

Multi-Model Deep Learning System for Image Curation

C. Lakshmi¹, Joice Anna Jacob², Nikhitha A. D³, Rashmi P. C⁴

¹²³B.E. Student, Department of Computer Science and Engineering, Vivekananda College of Engineering and Technology, Puttur, Karnataka, India

⁴Professor, Department of Computer Science and Engineering, Vivekananda College of Engineering and Technology, Puttur, Karnataka, India

Abstract – Large collections of photographs are commonly produced at weddings, celebrations and other events, making manual review time-consuming. This paper presents a hybrid, multi-module system for curating photographs containing human faces. The Django-based application supports batch upload of images and processes them using pretrained deep-learning components, computer-vision techniques, and rule-based scoring. MediaPipe is used for face detection and facial-landmark analysis, while DeepFace estimates facial emotion. OpenCV-based methods assess blur and sharpness, and handcrafted visual features are used to calculate an aesthetic score. Facial landmarks and geometric rules support pose categorization, while perceptual hashing identifies duplicate or near-duplicate images within a batch. The system combines available analysis scores using weighted scoring to determine image selection and organizes results into categories such as selected, rejected, blurry and duplicate. Images without a detected human face are excluded from the main curation workflow. The system stores analysis records and image files and presents batch results through a web interface. It provides a structured workflow to support the review and organization of event photographs.

Key Words: Human-Face Image Curation, Computer Vision, Deep Learning, MediaPipe, DeepFace, Facial Emotion Recognition, Blur Assessment, Aesthetic Scoring, Perceptual Hashing.

1. INTRODUCTION

Digital photography makes it possible to capture large numbers of images at events such as weddings, celebrations, and cultural programs. These collections may contain many photographs featuring people, including similar shots, blurred images, and variations in facial expression or pose. Reviewing and organizing each image manually can therefore require considerable time and effort. As the size of these collections keeps growing, there is a clear need for automated tools that can help users find the best photographs quickly.

Photo selection often involves assessing several characteristics, such as whether a human face is present, image sharpness, facial expression, pose, and overall visual quality. These assessments can be repetitive, and selection decisions may vary between reviewers. An automated system that assists with these tasks can provide a more

structured way to review and organize face-containing photographs.

Computer-vision and deep-learning techniques support different parts of this process. MediaPipe provides face detection and facial-landmark analysis, while DeepFace can estimate facial emotion. Other image characteristics, such as blur and aesthetic quality, can be assessed using image-processing methods and handcrafted visual features. Perceptual hashing can also help identify duplicate or near-duplicate photographs.

This paper presents a hybrid, multi-module image curation system for photographs containing human faces. The system combines pretrained analysis components with image-processing techniques and rule-based analysis. It uses facial landmarks and geometric rules to categorize pose, OpenCV-based measurements to assess blur and sharpness, and handcrafted features including sharpness, exposure, contrast, colorfulness, and saturation to calculate an aesthetic score. The system does not use a separately trained deep-learning model for aesthetic assessment.

The outputs of the analysis modules are combined using weighted scoring to help determine image selection and organization. Users upload image batches through a Django-based web application, which stores analysis records and presents the resulting categories. Images for which no human face is detected are excluded from the main curation workflow; therefore, faceless images are outside the system's current scope.

2. LITERATURE SURVEY

Automated image analysis has grown quickly with deep learning, but most studies address only one task at a time. The works reviewed below cover aesthetics, quality, duplicates, emotion, candid moments and blur.

[1] This study proposed an explainable aesthetic evaluation method that combines vision-language models, CLIP-IQA, LightGBM and SHAP. Its main strength is that each aesthetic score comes with an explanation of the factors behind it. Such transparency helps users understand why one photograph is rated higher than another. However, the study deals only with aesthetics and does not consider blur, facial emotion, pose or duplicate images, which are all needed for complete photo curation.

[2] This work developed an aesthetic quality assessment method that fuses multi-scale features using Spiking Neural Networks and Vision Transformers. Combining features from several scales helps the model capture visual characteristics more fully and improves prediction. The approach also shows that spiking networks can be paired with transformer-based feature extraction for image evaluation. It concentrates on aesthetic prediction alone and does not include blur detection, emotion analysis, pose assessment or duplicate identification.

