1. YOLO

Dataset Description:

1. **Source: Combination of 2 datasets:**
    - Dataset A: Provided by the project team; ~100 images were manually annotated using bounding boxes and gender labels (male / female).
    - Dataset B: Imported from Roboflow; pre-annotated with bounding boxes and gender classes.
2. **Annotation Format:**
    - YOLOv11-compatible .txt files for each image.
    - Each line in a label file:
        - `<class_id> <x_center> <y_center> <width> <height>`
3. **Object Classes:**
    - 0 → Male
    - 1 → Female
4. **Image Resolution:**

Varied across datasets; resized to 640×640 during training with padding where necessary.

5. **Dataset Statistics:**

Train: 1872

Validate: 225

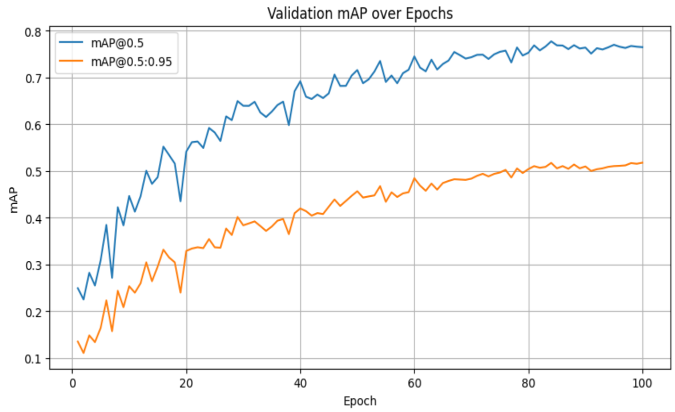
Model Training Details:

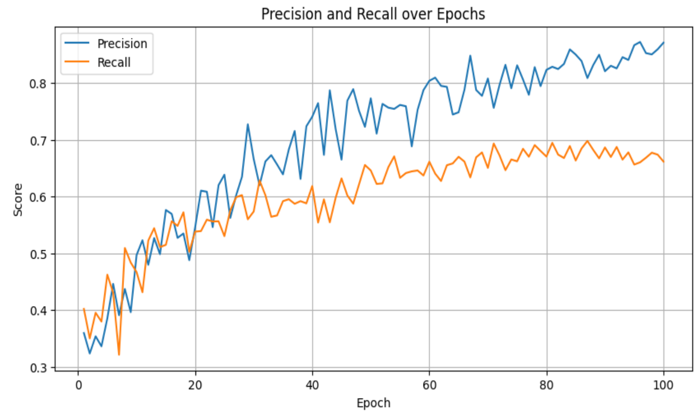
- **Model Used:** YOLOv11n (YOLO version 11 nano)
  Chosen for its light weight and fast inference while maintaining good performance on small datasets.
- **Training Framework:** PyTorch using Ultralytics YOLOv11
- **Training Location:** Kaggle Notebook
- **Epochs:** 100
- **Image size:** 640×640
- **Batch size:** 16
- **Device:** GPU (Tesla T4, 16 GB VRAM)

Training Performance Summary:

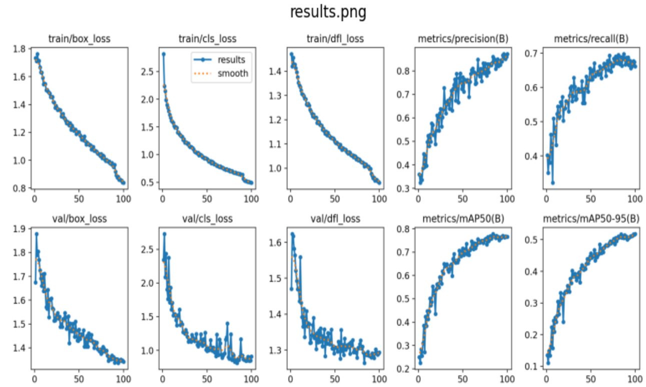
Class	Precision	Recall	mAP@0.5	mAP@0.5:0.95
All	0.857	0.689	0.778	0.518
Male	0.856	0.717	0.800	0.548
Female	0.858	0.661	0.755	0.487

