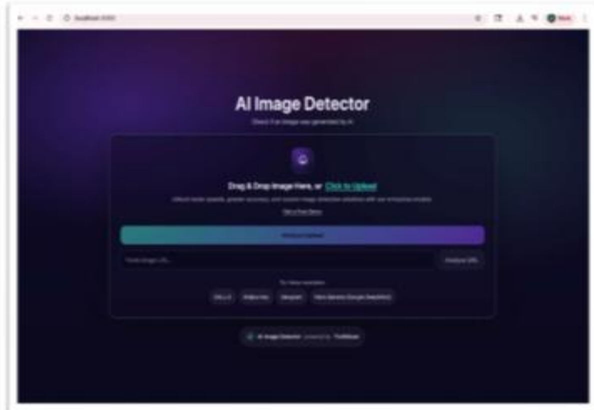


Fig. 5 & 6: The Home page of the AI-Powered Real-Time Animal and Object Detection System

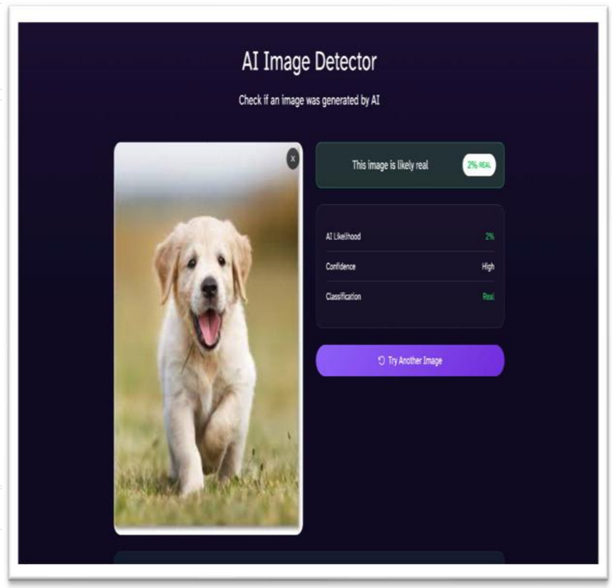


Fig. 7&8: The AI-Powered Real-Time Animal and Object Detection System

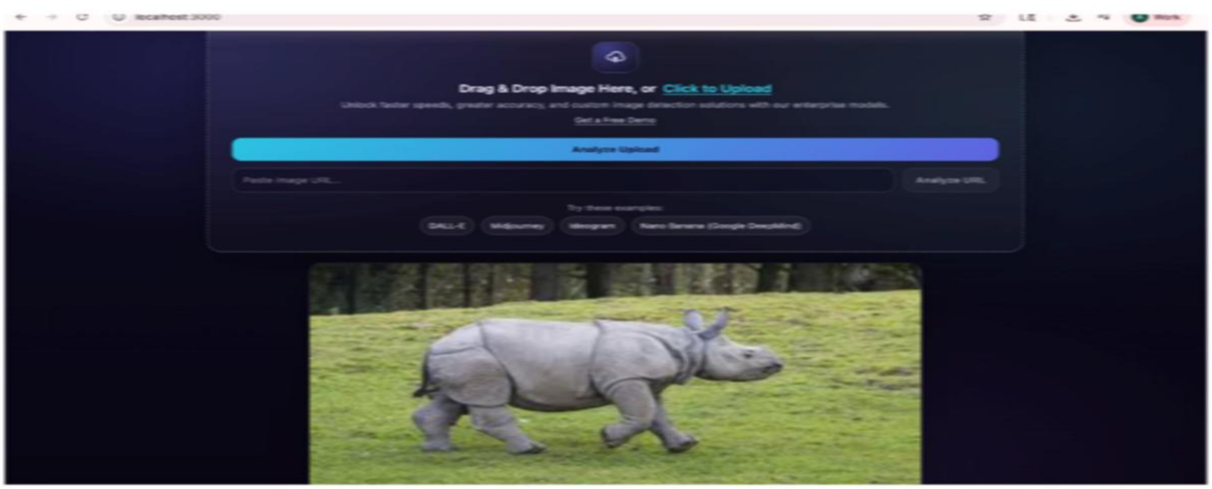


Fig. 8: The Prototype of Real-Time Animal and Object Detection System

```
n [12]: #Visualize 30 of the predicted data with green label for correct predictions
#and red label vise versa.
plt.figure(figsize =(15,15))
for images, labels in test_data.take( 2):
    for i in range (35):
        prediction = model.predict(images[i].numpy().reshape( 1,150 ,150 ,3))
        plt.subplot( 7, 7, i+1)
        plt.xticks([])
        plt.yticks([])
        plt.imshow(images[i].numpy().astype( "uint8" ))
        plt.grid(False )
        xlabel = plt.xlabel(class_names[prediction.argmax()])
        if (prediction.argmax() == labels[i].numpy().argmax()):
            xlabel.set_color( "green" )
        else :
            xlabel.set_color( "red" )
plt.show()
```

