CNN-Based Video Genre Classification System Using Frame Extraction, Visual Features, and Deep Learning Techniques

Dr. Pavithra A C[1] Mohammed Uwaiz Ahmed[2] Mohammed Zubair Durrani[3]
Syed Faisal Hashmi[4] Syed Mohammed Daniyaal[5]
[1,2,3,4,5]Department of Computer Science and Design
[1,2,3,4,5]ATME College of Engineering, India

Abstract — This work presents a CNN-based approach for automated video genre classification to address the growing need for scalable multimedia organization. Key frames are extracted from videos, preprocessed, and used to train a CNN model capable of learning visual patterns relevant to genre prediction. Optimization techniques such as data augmentation, batch normalization, and dropout are applied to improve generalization and reduce overfitting. The model is evaluated using standard performance metrics including accuracy, precision, recall, and F1-score, demonstrating clear improvements over traditional machine learning and handcrafted feature-based methods. Experimental results show that the proposed approach achieves reliable and consistent classification performance across multiple video genres. The system demonstrates strong potential for integration into real-world applications, including personalized recommendation systems, automated metadata generation, content indexing, and large-scale digital media management.

Keywords: Video Genre Classification, Convolutional Neural Networks (CNN), Key-Frame Extraction, Deep Learning, Multimedia Content Analysis, Automated Video Indexing

I. Introduction

The rapid expansion of digital media in recent years, supported by widespread smartphone use, high-speed internet connectivity, and large-scale streaming platforms, has resulted in unprecedented growth of video content. Platforms such as YouTube, Netflix, Amazon Prime, Hotstar, and various educational services generate vast quantities of videos daily. As a consequence, manual content annotation and genre categorization have become impractical, inconsistent, and non-scalable.

Automated video genre classification provides an effective solution by assigning a category such as action, drama, documentary, animation, or sports based on the visual and contextual characteristics of videos. Traditional machine learning approaches, which relied primarily on handcrafted features, struggled to generalize across diverse domains and large datasets. Recent advances in deep learning, particularly Convolutional Neural Networks (CNNs), have enabled automatic extraction of hierarchical visual representations, improving video understanding and classification accuracy significantly. CNN-based models have demonstrated strong performance across visual analysis tasks including image recognition, object detection, and action recognition, making them suitable for large-scale video classification applications.

II. Related Work

Earlier video classification methods used handcrafted features and traditional machine learning models, but these approaches lacked scalability and accuracy. With deep learning, CNN-based models became widely adopted due to their ability to automatically learn useful visual features. Researchers later introduced hybrid models such as CNN-LSTM and 3D CNNs to capture both spatial and temporal patterns. Recent work also explores multimodal learning and transfer learning to improve performance. Despite these advancements, many models remain computationally expensive, creating the need for simpler and efficient approaches. This project follows that direction by using a lightweight CNN model for effective video genre classification.

III. Methodology

A. System Architecture

The architecture consists of:

1) Input Layer: Raw videos from the UCF101 dataset are provided as input.
2) Frame Extraction: Key frames are extracted at regular intervals to reduce processing load while preserving meaningful information.
3) Preprocessing: Extracted frames are resized, normalized, and formatted for model training.
4) CNN Feature Extraction: The CNN identifies spatial features such as shapes, textures, and visual patterns.
5) Classification Layer: Fully connected layers and a Softmax function predict the appropriate genre.
6) Output: The final predicted label is generated for use in tagging or indexing applications.

B. Data Description

- Video frames extracted at fixed intervals serve as the main training samples.
- Frames are resized (e.g., 128×128) and normalized to ensure consistency.
- The dataset is split into training (80%) and validation (20%) sets.
- Each frame inherits its label from the source video for supervised learning.

C. Model Logic

- The model uses CNN layers to automatically learn spatial features from video frames.
- Dropout and augmentation are applied to improve generalization and prevent overfitting.
- The extracted features are processed and mapped to the corresponding genre label during training.
- The highest probability class is selected, and performance is evaluated using standard metrics.

Overall, the model learns visual patterns from frames and uses them to accurately classify video genres.

IV. RESULTS AND DISCUSSION

Experimental results show that the CNN model is able to learn meaningful visual patterns from the extracted frames and classify most video categories correctly. The training and validation results indicate stable learning behavior, suggesting that the preprocessing steps and model structure are appropriate for this task.