Training Results:

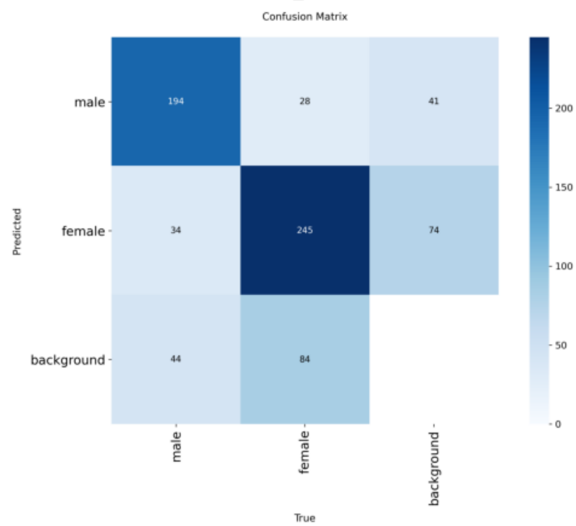




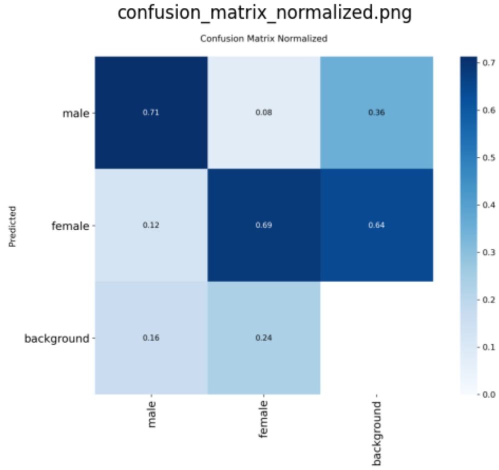
results.png



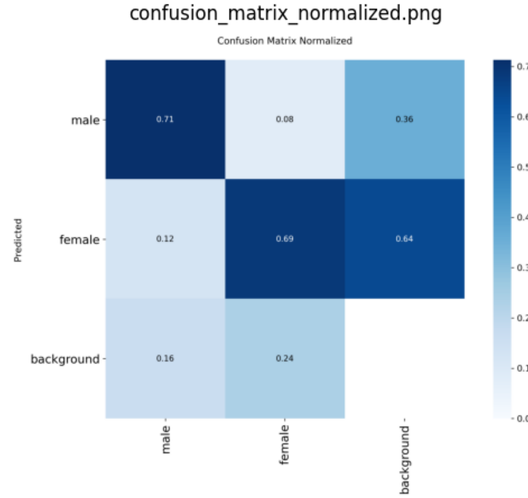
confusion_matrix.png



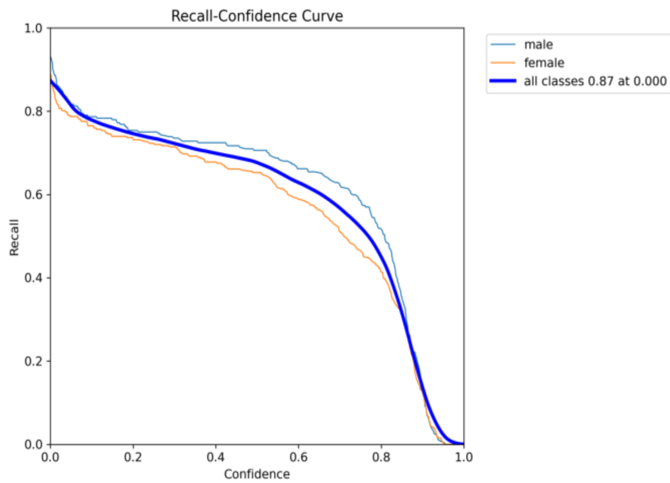
confusion_matrix_normalized.png



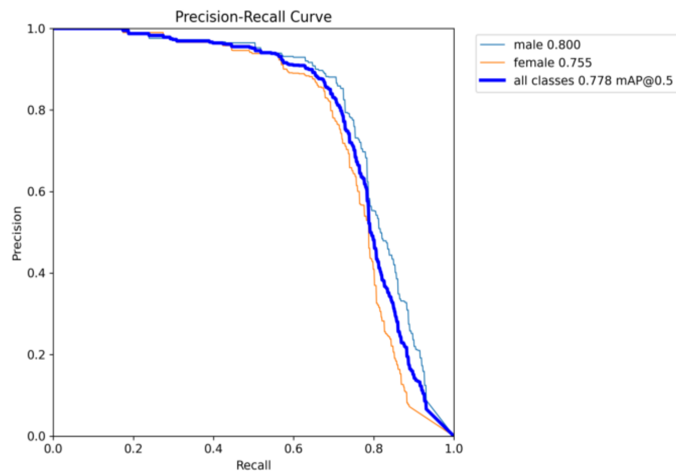
confusion_matrix_normalized.png



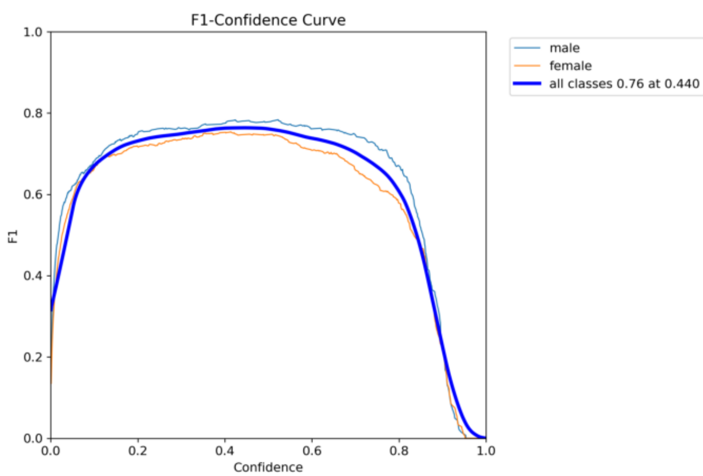
BoxR_curve.png



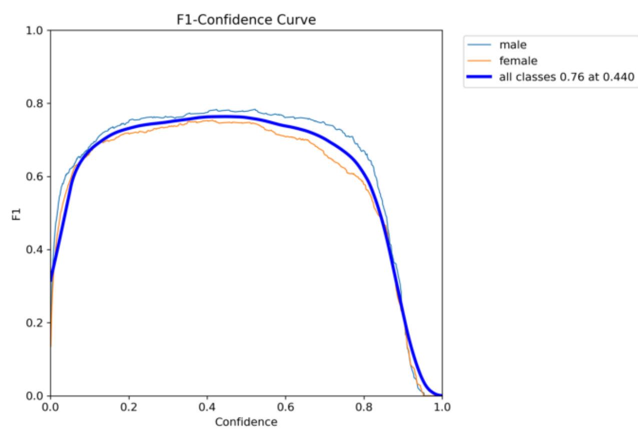
BoxPR_curve.png



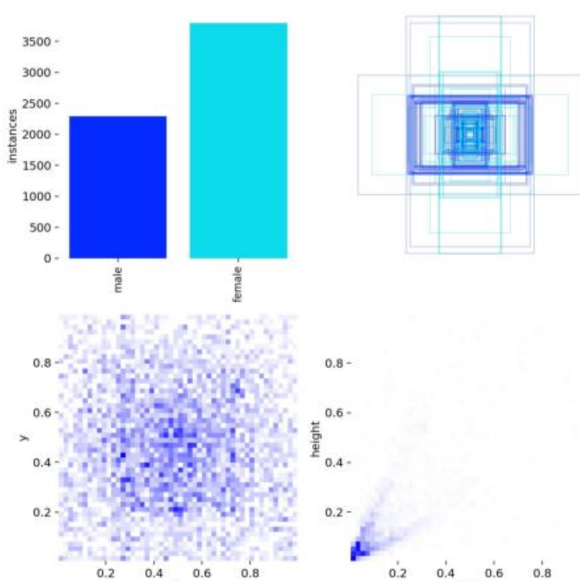
BoxF1_curve.png



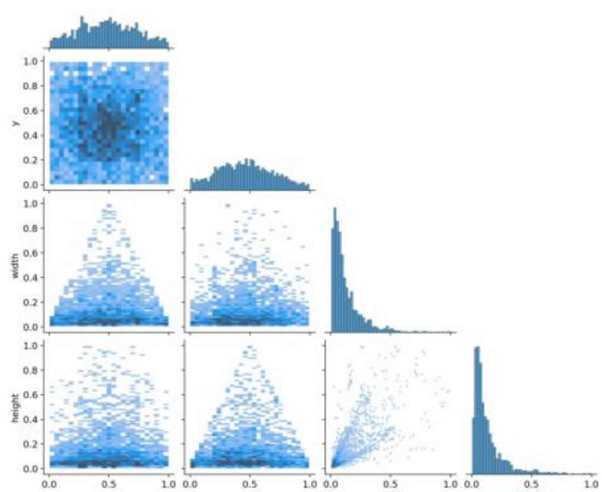
BoxF1_curve.png



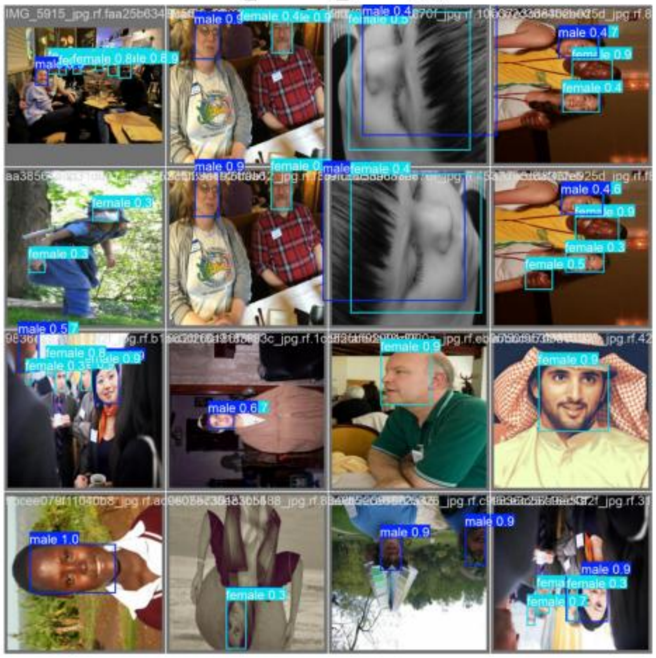
labels.jpg



