SSRG International Journal of Electrical and Electronics Engineering
ISSN: 2348-8379/ https://doi.org/10.14445/23488379/IJEEE-V13I4P101

Volume 13 Issue 4, 1-19, April 2026

Original Article

A Deep Learning based Framework for Actor Recognition and Screen Presence Analysis in Bollywood Films

Snehal Athghara[1], Apoorva Atre[2], Shweta Bagade[3], Gayatri Muttepawar[4], Shilpa Deshpande[5], Rashmi Apte[6], Mangesh Bedekar[7]

[1,2,3,4,5]*Computer Engineering Department, Cummins College of Engineering for Women, Pune, Maharashtra, India.*
[6]*Koushiki Innovision, Pune, India.*
[7]*School of Computer Science and Engineering, Dr. Vishwanath Karad, MIT World Peace University, Pune, India.*

[5]*Corresponding Author : shilpa.deshpande@cumminscollege.in*

Received: 01 January 2026 Revised: 08 February 2026 Accepted: 10 March 2026 Published: 30 April 2026

Abstract *- Quantitative and qualitative analysis of actor screen presence is a major aspect of film studies, media analysis, and performance evaluation. Existing approaches largely rely on manual annotation and have insufficient robustness under real-world cinematic conditions. This research proposes Actor Recognition and Screen Presence Analysis (AR-SPA), a deep learning-based framework designed to automate actor identification and screen-time analysis across full-length movies. AR-SPA operates by uniformly sampling frames and implementing a face-detector to localize actors using the You Only Look Once (YOLO) model. A discriminative recognition model is used to identify actors across the frames. This is achieved by experimentation on three different recognition models, namely, Convolutional Neural Network (CNN), CNN-Transformer, and Residual Network-Convolutional Block Attention Model (ResNet-CBAM). Experimental results highlight the tendency of the ResNet-CBAM model to consistently outperform the other conventional baseline models by achieving high Accuracy and F1-score across multiple testing conditions. Hence, this model is integrated into the AR-SPA framework for robust actor recognition. By aggregating these classified actor detections over the film's duration, the framework generates metrics such as total frame counts, proportional screen time, and actor dominance. This enables direct statistical comparison of actor prominence across sequels and long-running franchises. The framework is validated using a curated dataset of fifty films, from twenty movie series spanning across two decades. Through rigorous testing and validation, AR-SPA demonstrates high Accuracy and reliability in the face of challenges such as aging, dramatic lighting, and occlusions. The results of this research suggest that AR-SPA offers a scalable, reproducible tool for film scholars and industry analysts to track character evolution and performance trends over time.*

Keywords *- Convolutional Block Attention Module, Face Detection, Long-Form Video Analysis, Movie Series Analysis, Residual Network, Temporal Screen-Time Estimation.*

1. Introduction

There is a huge growth in the digital media industry, leading to the generation of massive amounts of visual content. Different film industries have gained popularity worldwide, contributing to the generation of movie data. Bollywood, constituting India's mainstream Hindi language cinema, is one of these industries. It is one of the largest film industries globally, with hundreds of films being released every year. These releases vary across diverse setups and locations, dramatic storytelling, and visually complex scenes. This eventually leads to a huge amount of Bollywood video data being produced. Each movie also differs in terms of cast and success at the box office. With the increasing availability of full-length movies on digital and streaming platforms, there is a growing interest in computational film analysis.

This particularly involves understanding the influence of visual elements on audience perception and film reception [1, 2]. While there are various aspects to film analysis, the correlation between the screen presence of an actor in the film and its effect on the audience is notable in many aspects. Screen presence plays a very significant role in shaping the narrative focus as well as audience attachment to the stars and their characters in the movies. Manually calculating the screen time of each actor in the movie is very time-consuming, error-prone, and impractical. This gives rise to the need for an automated, reliable, and robust framework that could make use of advanced methods in face detection and recognition with justifiable accuracy. Such a solution would be of great use to production houses, movie critics, film analysts, film researchers, and academicians, among others, for analyzing

the relationship between an actor's screen presence and audience engagement. This work would make it possible for all such stakeholders to provide visual and data-backed verdicts on a film's performance. Streaming platforms can also make the best use of this work to classify, organize, and recommend films to their users based on their favourite artists.

1.1. Research Gap

While significant research is carried out in the domain of face recognition and video analytics, Bollywood still remains an under-explored industry in this actor-centric analysis. Most existing actor recognition frameworks are tested predominantly on Hollywood movies and are not optimal. They rely on pretrained deep learning models such as FaceNet [3] and DeepFace [4]. These models are biased towards Western actors' datasets, since they are primarily trained on related datasets. They have been shown to exhibit demographic bias when applied to Indian actors [5-10]. Bollywood movies pose additional challenges due to the dramatic makeup, frequent costume changes, theatrics, and crowded scenes, aging across sequels, and significant changes in lighting and motion. Prior studies on video-based face recognition indicate that such unrestrained conditions considerably affect the recognition accuracy [11, 12].

Moreover, very limited datasets of Bollywood actors are readily available on the internet, posing a major challenge in developing reliable Bollywood-based image recognition systems. Furthermore, several state-of-the-art actor recognition frameworks achieve high accuracy, but are computationally heavy [6, 13-16].

This makes the frameworks unsuitable for long-duration video processing, thus rendering them impractical. Conversely, frameworks that are lightweight and time-efficient often compromise robustness under scenarios with varying lighting conditions, occlusions, and multiple poses that are quite common in Bollywood [6, 16-20]. Existing approaches, which rely on shot-based or event-centric metrics for screen presence analysis, fail to provide accurate quantitative results. This is attributed to the fact that shot durations vary significantly and do not directly correspond to the actual screen time [7]. Attention-based models identify participants based on event-based relevance rather than visual presence, hence limiting their application for actor screen presence measurement in movies [21]. This further highlights the requirement of a Bollywood-oriented framework that achieves a balance between speed and Accuracy in detecting and tracking actors in movies.

1.2. Need for Solution and Contribution

The quantification and analysis of the screen presence of actors in Bollywood films is significant in the domain of film analysis. Existing approaches often prove insufficient in specific areas and may not provide a comprehensive analysis of actor presence in movies. The lack of sufficient datasets available for Bollywood actors also contributes to the issues present in the film analysis domain. The datasets that are available often suffer from a lack of images for a custom recognition model training or low-quality images. Additionally, the statistical results obtained by such film analysis are a significant contribution to the film industry. With an increase in movie analysis performed in the Hollywood industry, it is essential to devise new approaches to contribute to Bollywood movie analysis.

To fulfill the requirement of a framework that identifies and recognizes Bollywood actors despite varying contexts and conditions, the Actor Recognition and Screen Presence Analysis (AR-SPA) framework is proposed. This framework integrates the deep learning-based face detection and recognition models for automated actor recognition. AR-SPA uses a supervised model for face recognition, training of which is done using a curated dataset containing 48 actors with about 22,000 images. The actor-detected frames obtained from the recognition model are fed into a statistical screen-presence analysis engine. Thus, the final output of AR-SPA is a visual analysis of the presence of constant actors throughout the movie. In this research, the AR-SPA framework is tested on 20 distinct Bollywood film series. A film series is a collection of related movies. These movies are released in succession and share the same universe, often consisting of recurring casts. For each series, AR-SPA gives a comprehensive overview of the screen presence of the lead actor and two supporting actors.

The main contributions that the proposed work offers are outlined below.

1. A domain-specific framework - To address the need for an automated and reliable solution for quantification of actor screen presence, the AR-SPA framework is proposed.
2. A curated Bollywood actor dataset - The AR-SPA dataset, used for a custom actor face classification model, is proposed. It is a custom dataset of 48 Bollywood actors, consisting of approximately 22,000 images, manually gathered from different resources. It is gathered keeping in mind the age, makeup, and costume variations in the actor's appearance.
3. A comparative model evaluation - This work highlights the overall comparison of three deep learning recognition models and identifies the best of those through experimentation and performance analysis.
4. Screen presence estimation - For the best identified model, this work provides the actor-centric screen presence metrics, including total detected frames, estimated temporal screen presence, and percentage contribution. This enables narrative and longitudinal analysis across individual movies and movie series.
5. Novel Actor Dominance Metrics for Narrative Analysis - This research proposes novel actor dominance metrics,

viz. Actor Dominance Ratio (ADR) and Dominance Shift indicators. These metrics quantify relative actor dominance within individual movies and across movie series. They enable the identification of narrative patterns such as stability of lead presence, emerging new supporting actors, and role transitions, which are not evident from screen-time alone. These metrics are further elaborated in section 4.2.1.

2. Literature Review

The domain of automated actors' detection and quantification of their screen presence has been extensively researched due to its use in various video-related applications and studies [22].

Previous work is more focused on face detection and spectral or graph-based clustering techniques, whereas in recent years, work is more focused on deep learning-based frameworks. However, these studies are primarily tested on Hollywood movies and actors.

Earliest approaches include spectral clustering techniques, which form identity clusters without naming the characters explicitly [23, 24]. Similar works using clustering include dynamic character graph formation via online face clustering [13] and clustering based on similarity scores [25]. These techniques require manual intervention for cluster identification and suffer from the possibility of noisy clusters. Work suggested in [5] uses face detection and tracking with Erdős-Rényi clustering to group actor faces to their character identities by creating a unique ID. Although this helps in reducing false positives and addressing missing detections, the approach struggles with unnamed and background casts with insufficient appearance for clustering. Several studies have used supplementary information like movie scripts and cast lists to improve actor identification, like the work in [14], which uses graph-based models that map the relation between detected faces and character names from the script and other information. Further, the approach by [26] uses generative appearance models to detect actors from movie scripts, and the Cast2Face framework [27] uses actor and face relation by propagating labels and using movie cast metadata. However, these approaches rely heavily on good computational power and robust textual resources. Work proposed in [6] demonstrates actor identification without using show scripts by using images available on the internet, and work in [28] uses self-supervised feature adaptation to have a robust character labeling, but it is computationally intensive for long videos or movies.

Later studies are using deep learning strategies such as Convolutional Neural Network (CNN) [17] for actor identification, computer vision for video image processing [18], and the DeepStar framework [7], which uses DeepFace [4] and appearance frequency to detect actors. Work suggested in [8] addresses the large volume of multimedia by proposing a robust actor recognition pipeline, which is further refined by [9]. Although these works propose scalable pipelines, they are trained and evaluated on a dataset focusing on Western media. Authors in [29] proposed a solution that jointly optimizes representation and classification for robust object tracking, but does not explore actor face tracking. Studies like the Video Character Tracker (ViCTer) framework [19] focus on semi-supervised and self-supervised learning techniques and combine face detection with multi-character tracking, but they still rely on clear face visibility and struggle with actor appearance change. Graph-based label propagation techniques used by [30] reduce manual effort but are susceptible to error in long videos. Work proposed in [21, 31] and the Watch Only Once (WOO) framework [32] integrate actor and action recognition. Automatic annotation of human actions as suggested in [33] further automates the task but lacks robustness to complex backgrounds and camera motion.