However, since the model uses a basic 2D CNN and relies only on single frames rather than motion sequences, it may struggle with classes that look visually similar. While the results are satisfactory for a simple implementation, using advanced methods such as 3D CNNs or sequence-based models could improve accuracy in future work.

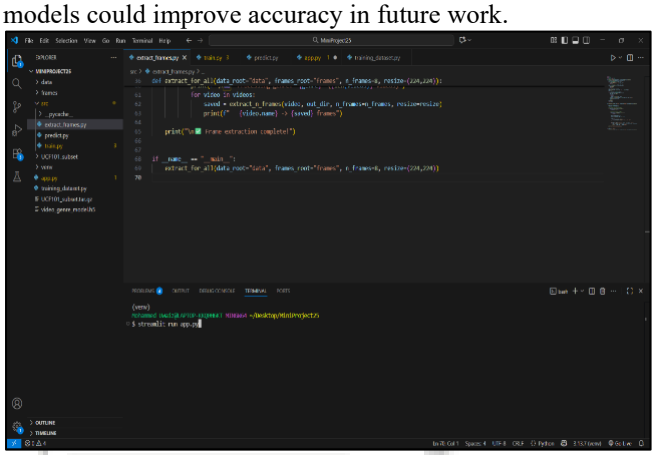

Fig. 1: Deploying the project locally

Deploying the project locally means setting up all required software, dependencies, and running the trained model directly on a personal system. This enables offline testing, faster execution, and full control over debugging and configuration without relying on a server or internet connection.

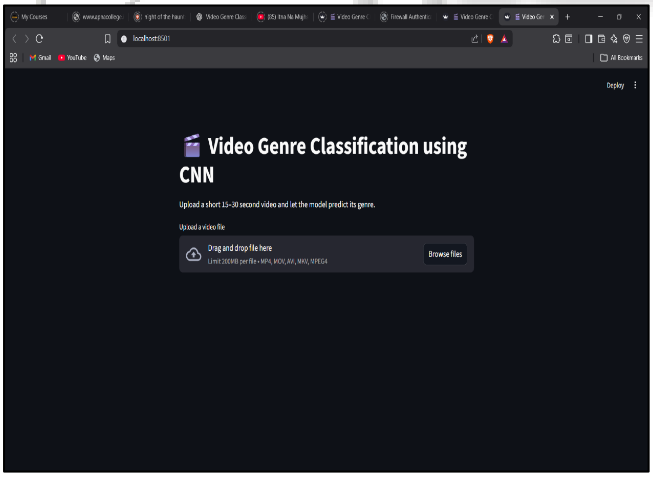

Fig. 2: Landing page on web

A landing page serves as the main interface where users can upload or select videos for testing. It shows basic project details and displays the predicted genre after processing, ensuring a simple and user-friendly experience.

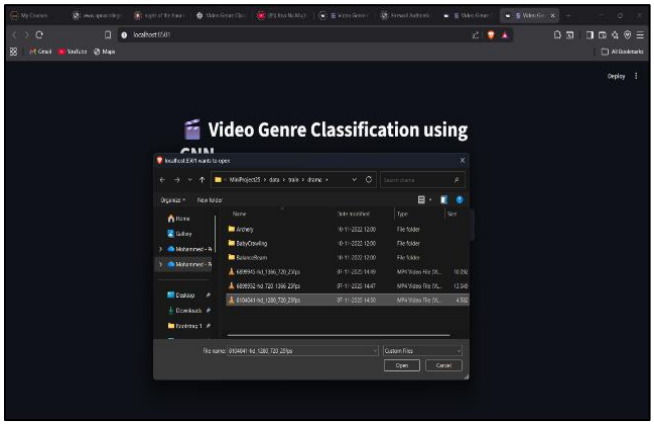

Fig. 3: Browsing directories to upload file

Browsing directories to upload a file lets users easily select videos from their device using a file picker. This makes the upload process simple, intuitive, and accessible, allowing the selected file to be processed for classification.