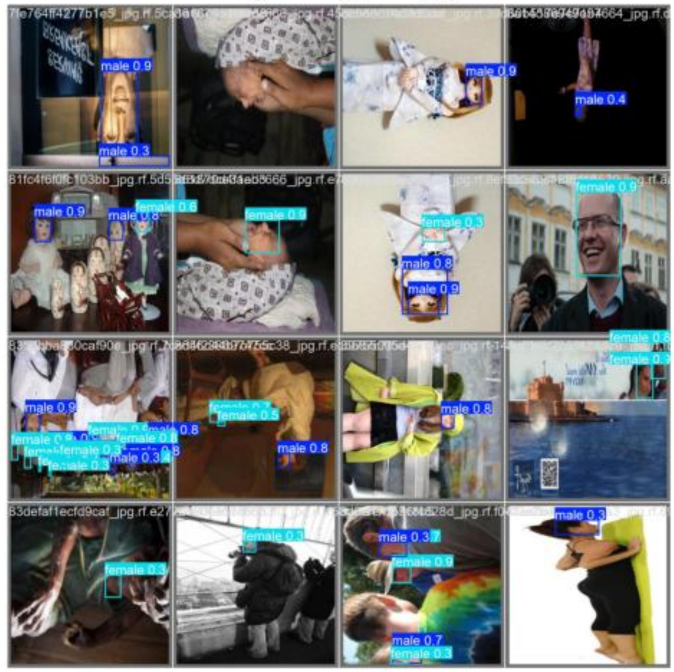
labels_correlogram.jpg



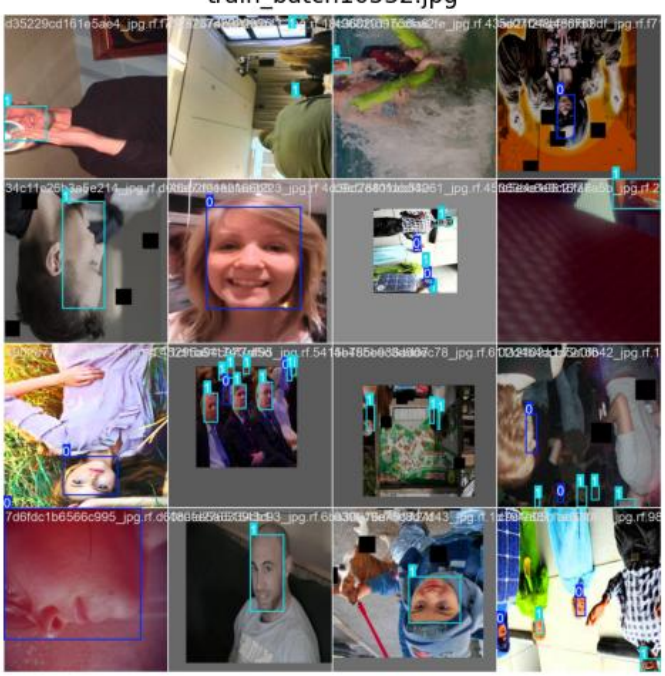
val_batch0_pred.jpg



val_batch1_pred.jpg



train_batch10532.jpg



train_batch10531.jpg

2. Classifier

Model Architecture:

Base Backbone: EfficientNet_V2_S, pre-trained on ImageNet1K (IMAGENET1K_V1 weights).

- Custom Classifier Head:
- Dropout(0.5)
- Linear(1280 → 512)
- ReLU
- Dropout(0.3)
- Linear (512 → 16) (final output layer for 16 celebrity classes)

This structure balances transfer learning benefits with custom fine-tuning capacity for the specific celebrity face task.

Dataset Details:

1. **Data Format:**
    - Data downloaded from Kaggle competition page.