Studies like actor-based Internet Protocol Television (IPTV) indexing framework [34] or mobile device-oriented systems [35] focus on domain-specific deployments. Authors in [20] present a tool for detecting characters from popular TV shows, and work in [15] addresses character detection in animated movies. Studies such as [16, 36] focus on practical system implementation, highlighting Graphics Processing Unit (GPU) based acceleration and integration of existing face recognition libraries, with limited emphasis on developing new recognition algorithms. Actor screen time analysis is explored in [37, 38], but they focus on entertainment videos and Bollywood music videos. Screen Time and Actor Recognition (STAR) framework [39] and work proposed in [40] provide a baseline by introducing a pipeline for face detection, tracking, and recognition using vision-based analysis and algorithms incorporating machine learning; they heavily rely on labeled datasets focused on Hollywood media.

In summary, the above literature review highlights that existing approaches exhibit several limitations. Most of the work is evaluated on Western-centric datasets such as Hollywood movies and TV shows, limiting their generalizability to regional cinema. The systems also lack a fully end-to-end architecture and rely on supplementary textual resources or are restricted to short video clips rather than full-length movies. Furthermore, while some studies address actor prominence, quantitative screen-time analysis is often limited to trailers or isolated segments, failing to scale to long videos like full-length movies. Collectively, these gaps highlight the need for a unified framework that combines accurate actor recognition with scalable, quantitative screen-presence analysis, particularly for underrepresented film industries like Bollywood.

Building upon the existing frameworks, this research proposes Actor Recognition and Screen Presence Analysis (AR-SPA), a unified framework that analyzes the screen

presence of actors in Bollywood films. This framework automatically identifies the target actors present in the movie frame. With these detections across the movie length, the screen presence of the actors is estimated.

A custom Bollywood actor image dataset is used due to the lack of sufficient datasets available publicly. The screen presence of the actors is quantified using various metrics, including two novel Actor Dominance metrics. Overall, the AR-SPA framework provides a comprehensive insight into screen presence analysis and contributes to film analysis.

2.1. Novelty of AR-SPA Framework

As summarized in Table 1, existing work is biased towards Western media and does not provide a comprehensive solution. The Indian Movie Face Database [41] gives a good foundation of a Bollywood-focused dataset, and an end-to-end pipeline for actor recognition and screen presence estimation is missing. The proposed AR-SPA framework introduces a custom-made dataset focusing on the Bollywood industry and comprising images of 48 Bollywood actors and 50 full-length Bollywood movies. Existing works discuss actor recognition and screen presence quantification as separate modules, so a unified framework with an end-to-end pipeline is missing. The proposed framework addresses this issue by presenting a complete pipeline that detects, recognizes, and quantifies the actor screen presence across Bollywood movies, as highlighted in the last row of Table 1. Additionally, by combining You Only Look Once (YOLO) based face localization with a Residual Network-Convolutional Block Attention Model (ResNet-CBAM) architecture, the AR-SPA framework effectively handles challenges such as aging effects, heavy makeup, lighting variations, and stylistic diversity common in Bollywood movies. Thus, the novelty of this work lies not only in architectural design but also in its methodological completeness, dataset originality, long-form analytical capability, and focus on underrepresented regional cinema, collectively addressing critical research gaps identified in existing literature.

Table 1. Qualitative Comparison of Related Work and Proposed Framework

Reference Group	Actor Face Recognition	Screen Presence Quantification	Movie Series / Long-form Analysis	Dataset Used
Face Clustering Based Methods [5, 13, 24]	Yes	No	No, work done on short video segments or isolated scenes	Western TV shows and Hollywood movie clips
Face-Name Association and Movie Script-Based [14, 26, 27]	Yes	No	Yes, but conditional on the availability and accuracy of textual resources	Hollywood movies, their scripts, and movie cast lists
Deep Learning Based Actor Recognition [7, 8, 17]	Yes	No	Yes, but do not explicitly compute screen presence statistics	Hollywood movies and TV shows
Screen Time Estimation Frameworks [39, 40]	Yes	Yes	No, work restricted to movie trailers and small movie clips	Western TV shows and Hollywood movie trailers
Proposed AR-SPA framework	Yes	Yes	Yes, work is evaluated on full-length Bollywood movies	Custom Bollywood dataset consisting of images of 48 actors and 20 movie series

3. Materials and Methods

3.1. Dataset Creation

3.1.1. Need for Dataset Creation

The Dataset acts as a foundation for research and analysis. As discussed in section 1.1, there are limited datasets of Bollywood actors' images available. This leads to lesser exposure and research with fewer analyses of Bollywood movies and trends in this industry, serving as a strong reason to tailor a Bollywood actor's specific dataset.

Such a dataset that considers factors like actors' age, makeup, and costume variations and is well annotated would be of great help to the research community and analysts. To fulfill this requirement, a new dataset creation is paramount.

3.1.2. Data Collection

This research makes use of a dual-component database, consisting of an image-based actor face dataset and a video-based movie dataset. This dual-dataset is essential for effectively training the actor recognition model and realistic evaluation of the screen presence analysis of the actors in various cinematic environments. The dataset comprises approximately 22,000 images across 48 Bollywood actors, with each actor represented by 400-500 images. It captures various real-world variations such as multiple angles, illumination conditions, expressions, aging changes, and makeup styles, reflecting the visual diversity typically encountered in Bollywood films. This is done by manually capturing the images from various publicly available videos like Bollywood actors' interviews, promotional content, and

advertisements, and movie clips available on YouTube. Additional images have been collected from Google Images, where visual material is sourced from various websites and media outlets. A limited number of images are also incorporated from publicly available Kaggle datasets; however, these collections are incomplete and cover only a small subset of Bollywood actors, lacking sufficient diversity. Figure 1 illustrates the schema of the AR-SPA face dataset. The movie dataset comprises 50 full-length movies from 20 distinct film franchises, each featuring multiple installments of the same storyline or characters. The movies used for this research are primarily released from the early 2000s to recent years. The series selection for this study is made keeping in mind that the primary cast, including the main lead and co-leads, remain consistent across the sequels. This makes the series suitable for longitudinal screen presence analysis. The movies used for testing are sourced from publicly accessible platforms, including official entertainment channels such as YouTube [42] and archival resources such as Internet Archive [43].

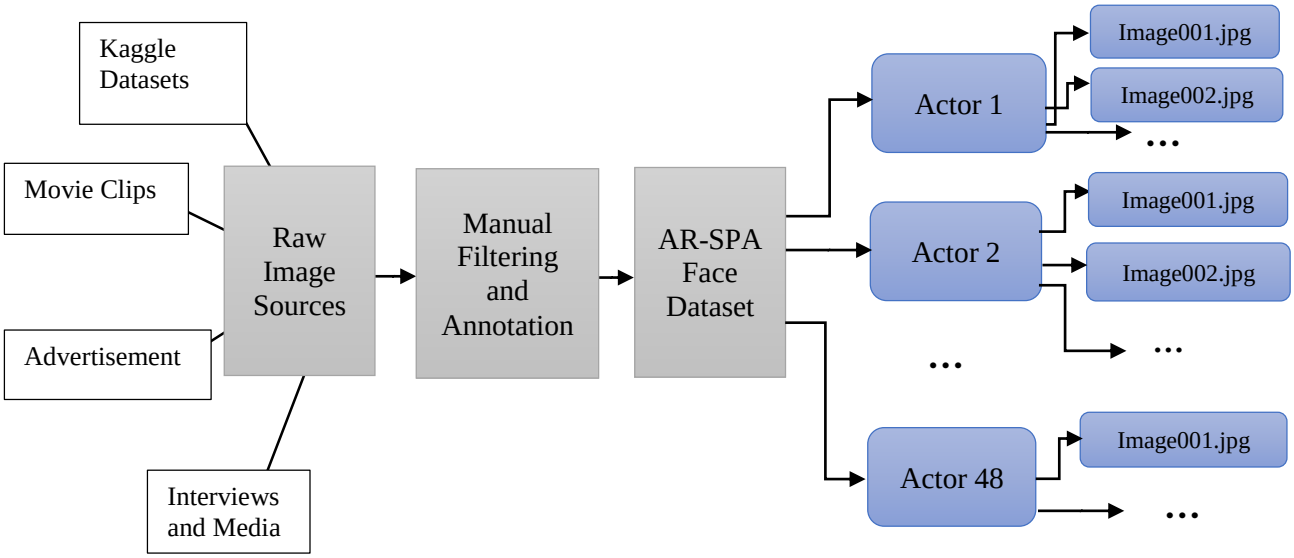


Fig. 1 AR-SPA face dataset schema

3.1.3. Data Preprocessing and Annotation

As discussed in section 3.1.2, the AR-SPA dataset is partially collected from public sources. Some of these images are observed to be blurry, dark, low resolution, or with occlusions. Such inconsistent and faulty images act as a barrier to the efficiency of actor recognition. This makes it vital to refine the dataset and include more appropriate images for a sufficiently large and good-quality dataset. To address these issues, following the dataset collection stage, the images undergo preprocessing and annotation. This step ensures data quality and consistency for actor recognition model development. The images that are collected from multiple sources are manually inspected to ensure the quality of the dataset.

This includes filtering duplicate entries, severely blurred frames, frames with extreme occlusions, and frames that lack sufficient discriminative information. All the images in the dataset are subsequently annotated with the respective actors. The images corresponding to a specific actor are organized into actor-specific directories, resulting in a sorted, well-labelled dataset of all actors. Images are manually verified to ensure that the images with makeup variations, age-related appearance changes, and other factors are filtered out. This step is particularly essential for Bollywood films as they have extreme diversity across the movies and actors. This labelled dataset is then standardized by resizing and normalizing the images for data consistency and generalization. The new image resolution after resizing the images is 224*224 pixels, and the cropped area ranges from 70-100% of the original image. They are normalized using the mean and standard deviation of the red, blue, and green channels of the image. The respective mean is subtracted from each channel, and the result is the proportion of the respective standard deviation. The set of values used for the mean and standard deviation of the red, blue, and green channels, respectively, in this research is (0.485, 0.456, 0.406) and (0.229, 0.224, 0.225). Augmentation of data is performed using basic techniques like horizontal flipping and brightness variations. This allows the model to become generalized and adapt to the varying cinematic conditions. This dataset can further be readily used in supervised learning.

3.1.4. Dataset Characteristics and Analysis

The dataset constructed reflects real-world cinematic conditions and shows diversity across spatial and temporal dimensions. The face dataset captures intra-class variability in images caused by age, hairstyle, makeup, facial hair,

accessories, and expressions, as well as inter-class similarity among actors with comparable facial features.

The movie dataset consists of a wide range of scenarios, such as multi-actor scenes, dynamic camera movements capturing close-range and wide-angle shots, occlusions, differences in illumination, and resolution changes. Actors appear throughout the movies in various costumes and environments, thus capturing a wide variety of shots across the movies in the same series.

Furthermore, a natural imbalance in the appearance of actors is observed since lead actors appear significantly more frequently than supporting actors, reflecting realistic movie screen-time distributions. The exceptions to this are multi-starrer comedy or action series where the screen-time of lead and supporting actors becomes comparable.