[3] This paper presented a no-reference image quality assessment algorithm built on visual perception characteristics. Because it estimates quality without a reference image, it suits real situations where no original is available. It can therefore be applied directly to photographs collected from many different cameras. A drawback is that the method may need parameter tuning to stay reliable under different imaging conditions, which makes it harder to use on mixed event photographs.

[4] This research described a photography evaluation system that uses CNNs together with image features to judge photographic quality. This reduces the need for manual assessment and gives a systematic way to evaluate photographs. The use of learned features alongside conventional ones also makes the evaluation more consistent across different images. However, the system focuses on quality evaluation and does not bring face, pose, emotion and duplicate checks into one curation pipeline.

[5] This study applied hashing techniques to remove exact duplicate images from large cloud repositories. Comparing hash values is simple and cheap to compute, and it saves storage and keeps collections tidy. It also scales well, since hashes can be compared without loading the full images each time. The approach only covers duplicate detection. It says nothing about sharpness, aesthetics, facial expression or pose, so it cannot decide which photograph in a set is the best one.

[6] This work studied near-duplicate detection by combining perceptual hashing with deep learning. Hashing provides fast comparison, while the deep-learning component captures more complex visual similarity, so the method finds images that are similar but not identical. This is useful for burst shots and repeated poses that are common at events. The cost is higher computation, and the work is limited to duplicate detection rather than full image curation.

[7] This paper proposed a real-time facial emotion recognition system using deep learning. It analyzes facial features and classifies emotional states quickly, which makes it useful where rapid expression analysis is needed. The fast response also allows it to be used on live video as well as still images. The emotion result, however, is treated as a separate output and is not linked with other image selection criteria such as sharpness or composition.

[8] This study addressed blur detection by combining traditional image features with machine learning to separate blurred images from sharp ones. Mixing conventional processing with learning gives a practical and systematic way to identify blur. The extracted features are simple, so the method remains easy to apply to large sets of images. The study stops at blur classification and does not extend to an overall evaluation of image quality or to choosing the best photographs.

[9] This work introduced CandidNet, a two-stream deep-learning network for detecting candid moments in photographs. The two streams work together to tell spontaneous images from posed ones, which is valuable for event photography. Candid shots often capture natural expressions that people prefer to keep in an album. The method relies on specialized datasets and focuses on candid detection only, without combining other quality measures into a single selection decision.

[10] This paper proposed a full-scene defocus blur detection method based on multi-level distillation learning. Distilling knowledge at several levels improves feature learning and gives accurate blur detection even in complex scenes. It can also separate sharp regions from defocused ones across the whole image, instead of giving a single decision. The work is concerned with locating defocused regions and does not address automatic image selection based on several quality criteria.

Overall, these studies show that each technique works well for its own task, but none of them combines face, blur, pose, emotion, aesthetic and duplicate analysis to select photographs. This gap motivates the hybrid multi-module system proposed in this paper.

3. PROPOSED SYSTEM

The proposed system is a Django-based web application that accepts an image folder as a batch and automatically analyzes the photographs. The major modules are listed below and their techniques are summarized in Table -1.

- 1) Image Upload and Batch Management
- 2) Image Conversion and Preprocessing
- 3) Face Detection
- 4) Blur and Sharpness Analysis
- 5) Pose and Candid Analysis
- 6) Aesthetic Assessment
- 7) Emotion Recognition
- 8) Weighted Scoring
- 9) Duplicate Detection
- 10) Image Organization and Result Display

