# Artificial Intelligence Techniques for Crowd Image Classification at Jabal Al-rahmah During Hajj

**Joud Ahmad Alhuthaly**
444002970

**Hanin Mesfer Almalki**
444003284

**Reem Ghazi Alosaimi**
444001268

**Nehal Hamed Alzhrani**
444001073

*Abstract*— **This study employs advanced machine learning techniques, including Convolutional Neural Networks (CNNs), Support Vector Machines (SVM), and Random Forest, to analyze real-time visual data from cameras and sensors at Jabal Al-Rahmah. These methods were selected for their effectiveness in handling complex image classification tasks and adapting to the challenges of dynamic and dense crowd scenarios. The system detects crowd density, movement patterns, and congestion points to proactively manage crowd flow and ensure pilgrim safety. The research aims to: (1) develop image analysis models using CNNs, SVM, and Random Forest to improve crowd density detection accuracy; (2) enhance the precision of pilgrim count and density estimates for better crowd management; (3) compare the performance of the four models to identify the most accurate in detecting movement patterns; and (4) provide recommendations to support future research and improve crowd management strategies. This work contributes to the field of image analysis using machine learning and offers practical insights for managing large-scale events.**

**Keywords: Crowd management, Hajj, Umrah, CNN, SVM**

## INTRODUCTION

Mount Arafat, also known as Jabal al-Rahma, is located southeast of Mecca and is one of the most significant religious sites in Islam. Standing at Arafat is a fundamental pillar of Hajj, where pilgrims gather on the $9^{th}$ of Dhu al-Hijjah to supplicate, recite the Quran, and pray. This day is considered the holiest day of Hajj, as sins are forgiven, and souls are freed from the fire.

The Ministry of Hajj and Umrah in Saudi Arabia is leading the way in integrating advanced smart systems to enhance the pilgrimage experience for millions of Muslims. A key focus of these efforts is facilitating the smooth performance of rituals, including the journey to Jabal Al-Rahmah (Mount of Mercy) in Arafat—one of the most spiritually significant sites during Hajj. The integration of smart systems has enabled real-time tracking, helping to guide pilgrims and alleviate crowd congestion during the pilgrimage.

Hajj, a pillar of Islam, requires all financially and physically capable Muslims to undertake the pilgrimage to Mecca at least once in their lifetime. When performed sincerely and without committing sinful acts, Hajj results in the forgiveness of sins. Similarly, Umrah, though not obligatory, is a profoundly spiritual journey associated with purification and divine blessings. Each year, millions of Muslims travel to Saudi Arabia to fulfill these sacred duties.

As the world's largest annual human gathering, Hajj presents significant challenges in crowd management, particularly during key stages like the journey to Jabal Al-Rahmah. To address these challenges, the Ministry has developed innovative solutions, such as leveraging real-time data to guide pilgrims along less congested routes. These advancements ensure the safety, comfort, and spiritual focus of pilgrims, showcasing how technology is revolutionizing both the logistical and spiritual dimensions of Hajj.

## LITERATURE REVIEW

Previous studies highlight the significance of automated techniques in managing large gatherings like Hajj, noting challenges such as overcrowding, limited real-time monitoring capabilities, and the complexity of ensuring safety in highly dynamic environments. For instance:

Bhuiyan et al. [1] proposed the Deep Hajj Crowd Dilated Convolutional Neural Network (DHCDCNNet), which integrates advanced augmentation techniques to create additional datasets tailored for Hajj scenarios. Their approach achieved high accuracy rates across multiple datasets, including 100% for the proposed Hajj-Crowd dataset using FCNN, VGGNet [2], and ResNet50 methodologies. This underscores the importance of robust models for analyzing high-density crowd images, overcoming challenges like extreme occlusion and perspective difficulties.