3.1.5. Challenges in Dataset Creation

Unlike Hollywood actors' datasets, which are readily available and have good quality images, datasets based on only Bollywood actors are very scarce and lack quality. Particularly, certain actors who have less representation in Bollywood do not have any curated datasets, giving rise to the need for manual collection of face images. Furthermore, many actors have maintained a long-standing presence in the film industry. This makes it imperative that facial changes due to aging are also captured. Another challenge is to ensure the dataset has diverse images, since it is prominently seen that Bollywood films use various costumes and dramatic makeup. To ensure that the images capture various lighting changes and face angles, they are sourced from multiple sources and over different time spans. Also, reliance on publicly available video content means that some frames are noisy or low resolution, increasing the manual effort needed to filter and select usable images.

3.2. Proposed Actor Recognition and Screen Presence Analysis (AR-SPA) Framework

This research introduces AR-SPA, an end-to-end framework for automated actor recognition and screen presence analysis in movies. This framework integrates a face detection model and an actor recognition model with a statistical engine that performs screen-time estimation and comparative analysis. The complete framework ensures robust performance under varied lighting conditions, expressions, poses, and cinematic transformations commonly seen in films.

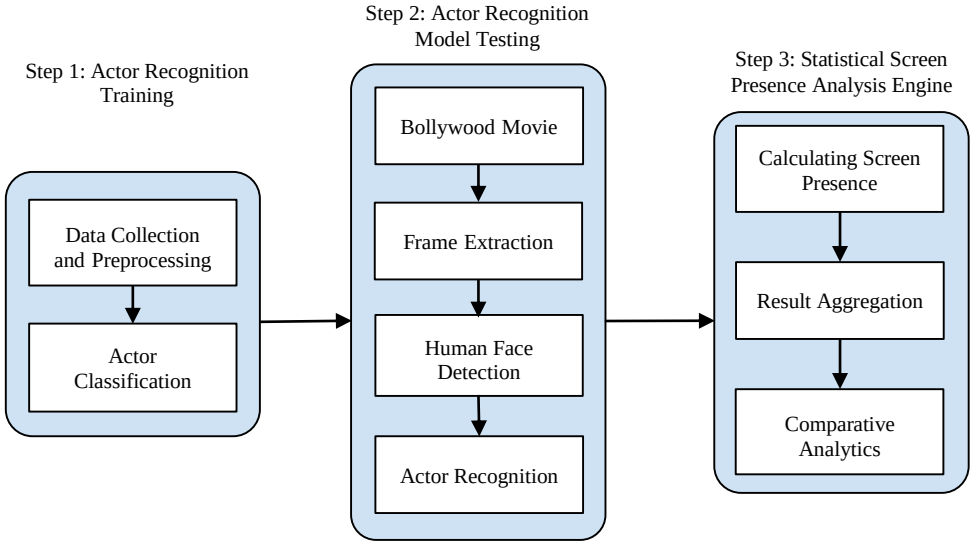


Fig. 2 Structural overview of AR-SPA framework

3.2.1. Overview of AR-SPA Framework

As observed in Figure 2, Step 1 involves data preprocessing and model training as discussed in Section 3.1. In Step 2, Bollywood movies undergo frame extraction. These frames are uniformly sampled so as to ensure consistency and resource management. From these sampled frames, each frame is preprocessed and passed to the YOLO face detection model [44]. This model accurately detects human faces in the received frames. These faces are cropped from the frame and fed into the face recognition model. The ResNet-CBAM recognition model has the spatial and channel attention via the convolution block attention model. This improves the model's robustness to detect faces even in cases of variations in lighting or setting. Once the cropped faces are received, the ResNet-CBAM model links each face with an actor. This face-actor mapping is done based on the confidence threshold of the model, which is constant throughout the movie. This threshold implies the confidence of the model in identifying the respective actor. In Step 3, all the valid face-actor associations are then considered further for calculating the screen time metrics of the target actors. These metrics include the total number of frames detected, the estimated temporal duration, and the proportion of screen presence. Each movie in the series is passed through these steps, and the results of

each movie are combined to give a comprehensive view of the screen time metrics. These statistics help in the comparative analysis of the target actors across the series and to identify their influence on the success of the movies. Overall, the AR-SPA framework integrates four major components, namely, frame extraction, face detection, actor recognition, and a statistical screen-presence analysis engine. These modules are arranged sequentially to ensure consistent and accurate processing across diverse video content.

Frame Extraction Module

This module carries out the time-based frame sampling and preprocessing. Initially, the input movie is broken down into an array of frames. This task is performed using Python libraries and inbuilt functions.

These frames are then sampled uniformly so as to reduce the computational load due to exhaustive frame processing. Each of these sampled frames undergoes preprocessing steps like resizing, normalization, and color space adjustments to standardize the input. This guarantees reduced variability and distortions in the frames that occur due to varied cinematic conditions and equips the frames for further stages.

Face Detection Module

With the standardized frames as input, this step is responsible for detecting the faces in each of these frames. All the frames are passed through a YOLO based face detection module. This model accurately identifies human faces and localizes them with bounding boxes. This is vital to ensure that only the relevant data in the frames is passed further to the recognition model. The detected, box-bound faces are cropped to maintain consistency.

Actor Recognition Module

The actor recognition module consists of both model training and testing as a part of the AR-SPA framework.

i. Model Training: Model training is a significant component of the initial stage of the framework. It involves feeding the pre-processed AR-SPA dataset into the actor classification model for training by adjusting various hyperparameters. The training happens under a controlled environment, and the checkpoints of the model are stored as part of the training and validation process. The final model is saved for testing and incorporation into the actor recognition pipeline.

ii. Model Testing: As part of model testing, the aligned face crops are fed into the actor recognition model. This model predicts the identity of the actor appearing in each frame. Only predictions that meet the minimum criteria for the confidence threshold are collected and sent to the final model. The confidence score is a probability value estimated by the recognition model. It reflects the certainty of the recognition output produced by the model. This selective filtration ensures that only high-certainty and relevant detections contribute to the final analysis.

Statistical Screen Presence Analysis Engine

All valid predictions from the recognition module are accumulated across the entire movie to derive screen presence statistics. For each actor, AR-SPA computes the total frames detected, estimated screen time (seconds and minutes), proportion of appearance relative to the movie, dominance ratio, and the comparative visibility across different movies in the series. These metrics are derived using frame counts and sampling rates, enabling accurate temporal estimates without processing every single video frame. Results from multiple movies are aggregated into structured tables to support longitudinal and comparative analysis.

3.3. Model used for Face Detection

Face detection is a crucial part of the proposed framework as it localizes the target region in the frames. It helps reduce the background noise and insignificant details in the frame, which negatively affect the recognition ability in further stages. The model primarily used for face detection in the AR-SPA framework is defined in the following subsection.

3.3.1. YOLO

YOLO is a face detection model highly known for its robustness to variations, precise bounding box localization, and high Recall. It has a convolutional architecture for feature extraction. This enables the division of the image into grid cells and direct prediction of the bounding boxes. For this study, the yolov8n model is employed as the first stage in the proposed framework due to its ease of integration across diverse systems and stability.

The nano variant was chosen to optimize computational resources as it yielded comparable performance with other larger variants. It localizes faces present in the captured video frames, enabling high-speed and accurate real-time video analysis. The detected face regions are extracted and processed before being fed to the classification network.

3.4. Models used for Actor Recognition

Accurate classification of the identified actor is crucial as it contributes to accurately detecting an actor's screen presence. To establish a reliable performance reference, one baseline and two hybrid architectures are implemented.

3.4.1. CNN

The architecture of the CNN model [45] contains sequential convolutional layers. The max-pooling and fully connected layers are present further. High-level spatial patterns are extracted from convolutional blocks, while pooling reduces dimensionality. The dense layer, which appears in the last layer of the network, allows nonlinear decision-making. The model provides a reference for performance without advanced feature gathering and residual

learning. It acts as a clean, interpretable benchmark against which the performance of the advanced architectures can be measured.

3.4.2. CNN-Transformer Hybrid

The model integrates a convolutional feature extraction mechanism with Transformer encoder blocks. CNNs excel at local spatial pattern extraction, while Transformer models capture long-range dependencies. The complementary features of both paradigms make it a perfect motivation for recognition purposes [10]. The hybrid architecture operates in two stages:

i. Convolutional Layer: This layer processes raw face images and extracts local visual descriptors such as edges, textures, and other facial patterns.

ii. Transformer Layer: The featured maps obtained from the previous layer are fed into encoder blocks. The self-attention mechanism with multiple heads helps the model to learn relevant global relationships between the facial regions. It captures the relational information missing in purely convolutional models.

The hybrid baseline is much more advanced than the traditional CNN and acts as a mid-tier reference point between simple CNNs and more advanced models such as ResNet-CBAM.

3.4.3. ResNet-CBAM Recognition Model

The Convolutional Block Attention Module (CBAM) [46] is integrated into the Residual Network (ResNet) [47] to create an enhanced recognition model.

Residual Network (ResNet)

The ResNet model is responsible for learning features from face images. These features can be used to discriminate between identities. It consists of convolutional layers, which are succeeded by batch normalization and Rectified Linear Unit Blocks. These layers are associated via residual connections that can bypass layers. The resulting stacked residual blocks are then grouped into stages, followed by a fully connected layer. The model can thus enable deep network training without the issue of vanishing gradients. It shows stable performance across large datasets. Hence, it serves as the backbone for the recognition model.

Convolutional Block Attention Module (CBAM)

CBAM is integrated after the ResNet block in the recognition model. It is used to further refine the learned features. It applies attention through two attention components - channel and spatial. The first component enables reweighting of the feature maps to prioritize information, while the spatial attention module highlights important facial regions. It brings focus to the appropriate facial features. Furthermore, it is highly robust and recognizes images accurately in case of expression changes and camera angles. The CBAM blocks are inserted after each layer of ResNet-50, a 50-layer variant of the ResNet model, enabling attention refinement at multiple depths. This architecture results in a compact, attention-weighted feature map that improves classification Accuracy while adding modest computational overhead.

4. Performance Evaluation

4.1. Experimental Setup

All the experiments conducted in this research are performed under fair and reproducible conditions. The experimental setup consists of configurations for dataset splitting, training configurations for the ResNet-CBAM model used in the actor recognition module, and the framework implementation details. These configurations remain consistent throughout the research to ensure fair evaluation. To facilitate model learning, train, validation, and test sets are formed by splitting the dataset. The split ratios vary across different models and datasets. As a part of this research, three different split configurations are evaluated. They are 75:15:10, 70:20:10, and 80:10:10. Experimental results demonstrate that the 75:15:10 dataset leads to the best performance and stable learning curves. This is attributed to the balanced ratio of training and validation datasets. The performance obtained through the 70:20:10 and 80:10:10 dataset splits is less accurate due to a lack of sufficient training data and validation data, respectively. These findings are complemented by the values of the metrics presented within Table 2.