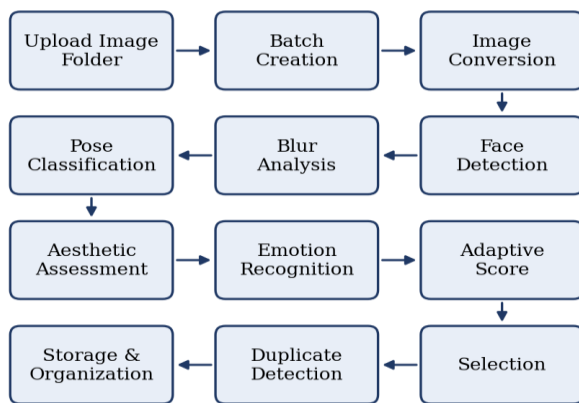


Fig -1: Overall processing flow of the proposed system

Table -1: Modules and techniques used

Module	Technique
Face detection	MediaPipe Face Detection
Emotion recognition	DeepFace (pretrained)
Blur / sharpness	OpenCV: Laplacian variance, Canny, Sobel
Aesthetic score	Handcrafted visual features
Pose / candid	MediaPipe Face Mesh + geometric rules
Duplicate detection	Perceptual hashing
Final selection	Weighted scoring

3.1 Image Upload and Preprocessing

The user uploads a folder containing photographs through the web interface. Django treats the uploaded folder as a single batch and creates a unique batch identifier. An analysis record is created for each image. During preprocessing, the image is converted into the required colour format and resized when necessary. A perceptual hash is also generated for later duplicate comparison.

3.2 Face Detection

MediaPipe Face Detection is used to determine whether a recognizable face is present in an image. The image is converted from BGR to RGB before detection. If no face is detected, the image is excluded from the curated collection and stored in the appropriate rejected category. Images with detected faces continue through the main analysis pipeline.

3.3 Emotion Recognition

The emotion-recognition module uses the pretrained DeepFace framework. DeepFace analyzes the detected facial region and estimates the dominant facial emotion. The detected emotion is mapped to a corresponding score, which is later used as one of the inputs to the weighted scoring

mechanism. If an emotion cannot be determined, the system records the result as Unknown.

3.4 Blur and Sharpness Analysis

Blur analysis is performed using OpenCV-based image-processing techniques rather than a separate deep-learning blur model. The Laplacian variance is used to estimate image sharpness. Based on the defined thresholds, images can be classified as Sharp or Slight Blur. For lower sharpness values, Canny edge density is used to identify intentional background blur or Bokeh. Sobel-based gradient analysis is then used to identify Motion Blur and Camera Shake. Motion Blur and Camera Shake are treated as hard rejection conditions.

3.5 Aesthetic Assessment

The aesthetic assessment module does not use a trained deep-learning aesthetic model. Instead, it calculates a score from handcrafted visual features: sharpness, exposure, contrast, colourfulness and saturation. The weighted aesthetic score is calculated as

$$A = 0.35S + 0.25E + 0.15C + 0.15F + 0.10T \quad (1)$$

where S represents sharpness, E represents exposure, C represents contrast, F represents colourfulness and T represents saturation. The resulting aesthetic score is normalized to the defined range used by the system.

3.6 Pose and Candid Analysis

MediaPipe Face Mesh is used to obtain facial landmarks. These landmarks are analyzed using rule-based geometric calculations involving facial width, eye positions and nose position. Based on these measurements, the system classifies images as Candid or Posed when a suitable face is detected. Images containing more than one detected face are categorized as Group.

3.7 Weighted Scoring

The outputs of the analysis modules are combined using predefined weights. The base weights used by the system are given in Table -2. Thus, the combined score can be represented as

Table -2: Base weights used in scoring

Component	Weight
Aesthetic score	0.40
Blur score	0.30
Pose/Candid score	0.15
Emotion score	0.15

$$\text{Score} = 0.40A + 0.30B + 0.15P + 0.15E \quad (2)$$

where A represents the aesthetic score, B represents the blur-related score, P represents the pose/candid contribution and E represents the emotion contribution. The resulting score is compared with a selection threshold of 6.5. Images with a score greater than or equal to 6.5 can be selected, provided that they do not satisfy a hard rejection condition such as Motion Blur or Camera Shake. Images without detected faces are also excluded from the curated collection.

3.8 Duplicate Detection

After image analysis and scoring, perceptual hashes are compared among images belonging to the same batch. Images with sufficiently similar hash values are identified as duplicate or near-duplicate photographs. Duplicate images are moved to the Duplicate category so that repeated photographs can be separated from the main curated collection.