The proposed research [2] introduces a robust CNN-based model for Hajj crowd counting and detection. The model predicts head locations in crowd images through an end-to-end system. Key processes include frame extraction from videos, multi-scale spatial feature extraction using a CNN-based prediction map, and density estimation via non-maximum suppression (NMS). The CNN architecture employs multi-layer convolution, adaptive feature maps, and scale-specific branches to handle diverse crowd scenarios, ensuring accurate crowd density predictions and high detection precision.

The proposed research [3] presents an AI-based crowd management framework for monitoring and predicting incidents during Umrah. It addresses key issues like crowd counting and behavior monitoring, offering quick response through anomaly detection. The framework is evaluated for efficiency, scalability, and accuracy, showing its potential for use in various.

## DATA EXPLORATION AND DESCRIPTION

Overview The dataset consists of manually collected images from various online sources, focusing on scenes at Jabal al-Rahmah during Hajj. Each image captures varying levels of crowd density, categorized into two classes: "Crowds" and "No Crowds." The total dataset contains 100 images, with an imbalanced class distribution—95 images labeled as "Crowds" and 20 images labeled as "No Crowds." This dataset is designed for crowd analysis and classification, offering an opportunity to explore density patterns and enhance automated image recognition techniques.


Fig. 1. Sample images from the dataset showing “crowds” (top) and “no crowds” (bottom) for analysis.

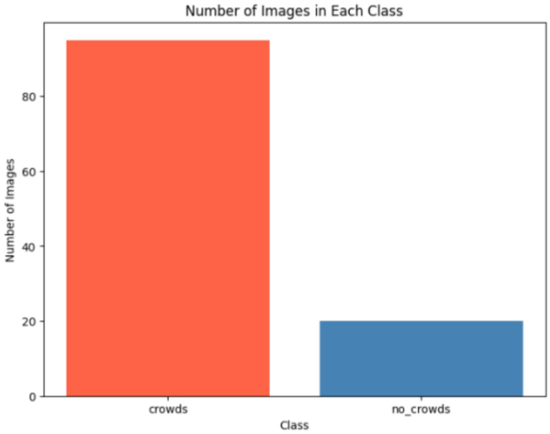


Fig. 2. Bar chart showing the distribution of images in the dataset.

## IMAGE CHARACTERISTICS

The images have been resized to dimensions such as 1024x1024 pixels for efficient training and processing, while some remain at 612x612 pixels, which may require additional preprocessing to standardize dimensions. All images are in JPG format, a widely used format for photographs that balances good quality with manageable file sizes. This dataset provides valuable opportunities for analyzing crowd patterns and enhancing automated image classification systems.

## VARIABILITY CHALLENGES

One of the main challenges in this dataset is the variability in crowd density. It was difficult to determine, based on visual inspection alone, whether a scene represented a dense crowd or not. The subtle differences in crowd size and arrangement posed a challenge in accurate classification, making the task of labeling images more complex. This variability adds to the complexity of building a reliable model to differentiate between the "Crowds" and "No Crowds" categories.

## METHODOLOGY

Data preprocessing is crucial for preparing the dataset for effective model training. The following techniques were employed:
**1. Image Resizing:** All images were resized to 224 × 224 pixels to standardize input dimensions across models, ensuring consistent processing and efficient training. This size is commonly used in deep learning, particularly in Convolutional Neural Networks (CNNs), as it strikes a balance between image quality and data size, helping preserve important details while reducing processing time and memory usage.

**2. Data Augmentation**: To increase dataset variability and reduce overfitting, several augmentation techniques were applied, including:

- **Rotation**: Randomly rotating images within a specified degree range.
- **Flipping**: Horizontally flipping images to introduce variability.
- **Grayscale Conversion**: Converting images to grayscale to simplify features.

These preprocessing steps aimed to improve model generalization and performance by diversifying the training data.


Fig. 3. Preprocessed images from the dataset.