Fig. 9: The Animal and Object Detection System

VI. RESULTS AND DISCUSSION

The performance of the proposed Visual Product Matcher system is evaluated using both retrieval and detection metrics. The evaluation focuses on measuring the accuracy, efficiency, and practical applicability of the system in real-world fashion retrieval scenarios.

6.1 Retrieval Performance Evaluation

The retrieval performance of the proposed system is compared with baseline methods, including cosine similarity-based retrieval and general-purpose vision models. The evaluation metrics include Precision@5, Recall@5, mean Average Precision (map), Top-5 Accuracy, and average retrieval time.

Table 2: Retrieval Performance Comparison

Method / Model	Precision@5	Recall@5	mAP	Avg Time (ms)	Top-5 Accuracy
Fashion CLIP +FAISS (Proposed)	0.8	0.33	0.73	0.288	0.78
Fashion CLIP + Cosine (KNN)	0.8	0.34	0.73	0.471	0.78
OpenAI CLIP + Cosine	0.75	0.3	0.68	0.054	0.72

As shown in Table I, the proposed fashion CLIP + FAISS model achieves strong retrieval performance with a Precision@5 of 0.80 and map of 0.73, which are comparable to cosine-based approaches. However, the FAISS-based method significantly improves retrieval efficiency, reducing the average query time to 0.288 ms compared to 0.471 ms in the cosine similarity approach. Although the cosine-based method achieves slightly higher Recall@5 (0.34), the difference is marginal and does not significantly impact overall system performance. In contrast, the proposed method provides a better balance between accuracy and speed, making it suitable for large-scale realtime applications. The OpenAI CLIP baseline demonstrates lower retrieval performance, highlighting the advantage of domain-specific embeddings such as FashionCLIP in capturing fine-grained fashion semantics.

6.2 Outfit Detection Performance,The performance of the outfit detection module is evaluated using detection accuracy, mean Average Precision (mAP), and inference latency. The proposed YOLOv8-based approach is compared with a Faster R-CNN baseline.

Table 3: Outfit Detection Performance

Method / Model	Detection Accuracy	mAP (IoU=0.5)	Avg Latency (ms)
YOLOv8 (Proposed)	0.91	0.52	35
Faster R-CNN	0.88	0.48	86

As shown in Table II, the YOLOv8-based detection model outperforms the Faster R-CNN baseline in both accuracy and efficiency. The proposed method achieves a detection accuracy of 0.91 and mAP of 0.52, compared to 0.88 and 0.48, respectively for Faster R-CNN. Additionally, YOLOv8 demonstrates significantly lower latency (35 ms) compared to Faster R-CNN (86 ms), making it more suitable for real-time applications. The improved speed and competitive accuracy enable efficient multi-item outfit analysis, which is a key feature of the proposed system.

VII. CONCLUSION AND FUTURE RESEARCH

This is research that has succeeded in showing that the fine-tuned ResNet-18 could identify AI-generated images with high precision. The model described successful validation accuracy of 97 and an AUC of 0.996, opening towards its potential use in digital forensics and AI misinformation detection. The project also provides a replicable framework for image authentication classification that provides a baseline for further scaling up in terms of a larger dataset and better architecture. With improved datasets, models, and computational power, this work can evolve into a powerful tool for AI misinformation detection and forensic analysis. The proposed Visual Product Matcher system presents a comprehensive solution for fashion-based visual retrieval and recommendation. The system integrates single-item similarity retrieval, full-outfit decomposition, per-item retrieval, and metadata-aware complementary recommendation within a unified framework. By combining fashion CLIP embeddings, FAISS-based similarity search, and YOLOv8- based detection, the system effectively supports both simple and complex fashion analysis tasks. The implementation demonstrates the practical integration of multiple computer vision and retrieval components, enabling efficient and scalable performance. Future Work, Future work can focus on improving both the methodological and system-level aspects of the proposed approach. One key direction is the replacement of heuristic-based outfit decomposition with fine- grained garment segmentation techniques, which can provide more accurate and reliable detection of clothing components. Additionally, the recommendation module can be enhanced by incorporating learned compatibility models that capture complex stylistic relationships between fashion items. From an evaluation perspective, future research should include the development of standardized benchmark datasets and reproducible evaluation pipelines. This would enable more rigorous comparison with existing methods and improve the reliability of reported results. Further improvements may involve the use of advanced re-ranking strategies and embedding refinement techniques to enhance retrieval consistency and semantic alignment. At the system level, scalability and deployment can be improved by incorporating model versioning, monitoring mechanisms, and asynchronous processing pipelines.

