DAYANANDA SAGAR COLLEGE OF ENGINEERING

(An Autonomous Institute Affiliated to VTU, Belagavi. Approved by AICTE & ISO 9001:2015 Certified)
Shavige Malleshwara Hills, Kumaraswamy Layout, Bengaluru-560111)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Deep Learning AAT Report 2024-25 (EVEN)

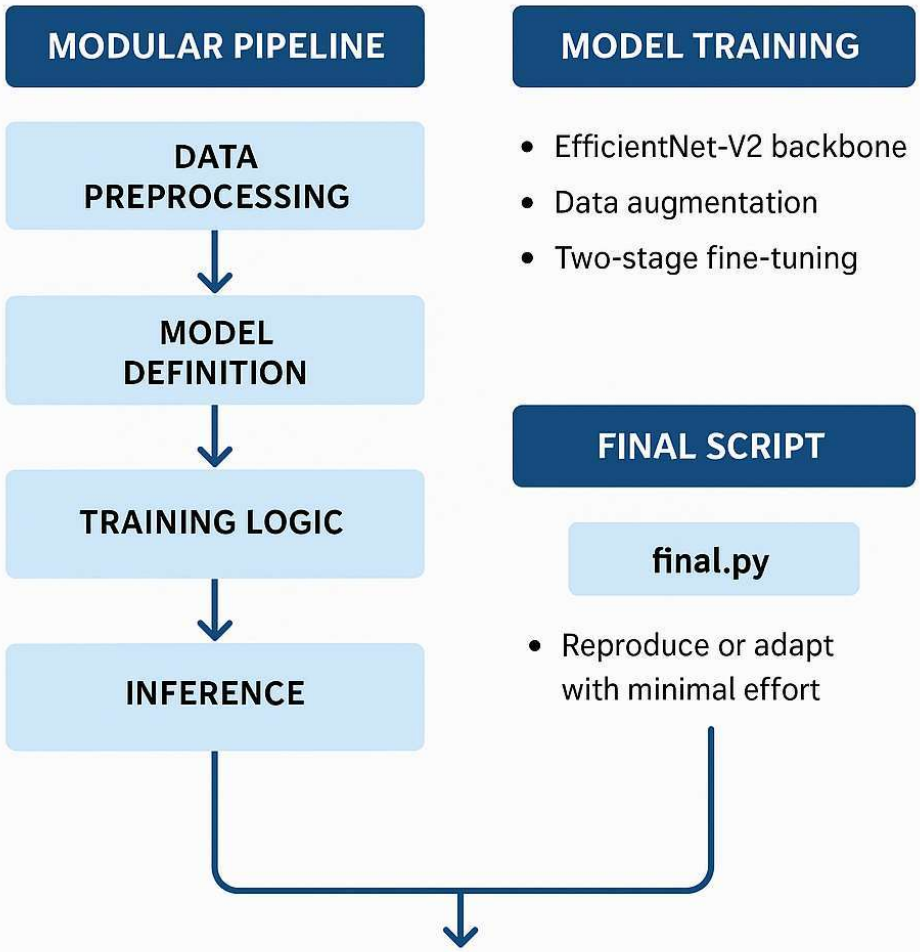
Course Faculty: Dr. Anusha Preetham

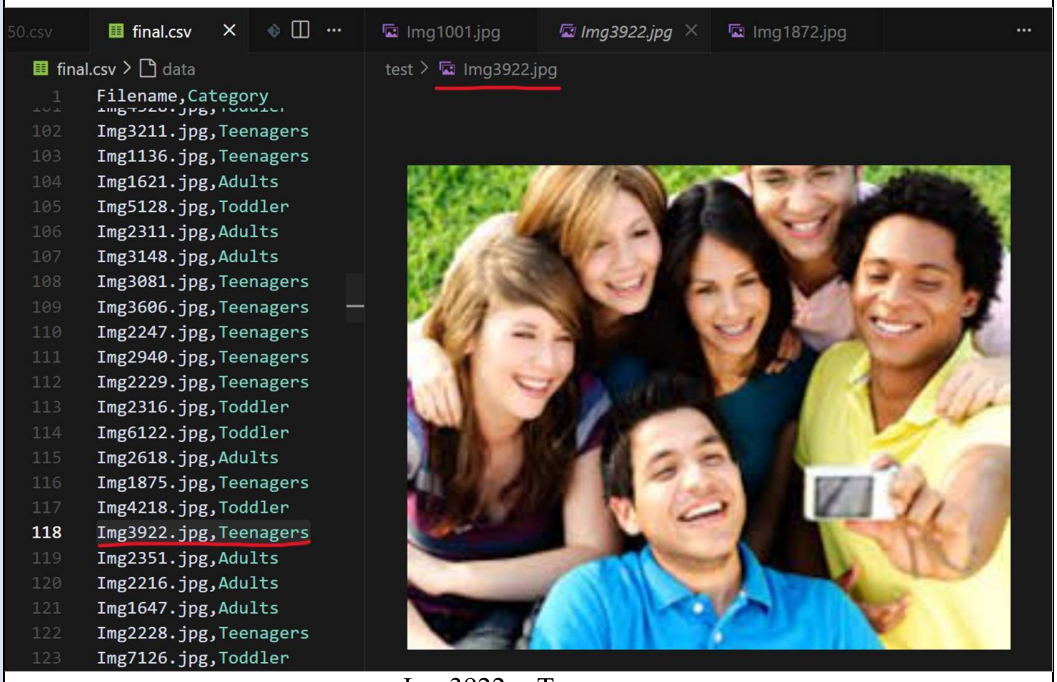
Course Name & code: Deep Learning 22CS63