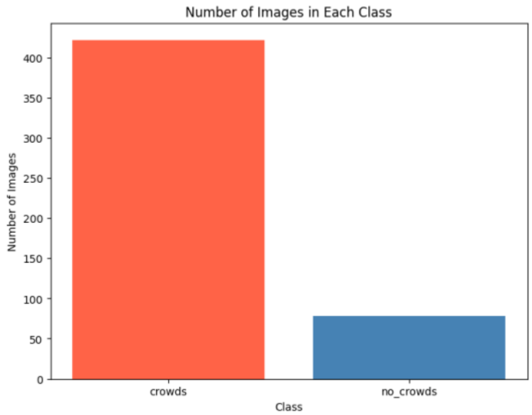


Fig. 4. Bar chart showing the number of images in each class after preprocessing.

After applying data augmentation and preprocessing techniques, the image illustrates the class distribution. The "crowds" class contains a significantly higher number of images compared to the "no crowds" class, reflecting the impact of data augmentation and preprocessing on the dataset.

**Assumption of solving this challenge using machine learning**

**Convolutional Neural Networks (CNNs)** are specialized deep learning models designed for image processing and computer vision tasks. Their architecture includes convolutional layers that excel at automatically extracting essential features from images, such as edges, shapes, and patterns. This feature extraction process is crucial for analyzing visual data and allows CNNs to identify and distinguish elements within images effectively.

When applied to crowd analysis, CNNs leverage their powerful feature extraction capabilities to recognize varying densities of people in images. This is particularly relevant in scenarios such as crowd management during events at Mount Arafat, where precise monitoring of crowd movements is essential for safety. By training on labeled datasets, CNNs can learn to differentiate between images with dense crowds and those with fewer individuals.

The training process of CNNs is efficient due to their reduced number of parameters compared to traditional models. This allows for faster training times and quicker convergence to accurate results. As a result, CNNs can provide real-time insights, enabling authorities to make informed decisions swiftly.

In the context of crowd management, CNNs can analyze images captured during high-traffic events at Mount Arafat to detect potential risks, such as overcrowding or unusual behavior. Their ability to distinguish between individuals, pathways, and surrounding elements enhances risk prediction and facilitates timely interventions. For instance, if a sudden gathering occurs, CNNs can issue early warnings, allowing for immediate crowd direction and safety measures.

**Random Forest** is a powerful machine learning algorithm that excels in tasks like classification by constructing multiple decision trees based on various features. When applied to image data, Random Forest is not directly responsible for processing raw pixel data like Convolutional Neural Networks (CNNs), but instead works by extracting important features such as color, texture, and shape from the image. Each decision tree in the random forest uses these features to make independent predictions, and the final classification is determined by aggregating the outputs of all the trees.

Random Forest is invaluable for analyzing images of crowded areas. By extracting features like density and shape from the images, Random Forest can classify regions as either crowded or non-crowded, aiding in real-time decision-making for crowd management.
1-High Accuracy: Random Forest effectively identifies crowded spaces based on attributes like the number of people and spatial configuration. 2-Resilience to Noise: The algorithm maintains performance despite environmental changes, such as variations in lighting and image quality, which is vital in dynamic settings.

By evaluating the importance of various features, Random Forest prioritizes key elements that improve the accuracy of crowd detection.

**SVM (Support Vector Machine)** is a machine learning algorithm that excels in classification and regression tasks by finding the optimal boundary between different classes in the data. It is especially effective in handling non-linear data using the kernel trick, making it well-suited for image classification and complex datasets.

SVM can be applied to detect and segment crowds in still images taken from Jabal Arafat, a crucial site during Hajj. By identifying repetitive textures and patterns that differentiate the crowd from the background, SVM effectively segments the image into crowd and non-crowd regions. Filters can be employed to capture different orientations and scales of these textures, which improves the detection of crowd-specific features. The SVM

classifier then clusters the regions based on these features, ensuring precise identification of areas with high crowd density. This technique plays an essential role in crowd management during Hajj.