Table 2. Comparison regarding ResNet-CBAM actor recognition model performance across various dataset split configurations

Train:Val:Test Split	75:15:10	70:20:10	80:10:10
Training Accuracy	0.9759	0.9681	0.9664
Validation Accuracy	0.9151	0.8871	0.8894
Testing Accuracy	0.9223	0.8968	0.8973
F1-score	0.9219	0.8965	0.8968
Selected	Yes	No	No

4.2. Evaluation Metrics

The performance of the AR-SPA framework is evaluated by using various evaluation metrics. A set of seven evaluation metrics, along with two novel metrics, is considered in this research. These metrics are classified into classification metrics and screen presence estimation metrics. The classification metrics, namely Accuracy, Precision, Recall, and F1-score, prove to be helpful in evaluating the ResNet-CBAM performance, which is used in the actor recognition module of the framework. The screen presence estimation metrics are used to evaluate the temporal estimation ability of the framework. They consist of Detected Frame Count, Screen-time Duration, Screen-time Percentage, and two novel Actor Dominance Metrics, namely Actor Dominance Ratio (ADR) and Dominance Shift.

Accuracy

Accuracy is used to estimate the overall reliability of actor recognition. The proportion of the number of actor face instances classified into their correct respective classes to the cumulative number of instances generates model accuracy. It is the measure of whether the model can rightly recognize the actor despite various changes in lighting, pose, expressions, and makeup. Higher accuracy of the model implies its correctness in predicting most appearances of the actor.

Precision

Precision refers to the reliability of positive actor predictions. It is estimated by calculating the ratio of rightly identified instances of an actor to the total count of the instances classified as that particular actor. Precision is critical while performing screen presence analysis since it indicates the amount of incorrect classifications. This is especially significant in frames where multiple actors are visible together, since it prevents misclassification due to visual similarity and occlusions.

Recall

Recall is used to estimate whether the recognition model correctly recognizes all instances of a given actor, given a video frame or a set of test images. It reflects how many actor face instances are correctly identified out of all the true actor face instances present. It is an indication of how many frames containing the actor will be truly detected in the movie with respect to all the frames in which the actor has been present during the movie's duration. A high Recall signifies that the framework is accurately detecting most actor appearances.

F1-score

F1-score is the key metric balancing the Precision and Recall of the model. It is the harmonic mean of the two, providing an overall score of the recognition performance. It is a significant aspect of the AR-SPA framework, as both incorrect actor detections and missed actor appearances can negatively affect the screen estimation metrics' accuracy by distorting them. This proves that the F1-score is a good estimate regarding the reliability of the model for comprehensive screen-time analysis.

Detected Frame Count

The Detected Frame Count is measured by estimating the cumulative sum of the number of all the frames in which the actor is identified. The video frame in which the confidence score of the identified actor crosses the target threshold is considered a detected frame. The summation of the number of all such frames across the movie duration gives the Detected Frame Count.

$$\mathrm{F(a_i)} = \sum_{\mathrm{k=1}}^{\mathrm{F_M}} \mathrm{D(a_i \in Frame_k)} \quad (1)$$

Here, $F(a_i)$ denotes the detected frame count for an actor a_i, F_M denotes the cumulative sum of the number of frames in a movie M, and D takes the value 1 if the actor a_i is detected in frame k, else 0.

Screen-time Duration

The Screen-time Duration estimated by the framework is the detected frame count converted to a time-based metric, such as seconds or minutes. It is the ratio of the Detected Frame Count to the frame sampling rate used in the framework. The frame sampling rate is the number of frames extracted per second from the movie. Screen-time Duration is used to quantify the screen presence of an actor in the movie.

$$\square(\square_{\square}) = \frac{\square(\square_{\square})}{\square_{\square}} \quad (2)$$

Here, $T(a_i)$ represents the screen-time duration (in seconds) for the actor a_i, $F(a_i)$ is the detected frame count as estimated from Equation (1), and f_s is the frame sampling rate.

Screen-time Percentage

The screen-time percentage refers to the percentage contribution of the actor's screen presence. It is computed relative to the total movie duration. It is mainly used to normalize screen-time estimates across different movies. It reflects the relative narrative presence of each character in the movie.

$$P(a_i) = \frac{T(a_i)}{F_M} \times 1 \quad (3)$$

Here, $P(a_i)$ is the screen-time percentage of an actor a_i, $T(a_i)$ is the screen-time duration as estimated in Equation (2), and F_M is the cumulative sum of frames in movie M.

4.2.1. Novel Actor Dominance Metrics

As a part of this research, novel Actor Dominance Metrics are defined. These metrics help in categorizing the relative narrative dominance and significance of the targeted actors across the movies.

Actor Dominance Ratio (ADR)

ADR is defined as the proportion of an actor's screen time relative to the cumulative screen time recorded of all target actors in a movie. It quantifies the dominance of one actor over another, as compared to the screen presence percentage, which refers to the absolute presence of an actor in the movie. At a series level, the Average ADR summarizes the persistent dominance across all installments.

$$ADR(a_i) = \frac{T(a_i)}{\sum_{j=1}^{N} T(a_j)} \quad (4)$$

Here, $ADR(a_i)$ is the Actor Dominance Ratio and $T(a_i)$ denotes the screen-time duration of the actor a_i, whereas the cumulative count of target actors in the movie is shown by the variable N.

Dominance Shift

The Dominance Shift quantifies the variation in actor prominence across the franchise. It is estimated as the variation between the maximum and minimum ADR of an actor across the series.

A higher dominance shift indicates that the actor's relevance has significantly changed throughout the series, while a lower shift indicates stability in actor visibility.

$$\Delta ADR(a_i) = ADR_{max}(a_i) - ADR_{min}(a_i) \qquad (5)$$

Here, $\Delta ADR(a_i)$ is the Dominance Shift, $ADR_{max}(a_i)$ is the maximum ADR and $ADR_{min}(a_i)$ in the minimum ADR for the actor a_i.

The screen-time estimation metrics are not compared against ground-truth values due to the unavailability of reference datasets and statistical analysis on Bollywood movies. Their consistency across the various experimental models and alignment with narrative roles provide a qualitative validation of the framework's reliability. The recognition model's performance metrics and screen-time estimation metrics together assist in the overall evaluation of the AR-SPA framework.

4.3. Results and Analysis

The results and analysis related to the comparison of the actor recognition models and the screen-time statistics of the actors are highlighted in this section. These statistics are obtained by applying the proposed AR-SPA framework to the entire movie dataset.

4.3.1. Results of Actor Recognition Model Comparison

This section presents a comparative evaluation of the actor recognition model ResNet-CBAM used in the proposed framework against baseline CNN and CNN-Transformer models. The comparison is performed across two stages, as shown in Figure 3.

Performance-based Model Comparison

This subsection presents a quantitative comparison of the actor recognition models using Accuracy, Precision, Recall, and F1-score, visualized using a metric-wise heatmap in Figure 4. Each row of the heatmap represents a recognition model, while the columns correspond to the performance metrics. They are obtained by experimentation performed on a test set held out during the training. The recognition capability of each model is analyzed under identical experimental conditions.

It can be observed from Table 3 and Figure 4 that the ResNet-CBAM model used in the proposed framework consistently outperforms the other baseline models. It displays the highest scores across all evaluation metrics. The attention mechanisms incorporated into the residual network make the model capable of capturing miniscule facial features and generalizing them to provide a robust recognition mechanism, which captures variations in pose, illumination and expression. The baseline models in comparison show significantly lower accuracy scores than the ResNet-CBAM model. With the lowest performance scores, the CNN model displays the least performance of all models. The findings indicate that the model is effective for basic image recognition, but it cannot handle complexities and variations that occur in real-world movie scenes. The CNN-Transformer hybrid model performs comparatively better. The self-attention modules integrated into the hybrid model allow it to capture global relationships. The ResNet-CBAM model is integrated into the proposed AR-SPA framework, serving as the foundation for actor face recognition and screen-time analysis on real movie content.

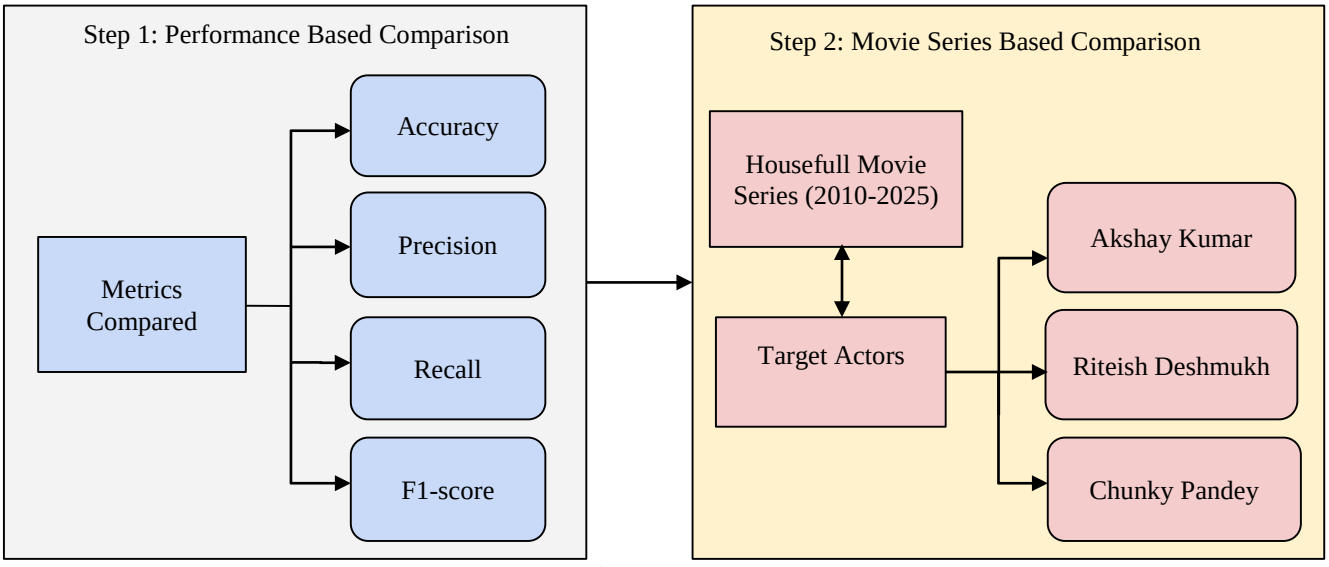


Fig. 3 Flow diagram of actor recognition model comparison

Table 3. Performance comparison between face recognition models

Model	CNN	CNN-Transformer	ResNet-CBAM
Validation Accuracy	0.6315	0.7108	0.9151
Testing Accuracy	0.5828	0.7035	0.9223
Generalization	Limited	Good	Excellent
Overfitting	Moderately Overfit	Slightly Overfit	No Overfit
Misclassified	41.72%	29.65%	7.77%