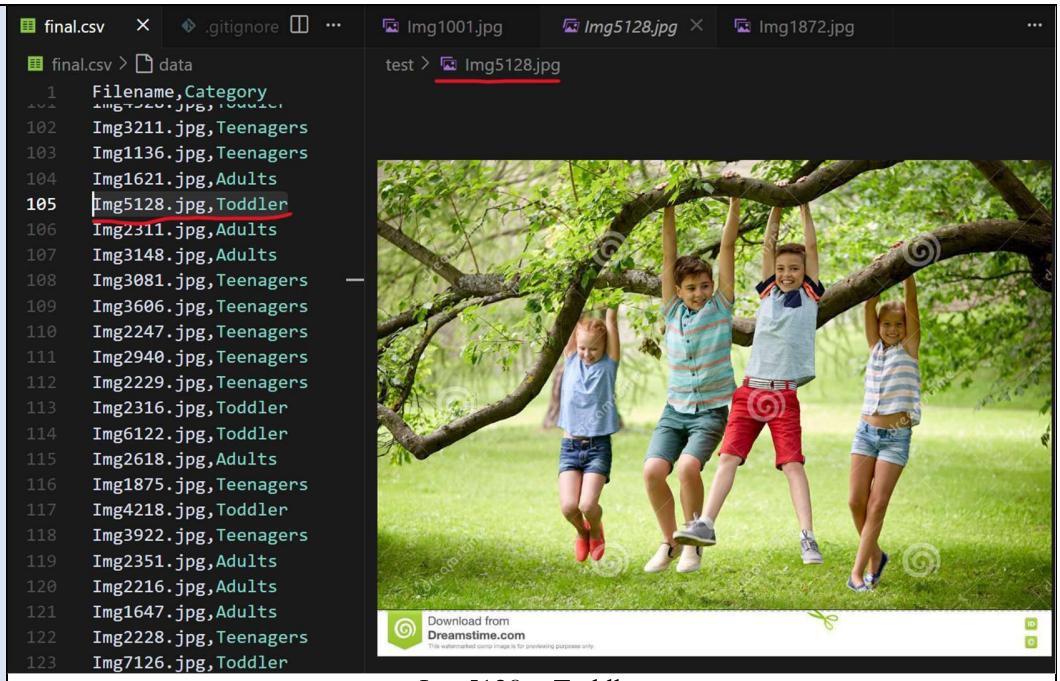
Semester: 6th Section: A

TITLE OF THE PROJECT	"Dance Form Classification using MobileNetV2 with Real-Time Image and Augmentation for Enhanced Generalization"				
STUDENT NAME	A M Tanushi	Aastha Kanaujia	Abhijit Biswal	Amit Kumar Singh	Ankit Kumar
USN	1DS22CS001	1DS22CS003	1DS22CS005	1DS22CS029	1DS22CS033
INDIVIDUAL CONTRIBUTION	Data Collection & Preparation	Data Augmentation & Input Pipeline	Model Arc2hitecture & Transfer-Learning Strategy	Training, Fine-Tuning & Evaluation	Inference, Result Analysis & Submission Generation
ABSTRACT OF THE WORK	This project presents an automated deep learning system designed to classify group photos from friendship gatherings into three distinct life-stage categories: Toddler, Teenagers, or Adults. It leverages a pre-trained EfficientNet-V2 backbone known for its high performance and computational efficiency. To improve generalization, the dataset undergoes extensive augmentation including horizontal flips, random crops, rotations, and colour jitter. The goal is to build a robust, lightweight classifier that performs well on diverse and unseen images while maintaining real-time inference capabilities. The training strategy follows a two-phase approach. In the first phase, the EfficientNet-V2 backbone is frozen, and only the classification head is trained to adapt the model quickly to the new dataset. In the second phase, the entire model is unfrozen and fine-tuned with a lower learning rate to improve task-specific accuracy. Validation is conducted using a stratified split and performance metrics like accuracy and confusion matrix, ensuring balanced learning across all three classes and preventing overfitting. After training, the model processes unseen test images to predict their respective life-stage categories. These predictions are compiled into a CSV file, formatted for competition submission. The final solution demonstrates strong performance with efficient computation, making it suitable for integration into larger AI systems like smart photo organization tools, demographic analytics, or real-time social media applications.				