4. SYSTEM IMPLEMENTATION

The system is implemented using Python and Django. Django handles user interaction, image-folder upload, batch creation, database operations, processing requests, result storage and result visualization. Image analysis is carried out by a persistent process pool with up to three worker processes. Each processing request handles up to 60 unprocessed images. Workers perform preprocessing and image analysis, then return results to the main Django process. The main process updates the database and integrates duplicate-detection results.

Table -3: Software Libraries and Methods Used

Purpose	Software / Method	Version
Face detection	MediaPipe Face Detection (BlazeFace)	0.10.7
Pose landmarks	MediaPipe Face Mesh	0.10.7
Emotion recognition	DeepFace	0.0.93
Deep learning backend	TensorFlow / Keras	2.12
Image processing	OpenCV	4.9.0.80
Arrays and image I/O	NumPy, Pillow	1.26.4
Sharpness	Laplacian variance	–
Edges and gradients	Canny, Sobel	–
Colourfulness	Colourfulness metric	–
Duplicate detection	pHash via ImageHash	4.3.2
Web framework	Django	5.0.6

The implemented flow is:

Upload Image Folder → Batch Creation → Image Conversion → Face Detection → Blur Analysis → Pose Classification → Aesthetic Assessment → Emotion Recognition → Weighted Score → Selection → Duplicate Detection → Storage and Organization.

The system therefore combines pretrained deep-learning components with OpenCV image processing, handcrafted visual scoring, rule-based landmark analysis and perceptual hashing. The analysis results stored for each photograph include face status, blur type, sharpness, pose, emotion, aesthetic score, perceptual hash, weighted score and selection status. These results are stored in the database and are displayed through the Django result interface. Software implementation and methods used is shown in Table -3.

5. RESULTS

The developed system was tested using photographs with different image sizes, lighting conditions, blur levels, facial expressions and poses. The testing included unit testing, integration testing, functional testing, system testing and acceptance testing. The test cases covered image upload, batch creation, face detection, blur classification, pose classification, aesthetic scoring, emotion recognition, weighted score calculation, image selection, categorization, duplicate detection and result display. Each module was first verified individually and then tested together with the other modules to confirm that data passed correctly between the stages.

According to the project testing the system was able to process image batches through the complete workflow, pass analysis results to the scoring and duplicate-detection stages, and store images according to their processing results. The system also handled unsupported and unsuitable images without interrupting the complete processing workflow. These results indicate that the implemented modules work together successfully for the tested image conditions. The major output categories include:

- Selected images
- Rejected images
- Blurry images
- Duplicate images
- Pose-based categories
- Aesthetic-based results

The result interface allows users to view processed images, inspect analysis information, monitor batch progress and review previous processing history

6. CONCLUSIONS

This paper presented a Django-based system for curating photographs that contain human faces. It combines MediaPipe for face detection and pose categorization, DeepFace for emotion estimation, OpenCV for blur and sharpness assessment, handcrafted aesthetic scoring, and perceptual hashing for duplicate detection. Their outputs are merged through weighted scoring to select and organize images. Functional testing showed that the modules work

together on the cases tested. The system currently handles only images with detectable faces, and its accuracy and generalizability still need to be evaluated on a documented, manually labelled dataset. Even so, it offers a practical, automated workflow that reduces the effort of sorting large photo collections.

7. FUTURE SCOPE

The system can be extended in several directions to improve its robustness, generalizability, and practical utility. First, the current aesthetic score, which relies on handcrafted or heuristic measures, could be benchmarked against a trained aesthetic-assessment model, such as one fine-tuned on large-scale datasets like AVA, to quantify how well it agrees with human perceptual judgments. Second, emotion and pose analysis could be strengthened for difficult cases, including group photographs with multiple subjects, faces that are partially turned away, and faces that are occluded by hands, sunglasses, masks, or other objects. This could be done by adopting more robust landmark estimators or multi-person pose models and by evaluating performance across varied lighting and camera angles. Third, near-duplicate detection could be evaluated more rigorously on images that have undergone cropping, resizing, compression, rotation, or minor edits, with the use of perceptual hashing or learned embedding-based similarity measures to improve tolerance to such transformations. Finally, the workflow could be made more flexible by allowing user-adjustable selection preferences, such as custom weights for sharpness, expression, or composition, and by extending support to non-face images such as landscapes, objects, and scenes, which are not handled by the current face-focused pipeline. Together, these enhancements would make the system more accurate, personalized, and applicable to a wider range of real-world photo collections.