    - Images and labels loaded from .npy files (faces_cropped.npy, labels_cropped.npy)
    - Images are already cropped face regions of shape (224, 224, 3), RGB.
2. **Label Remapping:**

Labels with value -1 (used for unknowns) were reassigned to class index 15 to maintain label consistency within 0–15.

Data Augmentations:

Data augmentation was applied using Albumentations to enhance generalization and robustness:

- HorizontalFlip (p=0.5)
- Rotate (limit=30°, p=0.5)
- HueSaturationValue (hue ±20, sat ±30, val ±20, p=0.7)
- RandomBrightnessContrast (brightness ±0.3, contrast ±0.3, p=0.7)
- CoarseDropout: 1–2 holes, each 10–20% of image area (black mask)
- Normalization: mean=[0.485, 0.456, 0.406], std=[0.229, 0.224, 0.225]
- Conversion to tensor with ToTensorV2()

Training Strategy:

Phase 1: Train Classifier Head Only

- Freeze all layers except the classifier head.
- Optimizer: AdamW(lr=1e-3, weight_decay=1e-4)
- Epochs: 100
- Loss Function: CrossEntropyLoss with MixUp
- Accuracy was monitored on a hold-out validation set (20%).

➢ Phase 2: Fine-Tune Last Feature Blocks + Head

- Only unfreeze the last 5 blocks of EfficientNet and the classifier.
- Optimizer: AdamW(lr=1e-4, weight_decay=1e-5)
- Epochs: 50
- Same loss and sampling strategy used.