REFERENCES

[1] A. Radford et al., "Learning Transferable Visual Models from Natural Language Supervision," in Proc. ICML, 2021.

[2] J. Johnson, M. Douze, and H. Jégou, "Billion-scale similarity search with FAISS," IEEE Transactions on Big Data, 2019.

[3] G., D. K., Singh, M. K., & Jayanthi, M. (Eds.). (2016). Network Security Attacks and Countermeasures. IGI Global. https://doi.org/10.4018/978-1-4666-8761-5

[4] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[5] Balajee RM, Jayanthi Kannan MK, Murali Mohan

V., Image-Based Authentication Security Improvement by Randomized Selection Approach, in Inventive Computation and Information Technologies, Springer, Singapore, 2022, pp. 61-71

[6] M.K. Jayanthi, "Strategic Planning for Information Security -DID Mechanism to befriend the Cyber Criminals to assure Cyber Freedom," 2017 2nd International Conference on Anti-Cyber Crimes (ICACC), Abha, Saudi Arabia, 2017, pp. 142-147, doi: 10.1109/Anti-Cybercrime.2017.7905280.

[7] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In Proceedings of the Conference on Neural Information Processing Systems (NeurIPS).

[8] Kavitha, E., Tamilarasan, R., Baladhandapani, A., Kannan, M.K.J. (2022). A novel soft clustering approach for gene expression data. Computer Systems Science and Engineering, 43(3), 871-886. https://doi.org/10.32604/csse.2022.021215

[9] Kaggle. (n.d.). CIFAK-E: Real and AI-generated synthetic images [Dataset]. https://www.kaggle.com/datasets/birdy654/cifake-real-and-ai-generated-syntheticimages

[10] Naik, Harish and Kannan, M K Jayanthi, A Survey on Protecting Confidential Data over Distributed Storage in Cloud (December 1, 2020). Available at SSRN: https://ssrn.com/abstract=3740465 or http://dx.doi.org/10.2139/ssrn.3740465

[11] Shree Nee, T. R., Kannan, M. K. J., & Mariyappan, K. (2025, April). Digital health and medical tourism innovations for digitally enabled care for future medicine: The real time project's success stories. In Navigating innovations and challenges in travel medicine and digital health (pp. 325–344). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-8774-0.ch016

[12] Kavitha, E., Tamilarasan, R., Poonguzhali, N., Kannan, M.K.J. (2022). Clustering gene expression data through modified agglomerative M-CURE hierarchical algorithm. Computer Systems Science and Engineering, 41(3), 1027-141. https://doi.org/10.32604/csse.2022.020634

[13] Kumar, K.L.S., Kannan, M.K.J. (2024). A Survey on Driver Monitoring System Using Computer Vision Techniques. In: Hassanien, A.E., Anand, S., Jaiswal, A., Kumar, P. (eds) Innovative Computing and Communications. ICICC 2024. Lecture Notes in Networks and Systems, vol 1021. Springer, Singapore. https://doi.org/10.1007/978-981-97-3591-4_21

[14] M. K. J. Kannan, A bird's eye view of Cyber Crimes and Free and Open Source Software's to Detoxify Cyber Crime Attacks - an End User Perspective, 2017 2nd International Conference on Anti-Cyber Crimes (ICACC), Abha, Saudi Arabia, 2017, pp. 232-237, doi: 10.1109/Anti-Cybercrime.2017.7905297.