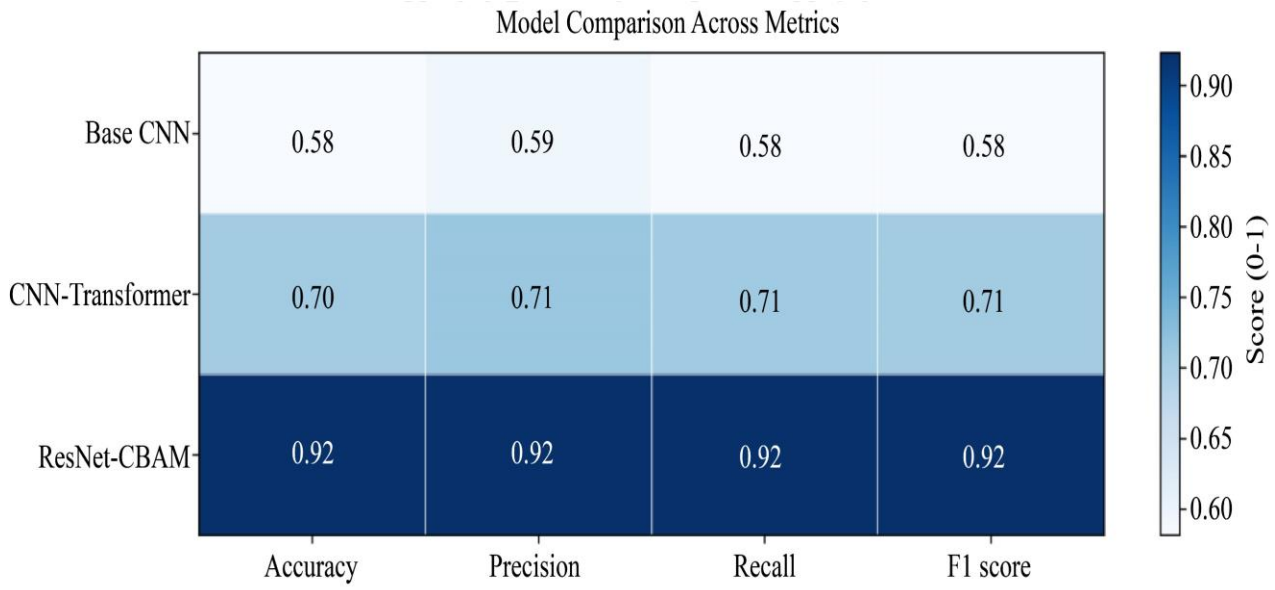


Fig. 4 Heatmap illustrating the performance comparison of CNN, CNN-Transformer, and ResNet-CBAM models

Model Comparison Based on Housefull Movie-Series Evaluation

The performance metrics discussed in the previous section highlight the capability of the recognition model. However, their behavior in real-world cinematic videos is influenced by various factors such as differences in shot angles, actors blended in crowd scenes, and camouflaging costumes. It is essential to assess their ability to provide accurate actor detections in such scenarios. For this purpose, the recognition models used in this research are tested on the Housefull movie series (2010-2025). This provides a comparative analysis of their performance in real-world scenarios.

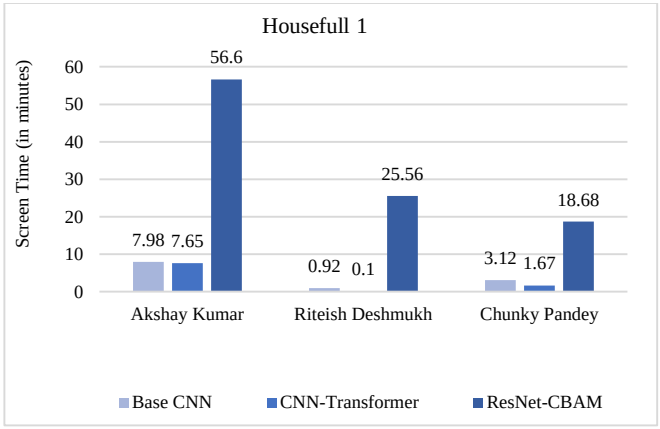
The Housefull series comprises five multi-starrer movies with multiple recurring actors. The three target actors for the screen-presence analysis are Akshay Kumar, Riteish Deshmukh, and Chunky Pandey. These actors appear consistently throughout the series. They serve as ideal actors for screen presence analysis due to their appearances as lead, co-lead, and supporting actors, respectively. This setup introduces real-world challenges as opposed to the standard test dataset used in performance metrics estimation. The test frames are affected by frequent scene transitions, wide-range shots, occlusions, changing lighting conditions, and changes in appearances throughout the movies. These changes also vary across all the movies in the series. Such conditions make the movies an ideal testing setup for the model performance evaluation. The comparative results obtained are illustrated in Figures 5 to 8. Figure 5 presents the movie-wise screen-time comparison for each actor.

The individual subplots represent the movies in the series. The joint bars illustrate the screen-time of all three target actors estimated using the three models. Figure 6 illustrates the actor-wise screen presence variation across the series. Each subplot denotes an actor and shows how their screen presence varies across the movies.

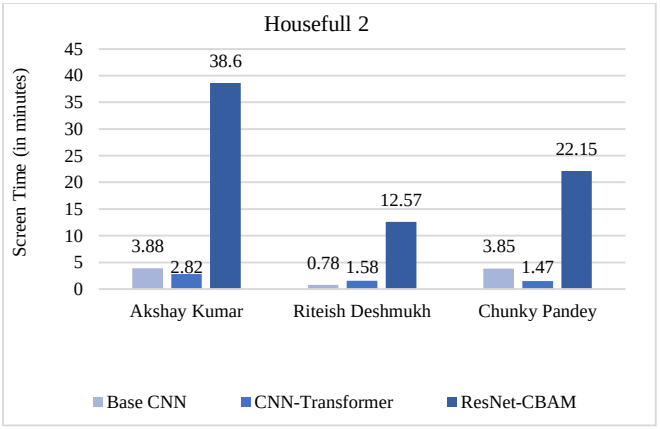
To further observe and analyze cross-model behavior, a global screen-time comparison heatmap (Figure 7) is used to visualize the comprehensive actor-wise and movie-wise screen presence across all models.

Figure 8 demonstrates the average screen time variation across all the movies as estimated by each model. From the illustrations in Figures 5 to 8, it is clear that the ResNet-CBAM model produces more stable and proportionally accurate screen-time estimates. These values closely align with the narrative presence of the actors in the respective films and are verified manually by evaluating the movie frames. In comparison, the conventional CNN model clearly under-

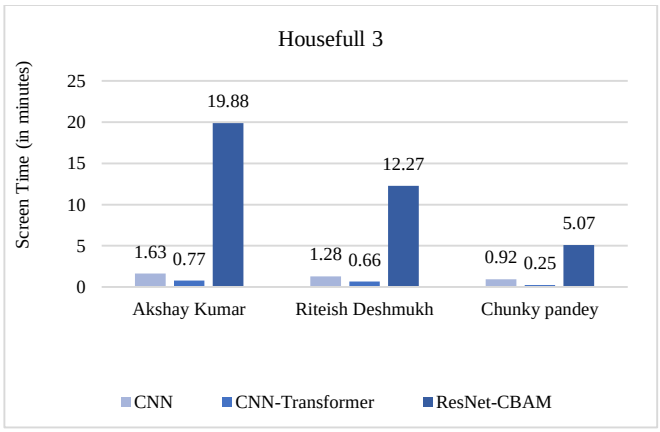
estimates the screen presence. Furthermore, the CNN-transformer hybrid model shows improvement, but allows misclassification and under-classification to a large extent. These illustrations reinforce the ability of the ResNet-CBAM model used in the proposed AR-SPA framework to produce balanced, consistent, and reliable screen-time estimates. This is due to the incorporation of attention-driven feature refinement, which enables better handling of pose variations, lighting changes, and brief actor appearances. Conversely, the baseline models exhibit irregular fluctuations and inconsistencies. These irregularities accumulate over long-duration videos and impact final screen-time statistics.



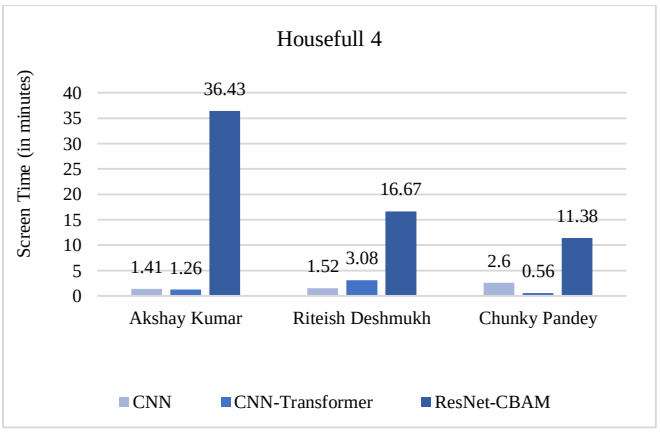
(a)



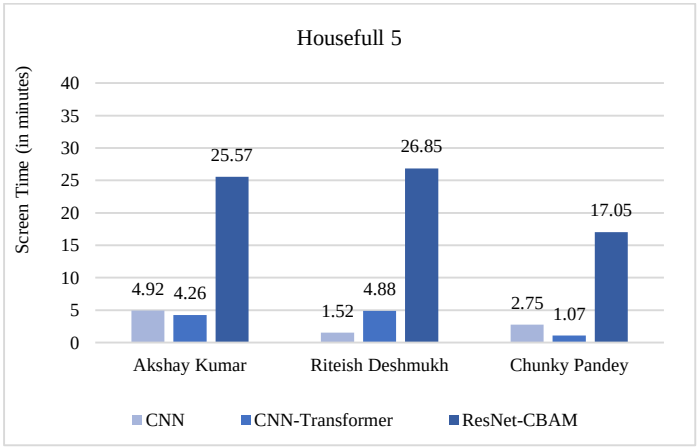
(b)



(c)



(d)



(e)

Fig. 5 Movie-wise actor screen-time comparison in the housefull series (2010-2025) using Proposed AR-SPA and Baseline Models: (a) Housefull 1(2010), (b) Housefull 2(2012), (c) Housefull 3(2016), (d) Housefull 4(2019), and (e) Housefull 5(2025).