## HYPERPARAMETERS TUNING

- Learning Rate in CNN: Set to 0.001 for optimal convergence.
- Image Resizing: The target size of all images was changed from 1024×1024 pixels to 224 × 224 pixels to improve training speed and model efficiency.
- Batch Size: 32 images per batch to balance training speed and memory usage.
- Epochs: Models were trained for 10 epochs, with early stopping criteria based on validation loss to prevent overfitting.
- Training and Validation Split: The dataset was divided into 80% training and 20% validation sets to monitor performance during training.

## EVALUATION METRICS

Model performance was evaluated using the following metrics:

- **Accuracy**: The proportion of correctly classified images.
- **Confusion Matrix**: To visualize the performance across different classes.
- **Classification Report**: This included precision, recall, and F1-score for a comprehensive assessment of model performance.

## RESULTS

### (CNNs) Model Performance

### Classification Report

| Category | Precision | F1-Score | Recall |
|---|---|---|---|
| *crowds* | 0.85 | 0.86 | 0.87 |
| no crowds | 0.15 | 0.14 | 0.13 |
| *Accuracy* | | 76 | |

**Crowds Class (High-Performance Class):**

- Precision: **0.85** indicates that out of all images predicted as "crowds, " are accurate.
- Recall: **0.87** reflects that the model correctly identifies 0.87 of actual crowd images.
- F1-Score: **0.86** balances precision and recall, showcasing a reliable performance for this class.

**No Crowds Class (Low-Performance Class):**

- Precision: **0.15** highlights the model's difficulty in correctly identifying non-crowd images.
- Recall: **0.13** shows that only a small proportion of actual non-crowd images are correctly classified.
- F1-Score: **0.14** suggests that improvements are needed to enhance the detection accuracy for non-crowd regions.

**Overall Accuracy**: 76% - this performance is heavily skewed by its strong results in the "crowds" category and poor performance in the "no crowds" category.

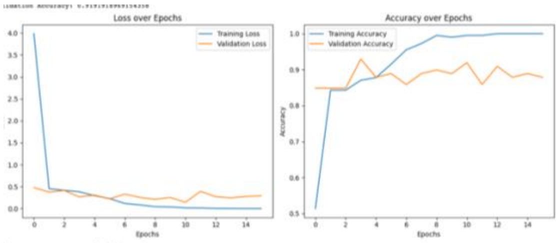

Fig. 5. The loss and accuracy of CNN model over epochs.

**CNN Model Performance Over Epochs**

The figure illustrates the performance of the Convolutional Neural Network (CNN) model over training epochs. The loss decreases steadily for both training and validation data, indicating effective learning and stability without overfitting. Similarly, the accuracy improves on the training data and remains stable on the validation data, reflecting the model's efficiency in classifying unseen data.

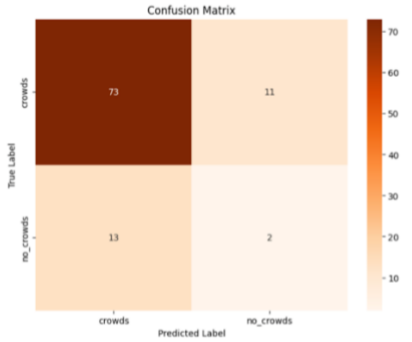


Fig. 6. CNN model confusion matrix.

The confusion matrix shows the model correctly classified 73 "crowds" and 2 "no crowds" but misclassified 11 "crowds" as "no crowds" and 13 "no crowds" as "crowds. " The model excels at detecting "crowds" but struggles with "no crowds," likely due to class imbalance or insufficient features. Balancing the dataset and enhancing feature extraction are suggested improvements.

(Random Forest) Model Performance

## Classification Report

| Category | Precision | F1-Score | Recall |
|---|---|---|---|
| *crowds* | 0.85 | 0.92 | 1.00 |
| no crowds | 0.00 | 0.00 | 0.00 |
| *Accuracy* | | 85 | |

**Crowds Class (High-Performance Class):**

- Precision: **0.85** - The model correctly identifies 85% of the instances labeled as "crowds."
- Recall: **1.00** - All instances of "crowds" in the dataset are identified (no false negatives).
- F1-Score**: 0.92** - Indicates a high balance between precision and recall for the "crowds" category.