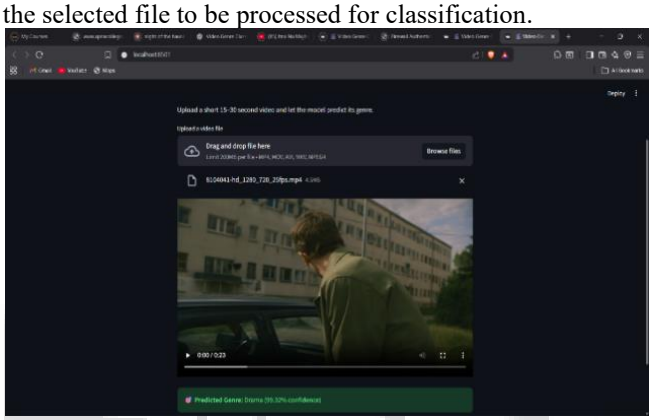

Fig. 4: Successful model prediction on upload of file

Successful model prediction means the uploaded video is processed correctly and its genre is accurately displayed. This confirms the system is working as expected and provides users with quick, clear feedback.

V. CONCLUSION

This project shows that a basic CNN model can classify video genres using extracted frames and spatial features. While the approach is simple, it provides reliable results and demonstrates that deep learning can automate genre recognition without manual feature engineering. The model also serves as a starting point for building more advanced video classification systems.

VI. FUTURE SCOPE

- Use advanced models like 3D CNNs to include motion information.
- Apply transfer learning to improve accuracy.
- Train with a larger dataset for better results.
- Make the model suitable for real-time use.
- Reduce model size so it can run on mobile or low-power devices.

REFERENCES

[1] Lakshmi, K.P., Solanki, M., Dara, J.S., Kompalli, A.B. (2024). Video genre classification using convolutional recurrent neural networks. *International Journal of*

Advanced Computer Science and Applications, 11(3), 156–162. Available at: https://doi.org/10.14569/IJACSA.2020.0110321

[2] Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., Fei-Fei, L. (2025). Large-scale video classification with convolutional neural networks. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1725–1732. Available at: https://doi.org/10.1109/CVPR.2014.223

[3] Yadav, A., Vishwakarma, D.K. (2024). A unified deep learning framework for movie genre classification. *Applied Soft Computing*, 96, 106624. Available at: https://doi.org/10.1016/j.asoc.2020.106624

[4] Rehman, A., Belhaouari, S.B. (2023). Deep learning for video classification: A review. *TechRxiv, Preprint*, 1–28. Available at: https://doi.org/10.36227/techrxiv.15172920.v1

[5] Shao, Y., Guo, J., Xu, M. (2024). Recognizing online video genres using ensemble deep convolutional learning. *Journal of Cloud Computing*, 13(6), 1–15. Available at: https://doi.org/10.1186/s13677-024-00664-2

[6] Ng, J.Y.H., Hausknecht, M., Vijayanarasimhan, S., Vinyals, O., Monga, R., Toderici, G. (2025). Beyond short snippets: Deep networks for video classification. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 4694–4702. Available at: https://doi.org/10.1109/CVPR.2015.7299101

[7] Jiang, Y-G., Wu, Z., Tang, J., Xue, X., Chang, S-F. (2025). Modeling multimodal cues in a hybrid deep learning framework for video classification. *ACM Multimedia*, 546–558. Available at: https://doi.org/10.1145/3123266.3123381

[8] Ünal, F.Z., Guzel, M.S., Bostanci, E., Acici, K., Asuroglu, T. (2023). Multilabel genre prediction using deep-learning frameworks. *Applied Sciences*, 13(15), 8665. Available at: https://doi.org/10.3390/app13158665

[9] Balaji, M., Mohan, T., Karthik, A. (2024). A multimodal deep learning approach for movie genre classification. *International Journal of Engineering Research and Technology*, 9(6), 112–119. Available at: https://doi.org/10.17577/IJERTV9IS060265

[10] Ramesh, M., Kumar, S., Singh, P. (2023). Sports video classification framework using enhanced deep learning. *Journal of Intelligent Systems*, 32(2), 221–232. Available at: https://doi.org/10.1515/jisys-2022-0021

[11] Ji, S., Xu, W., Yang, M., Yu, K. (2024). 3D convolutional neural networks for human action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1), 221–231. Available at: https://doi.org/10.1109/TPAMI.2012.59

[12] Simonyan, K., Zisserman, A. (2025). Two-stream convolutional networks for action recognition in videos. *Advances in Neural Information Processing Systems. https://arxiv.org/abs/1406.2199*