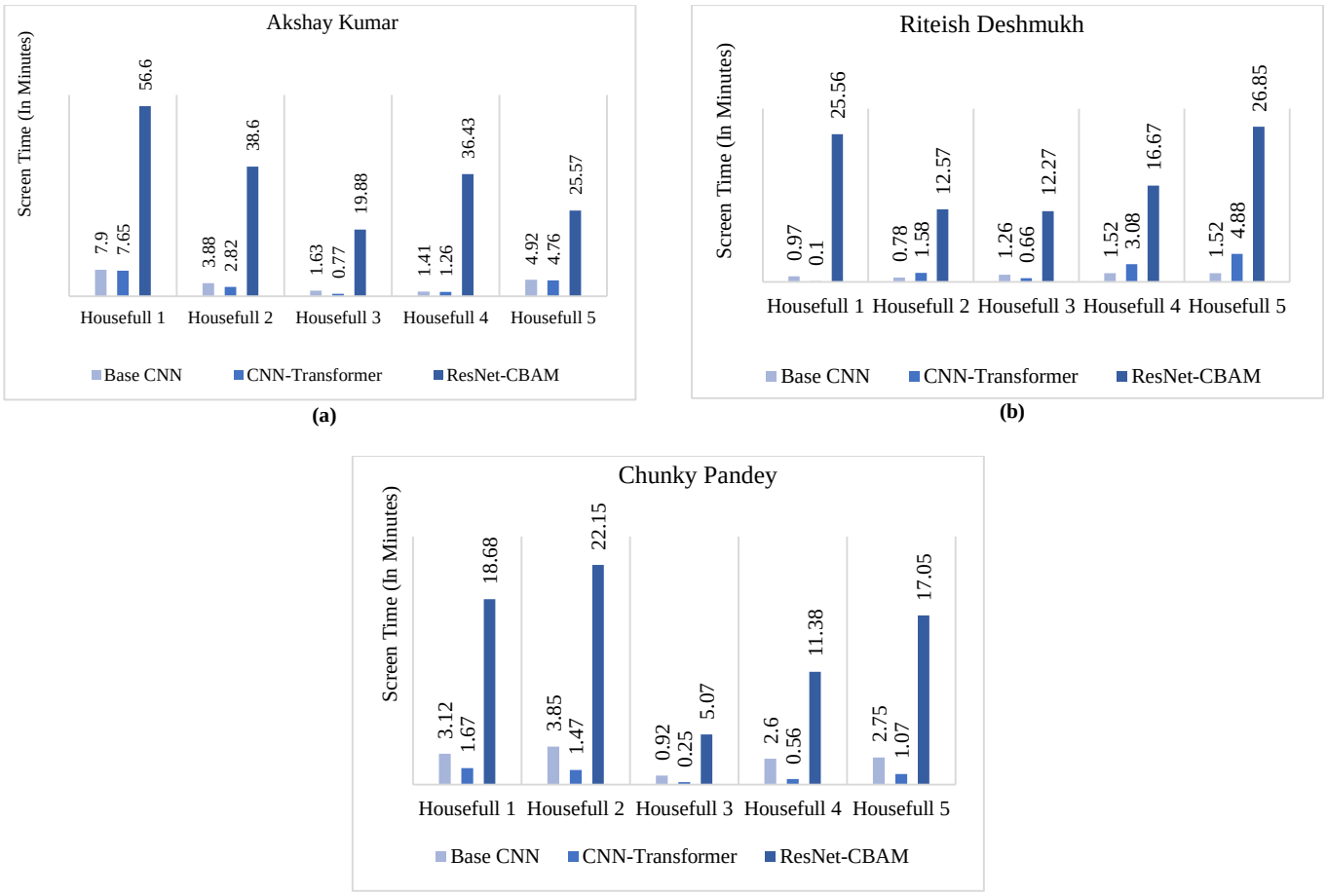


Fig. 6 Actor-wise screen presence comparison across the housefull series (2010-2025) using Proposed AR-SPA and Baseline Models: (a) Akshay Kumar, (b) Riteish Deshmukh, and (c) Chunky Pandey.

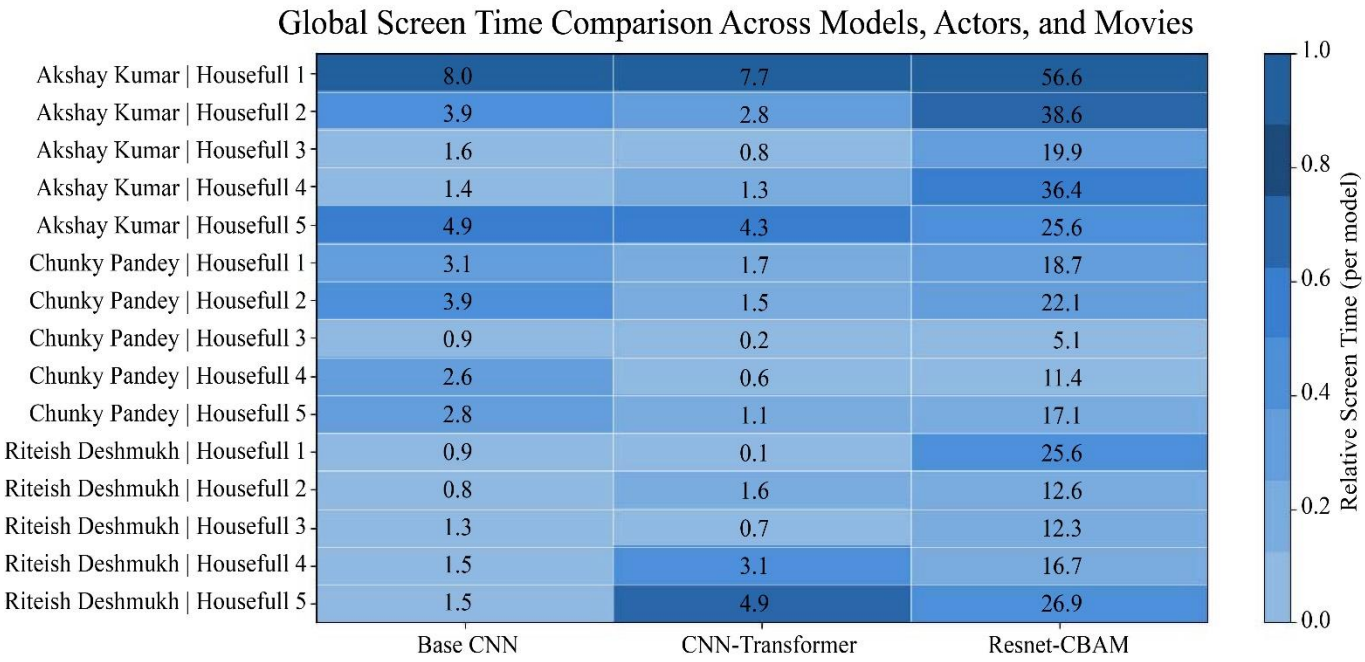


Fig. 7 Global heatmap illustrating actor screen presence across movies and recognition models

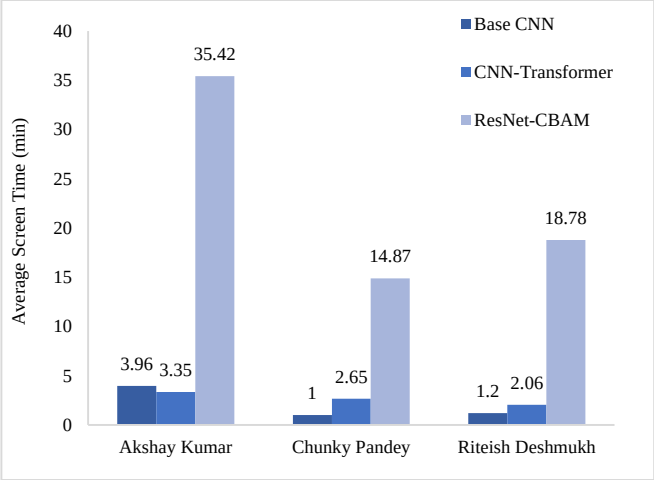


Fig. 8 Average screen presence of target actors across the housefull series (2010-2025) using all evaluated models

4.3.2. Results of Movie Screen Presence Analysis: Lead vs Supporting Actors

The results obtained from the AR-SPA framework with respect to individual movie analysis are discussed in this section. This includes the screen presence distribution among the lead and supporting actors in a movie. These results help in estimating the alignment of the calculated screen presence metrics with the narrative presence of the actors in the movies.

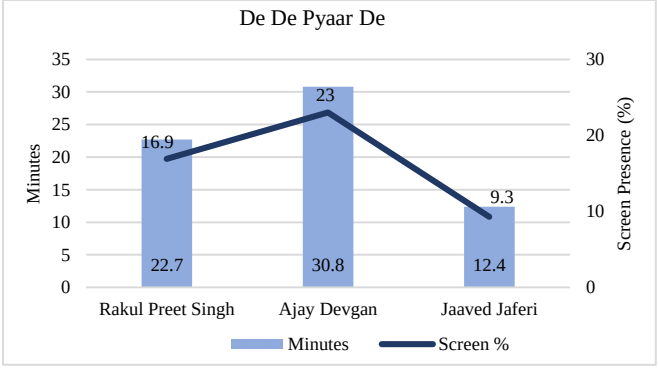
The actor's screen presence in each movie is quantified with the help of various metrics as discussed in the previous section 4.2. The findings are illustrated in Figure 9.

The figure consists of four sub-plots, each depicting the estimated screen-time statistics for an individual movie. The double axis in the graphs indicates the screen presence estimation with respect to the detected minutes that appeared and the percentage for three target actors in the movie.

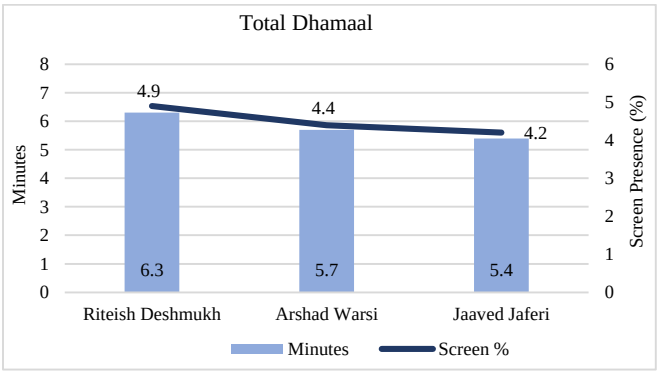
Figure 9 demonstrates a consistent trend across all movies, where the lead actor boasts significantly higher screen presence as compared to supporting actors. This aligns with the narration of the story and the importance of characters in the movie.

Certain deviations from this trend are seen in Figure 9(b) and Figure 9(c), where the difference between the screen time of the lead and supporting cast is very small. This is attributed to the multi-star nature of the movie.

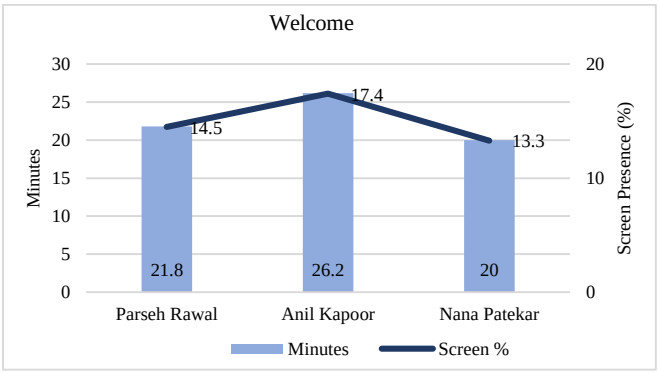
These findings establish that actor screen presence varies across movies based on various factors. These factors include the narrative importance of the character played by the target actor, the character's journey, and the changes in character arcs over time. AR-SPA effectively captures these findings and provides a solid ground for emphasizing actor prominence based on appearance.



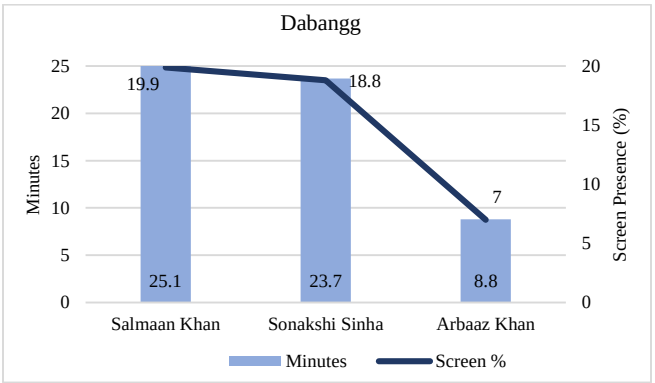
(a)



(b)



(c)



9 (d)

Fig. 9 Actor-wise screen presence analysis within individual movies: (a)De De Pyaar De (2019), (b) Total dhamaal (2019), (c) Welcome (2007), and (d) Dabangg (2010).