INTRODUCTION	Detecting the dominant age group in social photographs holds significant value across domains such as demographic research, targeted marketing, and sociological analysis. Group photos often capture key visual cues—like facial features, posture, and clothing—that can hint at the life stage of the individuals present. Automatically identifying whether a photo primarily contains toddlers, teenagers, or adults enables scalable data analysis across large unstructured image collections. However, traditional image classification methods demand vast labeled datasets and involve intricate feature engineering, which is both time-consuming and resource-intensive. To address these limitations, this project embraces transfer learning, leveraging EfficientNet-V2—a compact, high-performance convolutional neural network pre-trained on ImageNet. Instead of training from scratch, we reuse the network's learned visual representations and fine-tune them for our specific task. This drastically reduces training time and data requirements while preserving high accuracy. The model classifies group photos into three categories: Toddler, Teenagers, or Adults, based on the dominant age group in the image. The training pipeline follows a two-stage strategy for optimal learning efficiency. Initially, the EfficientNet-V2 backbone is frozen, allowing only the classification head to learn task-specific patterns. Once the top layers are trained, we unfreeze the full model and fine-tune it at a lower learning rate. This controlled approach helps retain general features from pre-training while adapting deeper layers to the new data. Data augmentation techniques—like flipping, rotation, and color jitter—enhance robustness to variability in lighting, orientation, and background conditions. A key goal of this project is to provide a reproducible and modular pipeline. The entire process—from data preprocessing and model training to prediction and CSV generation—is self-contained and designed for command-line execution. Clear configuration files and CLI scripts make it easy for users to retrain the model with new datasets, adjust hyperparameters, and run predictions on unseen images. This flexibility supports both experimentation and deployment in real-world scenarios. Ultimately, the system achieves high accuracy with modest computational requirements, making it suitable for use on standard GPUs or even edge devices. Its streamlined CLI and minimal setup effort make it accessible to researchers, engineers, and analysts alike. The final model can be deployed in applications ranging from automated image tagging and smart photo albums to audience segmentation tools for media and advertising.