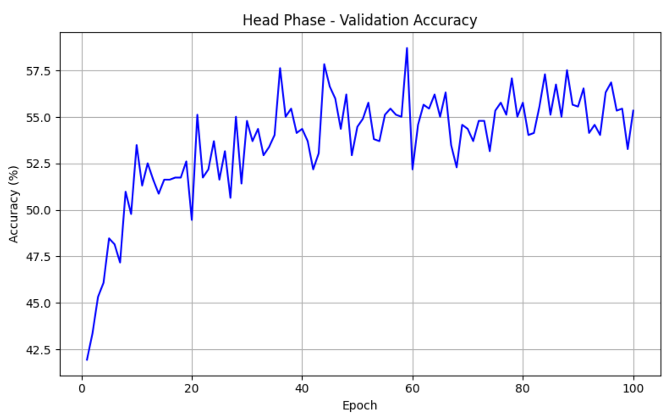
➢ MixUp Regularization

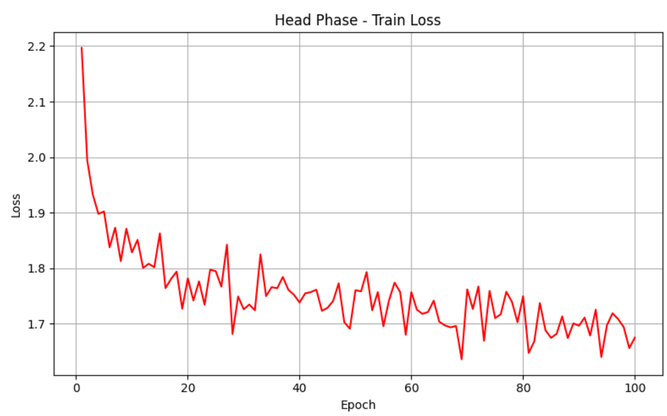
- MixUp was applied during training
- Hyperparameter: alpha=0.4 (Beta distribution)
- Loss computed as:

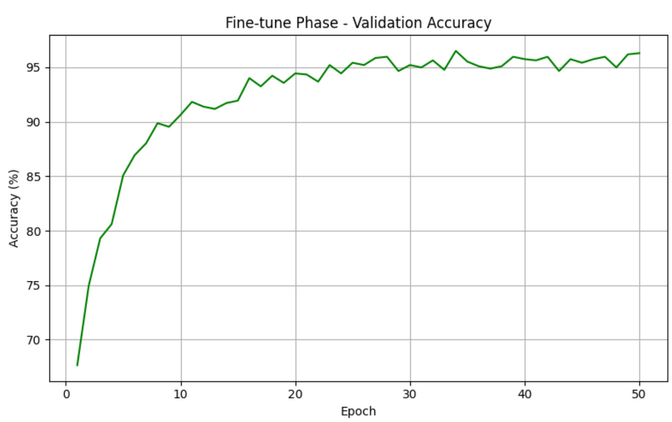
```
lam * CE(preds, y_a) + (1 - lam) * CE(preds, y_b)
```

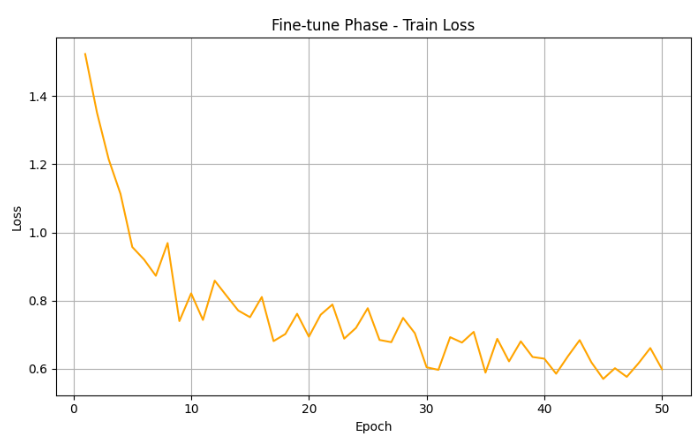
Training Results:

- Training ran for a total of 150 epochs across two phases.
- Model consistently achieved strong validation accuracy (>99%).
- This model was later used to classify cropped YOLO predictions during evaluation and visualization.