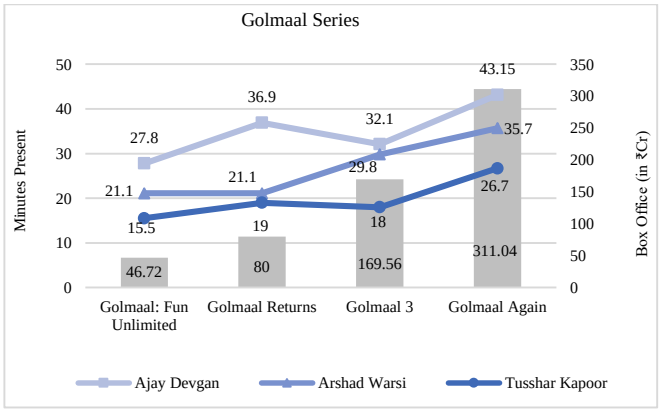
4.3.3. Results of Series-Level Screen Presence Analysis

While the previous subsection focuses on individual movies, this analysis extends to a series-level perspective. In this section, the evolution of the screen presence of lead and supporting actors across all installments within a franchise is examined. The objective is to capture longitudinal trends in actor prominence and understand their significance with regard to the series and its narration. Their relation to the commercial performance of the films is also analyzed wherever applicable. The movie series analyzed and tested as a part of this research comprises both movies with an ensemble cast as well as movies with a starring role.

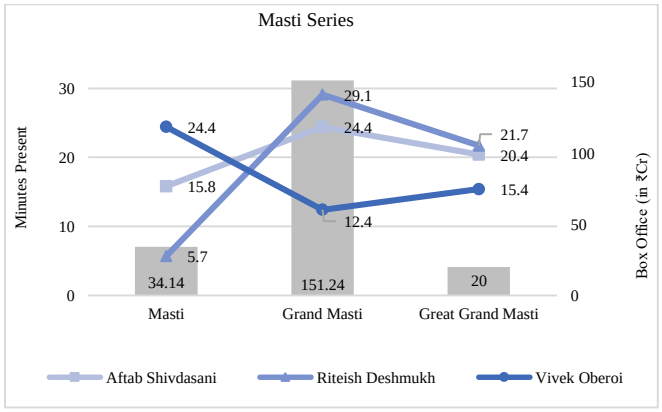
Ensemble-driven Movie Series

The narration of an ensemble-driven movie tends to revolve around multiple characters and their independent lives are inter-woven through a genre-driven, intricate story-line. This leads to almost equivalent focus on the entire ensemble cast, and the screen-time captured by the lead and supporting actors remains nearly constant. This pattern is observed throughout all the movies in the franchise, with a rise and fall of the individual screen-times due to plot, narration, and relative importance of the character in individual movies.

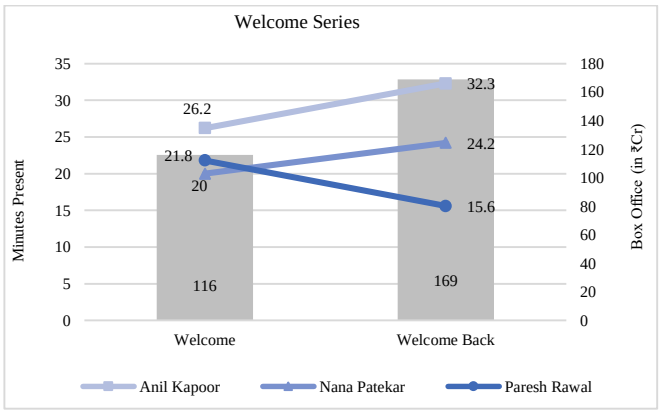
Figure 10 presents series-wise screen presence trends for four ensemble-driven movie series, where each subgraph corresponds to a distinct series. Line plots depict the screen presence of selected lead and supporting actors across successive movies, while background bar plots represent the corresponding box office collections. This combined visualization enables simultaneous assessment of narrative prominence and commercial outcomes.



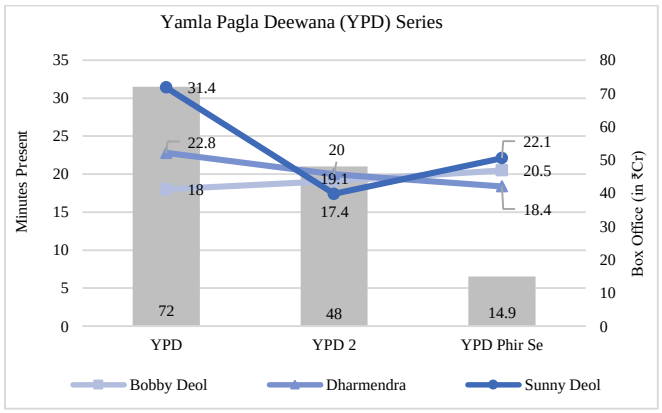
(a)



(b)



(c)



(d)

Fig. 10 Actor-wise screen presence trends across ensemble-driven movie series with box office: (a) Golmaal Series (2006-2017), (b) Masti Series (2004-2016), (c) Welcome Series (2007-2015), and (d) Yamla Pagla Deewana Series (2011-2018)

As inferred from Figure 10 (a), the screen presence of the three target actors in the Golmaal franchise remains nearly equivalent throughout the series. As the series progresses, different characters achieve prominence, leading to an increase in their calculated screen-times, while the overall trend remains stable. Figures 10 (b) and 10 (d) demonstrate how the focus of the story shifts across characters throughout the series. However, the screen presence distribution remains almost equal despite the occasional variations. Figure 10 (c) depicts the results of an ensemble movie, where the lead actor gains slightly more focus than the co-actors. The screen presence of some supporting actors decreases across the series due to changes in narration and flow of the story.

The inclusion of box office data provides insight into the relationship between actor presence and commercial success. While certain series show moderate correlation between increased actor visibility on the screen and higher box office returns, the pattern is not consistent across all franchises. These findings suggest that the screen presence of actors is not the sole factor of the commercial success of the series, but is also influenced by factors such as storytelling, direction, marketing, and audience reception.

Movies with Starring Roles

The movies with starring roles are often narrative-centric, with screen presence concentrated around a lead actor and, in some cases, a co-lead actor. Their presence typically remains stable or shows slight fluctuations based on the narrative focus and character growth.

Any roles other than these are supporting and occupy comparatively less screen time. The dominance of lead actors as compared to supporting actors throughout the franchise is illustrated in Table 4.

The screen presence statistics demonstrated in Table 4 are centered around four popular movie series starring a lead role. It can be inferred from the table that the lead and co-lead actors take the center stage, and the supporting actors contribute to the progression of the lead actor's narrative arc. The average dominance ratio of the lead actor remains the highest across all series, with the co-protagonist following with an almost equal prominence, if present. The supporting cast exhibits an estimated low dominance ratio, with certain exceptions like the Stree movie series. These exceptions occur due to the shift in narrative focus of the films.

The supporting actors often have comedic highlights or serve as key contributors to the actions of the lead characters. Conversely, the co-protagonist is seen as the highlight of the movie, achieving significance with lesser screen time through the individual presence and plot.

The dominance shift in the screen presence of the actors is analyzed effectively through the results. As observed in Table 4, the dominance shift of the lead actors is negligible. This implies the significance of these actors and their constant appearance throughout the entire series. Similarly, the dominance shift observed in the presence of the co-actors is very small.

This is because of their roles alongside the lead actors. On the other hand, due to the varying significance of supporting actors, the corresponding dominance shift is observed to be comparatively higher. This change is caused by the difference in contributions of supporting actors and the lead actors to the story of the film. The quantitative results and the trends highlighted from the above insights justify the importance of lead actors.

Table 4. Actor Screen-presence dominance statistics across movie series

Series	Actor	Total minutes	Avg ADR	Max ADR	Movie with the most dominant appearance	Dominance shift
Stree (2018-2024)	Rajkumar Rao	92.18	0.443	0.467	Stree 2	0.049
	Pankaj Tripathi	70.86	0.340	0.378	Stree 2	0.075
	Shraddha Kapoor	45.33	0.217	0.279	Stree	0.123
Dulhania (2014-2017)	Varun Dhawan	38.12	0.495	0.501	Badri Ki Dulhania	0.014
	Alia Bhatt	25.44	0.333	0.405	Humpty Sharma Ki Dulhania	0.144
	Sahil Vaid	13.48	0.173	0.238	Badri Ki Dulhania	0.131
Dabangg (2010-2019)	Salman Khan	74.17	0.473	0.509	Dabangg 3	0.073
	Sonakshi Sinha	57.97	0.363	0.411	Dabangg	0.113
	Arbaaz Khan	24.84	0.164	0.228	Dabangg 2	0.117
Son of Sardaar (2012-2025)	Ajay Devgan	46.53	0.747	0.799	Son of Sardaar 2	0.105
	Mukul Dev	10.88	0.169	0.253	Son of Sardaar	0.167
	Sanjay Mishra	5.15	0.084	0.115	Son of Sardaar 2	0.062

4.3.4. Qualitative Results: Actor Detection and Recognition

To complement the quantitative screen presence analysis, this section presents the qualitative frame-level samples, which highlight the effectiveness of the proposed framework. The bounding boxes as seen in Figure 11 are the actor faces detected by the face detection module of the AR-SPA framework. Furthermore, the actor recognition module identifies the actor and outputs a confidence score for each recognition, as observed in the top-right corner of each detected bounding box. The confidence score represents the recognition model's probability of the actor identity being accurate. For the AR-SPA framework, only classifications with a confidence score greater than 0.80 are included in the screen presence metrics estimation process.

(a) (b)

Fig. 11 Illustrative samples of actor detection and recognition in movie frames: (a) Grand masti (2013), and (b) Tiger zinda hain (2017) [Source: All frames are sourced from publicly available movie copies hosted on the internet archive: https://archive.org/]

5. Discussions

The proposed AR-SPA framework achieves better performance and results compared to the state-of-the-art techniques. AR-SPA combines actor recognition and screen presence quantification into an end-to-end pipeline. In contrast with existing techniques, which do not focus on screen presence quantification, the AR-SPA pipeline results in a comprehensive analysis of the actor's screen presence in Bollywood movies. The earlier works achieve good accuracy, but are prone to heavy computational load and time complexity. A large section of the existing frameworks is limited to short-video analysis, since the long-length movies are not compatible with the heavy framework. AR-SPA mitigates this by balancing the trade-off between accuracy and performance smoothly. The computational requirements of the proposed framework are very low. It does not require large storage space for shots and continuous scenes in the movies, as opposed to the state-of-the-art frameworks. The framework performs detection of scene continuity in the movies by keeping a simple check of the actor detection in subsequent frames. This check is used to identify the continuous shots and get accurate statistics for actor presence. The novel screen presence analysis metrics proposed in the work enhance the performance and evaluation of the AR-SPA framework. These metrics assist in gaining a deeper understanding of the actor's screen presence and their relative importance compared to other actors in the movies. The narration of the story becomes clear and easy to understand due to the detailed analysis of actor dominance and influence. The existing techniques do not focus largely on the aforementioned parameters. This emphasizes the higher performance of the AR-SPA framework in comparison with the state-of-the-art frameworks.