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to our project guide, Prof. Rashmi P. C., Department of Computer Science and Engineering, Vivekananda College of Engineering and Technology, Puttur, for her constant guidance, patient supervision and valuable suggestions throughout this work. Her encouragement during the design, implementation and testing stages helped us move forward with confidence, and her feedback on the preparation of this paper improved its clarity and structure.

We are thankful to the Principal of the institution for providing a supportive academic environment and the facilities needed to carry out this project. We also thank the Head of the Department of Computer Science and Engineering for the encouragement, motivation and departmental resources that made this work possible.

Our thanks go to all the teaching and non-teaching staff of the department for their timely help and kind cooperation.

The knowledge and skills gained from their lectures formed a strong base for this project.

We are grateful to our friends and classmates for sharing photographs for testing, giving honest opinions on the results and supporting us throughout the work. Finally, we thank our parents and family members for their love, patience and constant support, which kept us motivated until the completion of this project.

REFERENCES

- 1) Supatta Viriyavisuthisakul, Shun Yoshida, Kaede Shiohara, Ling Xiao and Toshihiko Yamasaki. "Explainable AI for Image Aesthetic Evaluation Using Vision-Language Models." In 2025 Conference on Artificial Intelligence x Multimedia (AIxMM), 2025. doi: 10.1109/AIxMM62960.2025.00017.
- 2) Yang Liu, Shuo Zhang, Tengyu Fan and Minghu Fan. "Image Aesthetic Quality Assessment Method Based on Multi-scale Fusion Spiking Neural Network Vision Transformer." In Advanced Intelligent Computing Technology and Applications, Lecture Notes in Computer Science, vol. 15842, pp. 101–112, 2025. doi: 10.1007/978-981-96-9863-9_9.
- 3) Xu Wang et al. "Improved No-Reference Image Quality Assessment Algorithm Based on Visual Perception Characteristics" 2025.
- 4) Gabriel Garcia Budia et al. "CNN-based photography evaluation system using image features." 2025.
- 5) G. Subrahmanya Swamy, B. PruthviRaj Goud, K. Sudheer Reddy, K. Praveen Kumar and M. Sivasai. "Exact Image Deduplication using Hashing Techniques on Cloud Repositories." In 2025 5th International Conference on Intelligent Technologies (CONIT), 2025. doi: 10.1109/CONIT65521.2025.11167673.
- 6) Yash Jakhar and Malaya Dutta Borah. "Effective Near-Duplicate Image Detection Using Perceptual Hashing and Deep Learning." In Information Processing & Management, vol. 62, no. 4, Art. no. 104086, 2025. doi: 10.1016/j.ipm.2025.104086.
- 7) R. Chaitra and K. Vivekananda Bhat. "Real-Time Facial Emotion Recognition Using Deep Learning Approach." In 2025 Emerging Technologies for Intelligent Systems (ETIS), 2025. doi: 10.1109/ETIS64005.2025.10960833.
- 8) Shweta Jaiswal, Meenakshi Chandra, P. Mukilan, Amit Kumar Upadhyay and Madhuri Sharma. "Machine Learning for Detecting Blurred Images." In 2024 4th International Conference on Advancement in Electronics & Communication Engineering (AECE), 2024. doi: 10.1109/AECE62803.2024.10911165.
- 9) Gaurav Ramola, Nikhar Maheshwari and Sudha Velusamy. "CandidNet: A Novel Framework for Candid

Moments Detection.” In 7th International Conference on Computer Vision & Image Processing (CVIP), pp. 166–180, 2022.

- 10) Wenda Zhao, Fei Wei, Haipeng Wang, You He and Huchuan Lu. “Full-Scene Defocus Blur Detection With DeFBD+ via Multi-Level Distillation Learning.” In IEEE Transactions on Multimedia, vol. 25, pp. 9228–9240, 2023. doi: 10.1109/TMM.2023.3248162.