**No Crowds Class (Low-Performance Class):**

- Precision, Recall, and F1-Score for "no crowds" are all **0.00**, indicating the model completely fails to classify this category.

**Overall Accuracy**: Accuracy: 85% - The model correctly classifies 85% of all instances, but this high accuracy is heavily influenced by the model's strong performance in the “crowds" category and complete failure in the "no crowds" category.

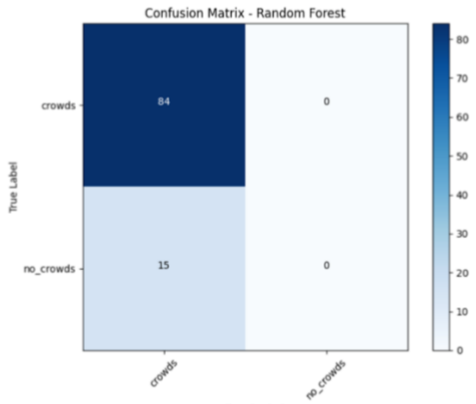


Fig. 7. Random forest confusion matrix.

The confusion matrix shows the model correctly classified 84 "crowds" but misclassified 15 "no crowds" as "crowds." The model excels at detecting "crowds" but completely fails to detect "no crowds", likely due to class imbalance or insufficient features. Balancing the dataset and enhancing feature extraction are suggested improvements.

(SVM) Model Performance

## Classification Report

| Category | Precision | F1-Score | Recall |
|---|---|---|---|
| *crowds* | 0.82 | 0.88 | 0.95 |
| no crowds | 0.64 | 0.42 | 0.31 |
| *Accuracy* | | 80 | |

**Crowds Class (High-Performance Class):**

- Precision: **0.82** - The **precision** is strong; it is indicating relatively few false positives.
- Recall: **0.95** - is very high (0.95), meaning the model effectively identifies most actual crowd regions with very few misses (false negatives).
- F1-Score: **0.88** - It is highlighting the balance between precision and recall, making this a high-performance class.

**No Crowds Class (Low-Performance Class):**

- Precision: **0.64** - This is classified as a low-performance class due to the imbalance in precision and recall.
- Recall: **0.31** - is particularly low, meaning the model misses many actual no-crowd (high false negatives).

- F1-Score: **0.42** - is significantly lower than for the crowds class, indicating weak overall performance for this category.

**Overall Accuracy**: Accuracy: 80% While the accuracy is decent, the imbalance between the performance of the two classes suggests that improvements are needed, particularly for the "No Crowds" class, to make the model more reliable.

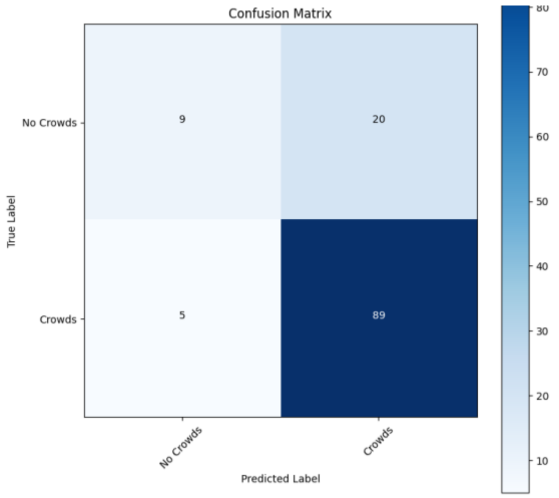


Fig. 8. SVM confusion matrix.