6. Conclusion

This research work successfully delivers AR-SPA - a complete framework for actor detection and recognition in Bollywood movies. This framework gives statistical insights into the presence of an actor in a movie and the variation in his presence across the movie series. The experimental results obtained from this framework are accurate and aligned with the real-world data. Unlike the state-of-the-art approaches, the AR-SPA framework takes into consideration the cinematic diversity and efficiently handles various conditions like changes in appearance and illumination, and occlusion, which are prominent across Bollywood movies. One of the major contributions of this research is the Bollywood actors dataset containing images of 48 actors. The comparative evaluation of three actor recognition models demonstrates the higher performance of the ResNet-CBAM model. This model is tested on a total of 50 Bollywood movies and achieves an Accuracy of 92%, highlighting its efficiency in detecting Bollywood actors. Furthermore, two novel screen presence analysis metrics, namely Actor Dominance Ratio (ADR) and Dominance Shift, are proposed to enable assessment of the narrative presence of the actors. AR-SPA helps dive deep into the aspect of variations in audience engagement and box office collections with changes in actor screen presence, supported with statistics, which would potentially result in tangible benefits through informed decisions. For future research, enhancement, and experimentation, the proposed AR-SPA dataset is scalable and can be expanded further to include additional actors. This research can also be extended to include the aspects of influence of social media, marketing, publicity stunts, sentiment analysis, and other factors that contribute to the box office collection of movies.

References

[1] Chunfang Li et al., “PyCinemetrics: Computational Film Studies Tool based on Deep Learning and PySide2,” *SoftwareX*, vol. 26, pp. 1-8, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[2] James E. Cutting, Jordan E. DeLong, and Christine E. Nothelfer, “Attention and the Evolution of Hollywood Film,” *Psychological Science*, vol. 21, no. 3, pp. 432-439, 2010. [CrossRef] [Google Scholar] [Publisher Link]

[3] Florian Schroff, Dmitry Kalenichenko, and James Philbin, “FaceNet: A Unified Embedding for Face Recognition and Clustering,” *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, pp. 815-823, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[4] Yaniv Taigman et al., "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, pp. 1701-1708, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[5] SouYoung Jin et al., "End-to-End Face Detection and Cast Grouping in Movies using Erdös-Rényi Clustering," *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, pp. 5286-5295, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[6] Arsha Nagrani, and Andrew Zisserman, "From Benedict Cumberbatch to Sherlock Holmes: Character Identification in TV Series without a Script," *arXiv preprint*, pp. 1-13, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[7] Ijaz Ul Haq et al., "DeepStar: Detecting Starring Characters in Movies," *IEEE Access*, vol. 7, pp. 9265-9272, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[8] Abhinav Aggarwal et al., "Robust Actor Recognition in Entertainment Multimedia at Scale," *Proceedings of the 30th ACM International Conference on Multimedia*, New York, NY, USA, 2079-2087, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[9] Tomas Kerepecky et al., *Automated Actor Recognition in Video Content*, Data Science in Applications, Springer, Cham, pp. 3-22, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[10] Lin He, Lile He, and Lijun Peng, "CFormerFaceNet: Efficient Lightweight Network Merging a CNN and Transformer for Face Recognition," *Applied Sciences*, vol. 13, no. 11, pp. 1-14, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[11] Huafeng Wang, Yunhong Wang, and Yuan Cao, "Video-based Face Recognition: A Survey," *World Academy of Science, Engineering and Technology*, pp. 293-302, 2009. [Google Scholar]

[12] Stefanos Zafeiriou, Cha Zhang, and Zhengyou Zhang, "A Survey on Face Detection in the Wild: Past, Present and Future," *Computer Vision and Image Understanding*, vol. 138, pp. 1-24, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[13] Prakhar Kulshreshtha, and Tanaya Guha, "Dynamic Character Graph Via Online Face Clustering for Movie Analysis," *Multimedia Tools and Applications*, vol. 79, no. 43-44, pp. 33103-33118, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[14] Jitao Sang, and Changsheng Xu, "Robust Face-Name Graph Matching for Movie Character Identification," *IEEE Transactions on Multimedia*, vol. 14, no. 3, pp. 586-596, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[15] Hayeon Kim et al., "Character Detection in Animated Movies using Multi-Style Adaptation and Visual Attention," *IEEE Transactions on Multimedia*, vol. 23, pp. 1990-2004, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[16] Salwa Shakir Baawi, Farah Jawad Al-Ghanim, and Nisreen Ryadh Hamza, "Categorization of Celebrity Photos based on Deep Machine Learning for Feature Extraction and Classification," *Wasit Journal of Computer and Mathematics Science*, vol. 4, no. 1, pp. 1-16, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[17] Alexander Gabourie, and Connor McClellan, "Actor Identification with Deep Learning," *Stanford University, CS230 Deep Learning*, 2018. [Google Scholar]

[18] Xiawei Lu, "Digital Media Video Image Data Processing based on Computer Vision," *International Journal of Computer Applications in Technology*, vol. 78, no. 2, pp. 101-111, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[19] Zilinghan Li et al., "ViCTer: A Semi-Supervised Video Character Tracker," *Machine Learning with Applications*, vol. 12, pp. 1-14, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[20] Joshua Baldwin, and Ralf Schmälzle, "A Character Recognition Tool for Automatic Detection of Social Characters in Visual Media Content," *Computational Communication Research*, vol. 4, no. 1, pp. 351-371, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[21] Vignesh Ramanathan et al., "Detecting Events and Key Actors in Multi-Person Videos," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3043-3053, 2015. [Google Scholar] [Publisher Link]

[22] Minchul Kim, Anil Jain, and Xiaoming Liu, "50 Years of Automated Face Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1-20, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[23] S. Foucher, and L. Gagnon, "Automatic Detection and Clustering of Actor Faces based on Spectral Clustering Techniques," *Fourth Canadian Conference on Computer and Robot Vision (CRV '07)*, Montreal, QC, Canada, pp. 113-122, 2007. [CrossRef] [Google Scholar] [Publisher Link]

[24] Christina Chrysouli, Nicholas Vretos, and Ioannis Pitas, "Face Clustering in Videos based on Spectral Clustering Techniques," *The First Asian Conference on Pattern Recognition*, Beijing, China, pp. 130-134, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[25] Remigiusz Baran, Filip Rudzinski, and Andrzej Zeja, "Face Recognition for Movie Character and Actor Discrimination based on Similarity Scores," *2016 International Conference on Computational Science and Computational Intelligence (CSCI)*, Las Vegas, NV, USA, pp. 1333-1338, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[26] Vineet Gandhi, and Remi Ronfard, "Detecting and Naming Actors in Movies using Generative Appearance Models," *2013 IEEE Conference on Computer Vision and Pattern Recognition*, Portland, OR, USA, pp. 3706-3713, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[27] Guangyu Gao et al., "Cast2Face: Assigning Character Names onto Faces in Movie with Actor-Character Correspondence," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 12, pp. 2299-2312, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[28] Krishna Somandepalli, Rajat Hebbar, and Shrikanth Narayanan, "Robust Character Labeling in Movie Videos: Data Resources and Self-Supervised Feature Adaptation," *IEEE Transactions on Multimedia*, vol. 24, pp. 3355-3368, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[29] Qing Wang et al., "Object Tracking with Joint Optimization of Representation and Classification," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 4, pp. 638-650, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[30] Vijay Kumar, Anoop M. Namboodiri, and C.V. Jawahar, "Face Recognition in Videos by Label Propagation," *2014 22nd International Conference on Pattern Recognition*, Stockholm, Sweden, pp. 303-308, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[31] P. Bojanowski et al., "Finding Actors and Actions in Movies," *2013 IEEE International Conference on Computer Vision*, Sydney, NSW, Australia, pp. 2280-2287, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[32] Shoufa Chen et al., "Watch Only Once: An End-to-End Video Action Detection Framework," *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, QC, Canada, pp. 8158-8167, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[33] Olivier Duchenne et al., "Automatic Annotation of Human Actions in Video," *2009 IEEE 12th International Conference on Computer Vision*, Kyoto, Japan, pp. 1491-1498, 2009. [CrossRef] [Google Scholar] [Publisher Link]

[34] Jae Young Choi et al., "An Automatic Face Indexing Framework for Actor-based Video Services in an IPTV Environment," *2010 Digest of Technical Papers International Conference on Consumer Electronics (ICCE)*, Las Vegas, NV, USA, pp. 81-82, 2010. [CrossRef] [Google Scholar] [Publisher Link]

[35] Lai-Tee Cheok et al., "Automatic Actor Recognition for Video Services on Mobile Devices," *2012 IEEE International Symposium on Multimedia*, Irvine, CA, USA, pp. 384-385, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[36] Himanshu Nainwal, and Koshti Vanshika Shaileshbhai, "An Approach for Face Detection and Face Recognition using OpenCV and Face Recognition Libraries in Python using GPU," *2024 8th International Conference on Computational System and Information Technology for Sustainable Solutions (CSITSS)*, Bengaluru, India, pp. 1-6, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[37] Nithya Manickam, and Sharat Chandran, "Fast Lead Star Detection in Entertainment Videos," *2009 Workshop on Applications of Computer Vision (WACV)*, Snowbird, UT, USA, pp. 1-6, 2009. [CrossRef] [Google Scholar] [Publisher Link]

[38] Sher Muhammad Doudpota, and Sumanta Guha, "Automatic Actors Detection in Musicals," *Seventh International Conference on Digital Information Management (ICDIM 2012)*, Macau, Macao, pp. 226-231, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[39] Tomas Kerepecky et al., "STAR: Screen Time and Actor Recognition in Video Content," *Pattern Recognition: 46th DAGM German Conference, DAGM GCPR 2024*, Munich, Germany, pp. 270-284, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[40] Christina Lin, Stephany Liu, and Samuel Wong, "Counting Actor Screen Time in Movies," *Stanford University, CS230, Deep Learning*, 2019. [Google Scholar]

[41] Shankar Setty et al., "Indian Movie Face Database: A Benchmark for Face Recognition Under Wide Variations," *2013 Fourth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)*, Jodhpur, India, pp. 1-5, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[42] YouTube, 2026. [Online]. Available: https://www.youtube.com

[43] Internet Archive: Digital Library of Free & Borrowable Texts, Movies, Music & Wayback Machine, Archive.Org, 2026. [Online]. Available: https://archive.org

[44] Joseph Redmon et al., "You Only Look Once: Unified, Real-Time Object Detection," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 779-788, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[45] Prashant Udawant et al., "A Systematic Approach to Face Recognition using Convolutional Neural Network," *2024 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC)*, Bhubaneswar, India, pp. 1-6, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[46] Sanghyun Woo et al., "CBAM: Convolutional Block Attention Module," *Computer Vision-ECCV 2018: 15th European Conference*, Munich, Germany, pp. 3-19, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[47] Kaiming He et al., "Deep Residual Learning for Image Recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 770-778, 2016. [CrossRef] [Google Scholar] [Publisher Link]