DESIGN	
PLATFORM USED	• Programming language: Python 3.10+ • Deep-learning library: TensorFlow 2 (Keras API) with EfficientNet-V2 models • Data manipulation & utilities: NumPy, pandas, Pillow, matplotlib • Hardware: GPU-accelerated execution if available (memory-growth enabled), otherwise CPU • Package management: requirements.txt for deterministic environment recreation
CODE LINK (GITHUB/ GOOGLE DRIVE)	https://github.com/amitkumarsingh01/Friendship_goal_DL.git

CONCLUSION	The project architecture is built with modularity at its core, clearly separating data preprocessing, model definition, training logic, and inference into independent components. This structured design not only simplifies code maintenance but also enables quick experimentation with different configurations or datasets. Such a separation ensures that each part of the pipeline can be debugged, extended, or reused without affecting the rest of the system, making it ideal for both research and production settings. To maximize performance on a relatively small dataset, the pipeline employs judicious data augmentation techniques such as random flips, rotations, and color distortions to simulate real-world variability. Combined with a two-stage fine-tuning process, where the EfficientNet-V2 backbone is first frozen and then gradually unfrozen for full fine-tuning, the model achieves robust generalization across all three classes—Toddler, Teenagers, and Adults. This careful training strategy allows the model to extract meaningful age-related features without overfitting. The entire workflow is wrapped in a user-friendly script called final.py, which handles everything from training to CSV submission generation. This script makes it simple for other researchers or practitioners to reproduce results, test new hypotheses, or adapt the model to related classification problems, such as estimating age ranges in individual portraits or family photos. With minimal configuration, users can deploy the pipeline end-to-end, enabling scalable, real-world use of deep learning for demographic image analysis.
SNAPSHOTS OF RESULTS	 Img3922 = Teenagers



Img5128 = Toddler

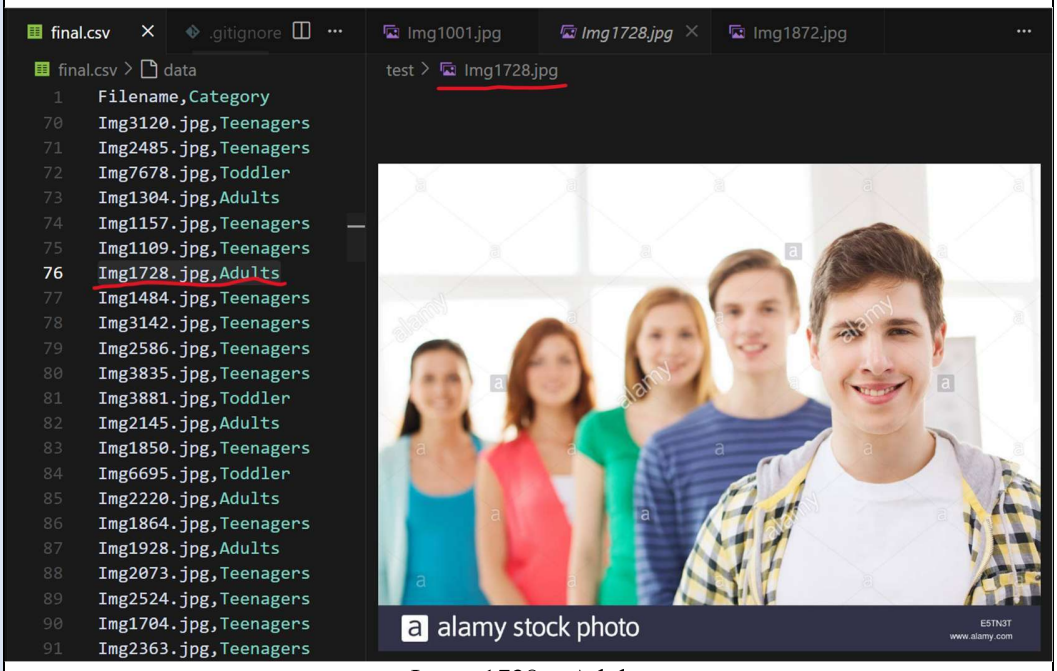


Image1728 = Adults