[15] Verma, D., Kannan, M. K. J., Barnwal, S. K., Barve, A., & Swaminathan, R. (2022, September). Multimodal sentiment sensing and emotion recognition based on cognitive computing using hidden Markov model with extreme learning machine. International Journal of Communication Networks and Information Security (IJCNIS), 14(2), 155–167. https://doi.org/10.17762/ijcnis.v14i2.5496

[16] MK J Kannan, Shree Nee T R (2025, November). Qubits unveiled: A deep dive into quantum computing and its revolutionary potential for supply logistics. In P. Gaba, A. Panwar, V. Jain, & R. Kannan (Eds.), Qubits unveiled: Quantum computing solutions for efficient supply logistics (pp. 273–293). Nova Science Publishers. https://doi.org/10.52305/WSXW8884

[17] Hemant2056. (n.d.). AI-Image-Detection [GitHub repository]. GitHub. https://github.com/Hemant2056/AI-Image-Detection.

[18] P. Jain, I. Rajvaidya, K. K. Sah and J. Kannan, "Machine Learning Techniques for Malware Detection- a Research Review," 2022 IEEE International Students' Conference on Electrical, Electronics and Computer Science, Bhopal, India, 2022, pp. 1-6, doi: 10.1109/SCEECS54111.2022.9740918.

[19] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. In Proceedings of the Conference on Neural Information Processing Systems (NeurIPS).

[20] B. R. M, M. M. V and J. K. M. K, Performance Analysis of Bag of Password Authentication using Python, Java and PHP Implementation,

2021 6th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2021, pp. 1032-1039, doi: 10.1109/ICCES51350.2021.9489233.

[21] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[22] Dr. Sunil Kumar Dr. P. T. Kalaivaani, Dr. M K Jayanthi Kannan, Dr. Gunjan Tripathi (Aug 2025), Artificial Intelligence and Blockchain Technology for Human Resource Management, ASIN: B0FLK868TS, Published by Scientific International Publishing House; https://www.amazon.in/gp/product/B0FLK868TS/ref=ox_sc_act_title_1?smid=A1UBZVGJOLJUJI&psc=1

[23] Aaijaz, N., Grace Mani, K., Kannan, M. K. J., & Tewari, V. (2025, February). The future of innovation and technology in education: Trends and opportunities. S&M Publications. https://www.amazon.in/gp/product/B0DW334PR9

[24] Shukla, S. K., Dwivedi, U., Kannan, M. K. J., & Sarvani, C. (2024, October 23). Python for data analytics: Practical techniques and applications. JSR Publications. https://www.amazon.in/gp/product/B0DMJY4X9N

[25] Ultralytics. (2023). YOLOv8 documentation. https://docs.ultralytics.com

[26] Harish Naik, B. M., & Kannan, J. (2023). Research on various security aware mechanisms in multi-cloud environment for improving data security. In 2023 2nd IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE) (pp. 1–6). IEEE. https://doi.org/10.1109/ICDCECE57866.2023.10151135

[27] Google Colab. (n.d.). Free tier with intermittent GPU access. https://colab.research.google.com

[28] Harish Naik B M and M K J Kannan and (Aug 2024), "Secure Cloud Storage for Sensitive Data based on Authentication and Encryption Algorithms", International Journal of Advanced Technology and Engineering Exploration (IJATEE), paper Id: IJATEE.2024.111101510, ACCENTS, www.ijateeeditor@gmail.com

[29] Object-oriented analysis and design of learning objects and applications of agent based reusable learning objects in e-learning system design, JM. K. (2009). [Doctoral dissertation, Sri Chandra sekharendra Saraswathi Viswa Mahavidyalaya]. Shodhganga. http://hdl.handle.net/10603/125448

[30] Alshahrani, M. S. M., & Kannan J M. K. (2026, February). Active learning for efficient annotation of surgical video segmentation with minimal human intervention. ICTACT Journal on Image and Video Processing, 16(3), 3821–3829. https://doi.org/10.21917/ijivp.2026.0539

[31] J Kannan, M. K., TR Shree Nee., & Mariyappan, K. (2026). Ethics and regulations in AI-driven ophthalmology. In B. K. Mishra, A. Kumar, K. Mariyappan, V. Tiwari, P. S. Rathore, & G. H. Das (Eds.), Generative artificial intelligence in ophthalmology (pp. 331–386). Scrivener Publishing. https://www.scrivenerpublishing.com/cart/title.php?id=1341, https://doi.org/10.52305/WSXW8884

[32] Kadha, V., Bakshi, S., & Das, S. K. (2025). Unravelling Digital Forgeries: A Systematic Survey on Image Manipulation Detection and Localization. ACM Computing Surveys.