The confusion matrix shows that the model performs well in identifying the "Crowds" class, correctly classifying 89 instances, with only 5 misclassifications as "No Crowds. " However, the model struggles with the "No Crowds" class, correctly classifying only 9 instances while misclassifying 20 instances as "Crowds." This indicates a strong performance for detecting "Crowds" but significant difficulty in distinguishing "No Crowds"

## Model Comparison for Image Classification

**CNN Model:** The CNN model was shown to be able to accurately identify images containing "crowds" It performed well in classifying these images, successfully identifying most of the correct instances. However, the main problem arose with images without crowds, where the model had difficulty classifying them effectively. This was primarily due to a large imbalance between classes, where the "crowd" class was more frequent than the "no crowd" class, causing the model to focus more on detecting crowds. Although attempts were made to address this imbalance through various data balancing techniques during training, further improvements are needed.

**Random Forest Model**: The Random Forest model excelled at identifying the "crowd" class, with no misclassifications of crowd images. However, it had a major problem with the "no crowd" class, where it failed to classify these images completely. This problem can be attributed to the structural nature of the model, which resulted in a bias towards the "crowd" class. Despite its success in classifying crowds, the failure to identify "no crowds."

**SVM Model:** The SVM model was effective in recognizing images containing crowds, showing high accuracy in this regard. However, the "no crowd" category was a weak point, as the model failed to classify many of these instances. This suggests that the model was not robust enough to accurately identify the less frequent category. Despite these challenges, overall performance was still decent, with good ability to classify crowd images. However, improving the "no crowd" category classification remains critical.

## Challenges and opportunities for improvement

Class imbalance: The main issue with all models was the unbalanced data distribution, with the "crowd" category making up the majority of the data.

This imbalance negatively impacted the models' ability to classify the "no crowd" category. It is important to achieve better data balance, either by increasing the "no crowd" category or assigning it higher weights during training.

**Improving feature extraction:** To improve model performance, it is essential to improve how features are extracted from images. Advanced techniques such as using deep neural networks or methods such as principal component analysis for image analysis can help the model better distinguish between different classes.

**Future improvements:** While techniques such as data augmentation have been used to address class imbalances, there are greater opportunities to improve performance by using more robust models that are better suited to image processing through deep learning. Additionally, incorporating temporal context or using a time series (such as the times images were taken) can help the model recognize patterns of increased crowd density during specific times of the day. This will allow the model to learn not only from the images themselves but also from the temporal context in which they were taken.

## CONCLUSION

The use of artificial intelligence techniques such as Convolutional Neural Networks (CNN), Support Vector Machines (SVM), and Random Forest in analyzing crowd images is an important step toward improving crowd management and organization in areas like Jabal Al-Rahmah during Hajj.

Although the results achieved were good, there is room for improvement in classification accuracy, particularly in less frequent crowd categories. In the future, we aim to enhance these models to make them more efficient and effective, improving their ability to handle the challenges posed by dynamic crowd environments. Enhancing these techniques will contribute to better and faster crowd management, promoting the safety of pilgrims and ensuring a smoother and more efficient Hajj season.

## ■ REFERENCES

[1] Bhuiyan, R. et al. (2022) Deep dilated convolutional neural network for crowd density image classification with dataset augmentation Hajj pilgrimage, Sensors (Basel, Switzerland). Available at:https://pmc.ncbi.nlm.nih.gov/articles/PMC9320336/

**[2]** *Transfer learning using VGG-16 with deep convolutional neural network for classifying images*. Available at: https://www.ijsrp.org/research-paper-1019.php?rp=P949194 .

[3]Hajj pilgrimage video analytics using CNN (no date) Semantic Scholar. Available at: https://www.semanticscholar.org/reader/69b5a9930e7daf5e481793e08b77652f50c20008

[4] Ali, H., Khan, M., and Ahmed, S., 2023. Crowd Management Intelligence Framework: Umrah Use Case. Proceedings of the IEEE International Conference on Artificial Intelligence and Data Analytics (CAIDA). Available at: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=10381717)phy