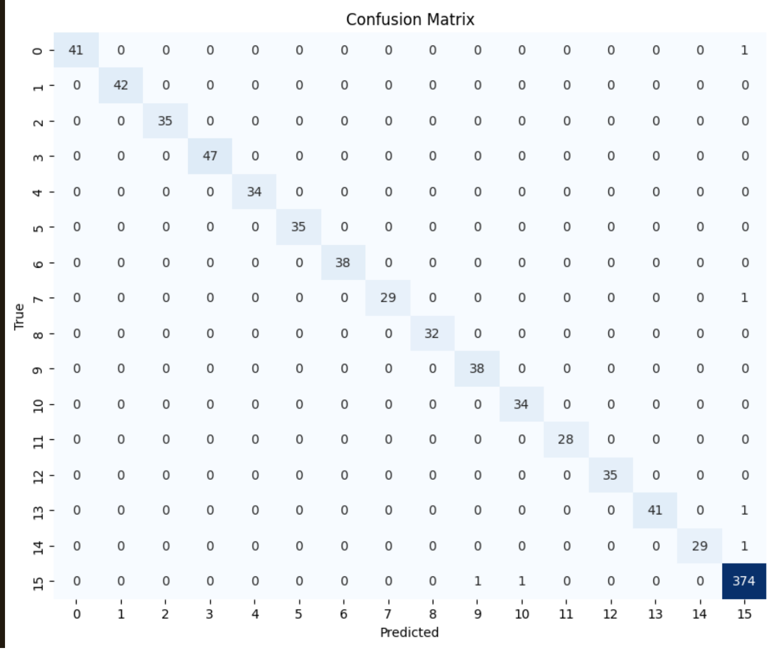







```
Top-1 Accuracy: 99.35%
Top-3 Accuracy: 100.00%
Macro F1 Score: 0.9935

→ Final results on validation set:
   • Top-1 Acc: 99.35%
   • Top-3 Acc: 100.00%
   • Weighted F1: 0.9935
```

```
Classification Report:
              precision    recall  f1-score   support

           0     1.0000    0.9762    0.9880        42
           1     1.0000    1.0000    1.0000        42
           2     1.0000    1.0000    1.0000        35
           3     1.0000    1.0000    1.0000        47
           4     1.0000    1.0000    1.0000        34
           5     1.0000    1.0000    1.0000        35
           6     1.0000    1.0000    1.0000        38
           7     1.0000    0.9667    0.9831        30
           8     1.0000    1.0000    1.0000        32
           9     0.9744    1.0000    0.9870        38
          10     0.9714    1.0000    0.9855        34
          11     1.0000    1.0000    1.0000        28
          12     1.0000    1.0000    1.0000        35
          13     1.0000    0.9762    0.9880        42
          14     1.0000    0.9667    0.9831        30
          15     0.9894    0.9947    0.9920       376

    accuracy                         0.9935       918
   macro avg     0.9960    0.9925    0.9942       918
weighted avg     0.9935    0.9935    0.9935       918

Top-1 Accuracy: 99.35%
Top-3 Accuracy: 100.00%
```



Conclusion:

This project involved two integrated models: a YOLOv11n object detector for gender detection and an **EfficientNet-based classifier** for celebrity recognition.

1. The **YOLOv11n model**, trained on a custom gender-labeled dataset (~2,000 images), achieved solid detection performance with:
   - Precision: 85.7%
   - Recall: 68.9%
   - mAP@0.5: 77.8%
   - mAP@0.5:0.95: 51.8%

These metrics indicate reliable detection, particularly for the male class, with lightweight architecture optimized for fast inference on small datasets.

2. The **EfficientNet_V2_S** classifier, fine-tuned in two stages using MixUp regularization and strong augmentations, reached >99% validation accuracy. It effectively learned to classify 16 celebrity identities from cropped facial images, even handling unknowns by remapping them to a dedicated class.

Together, these models form a robust pipeline for gender and identity recognition with efficient computation, high accuracy, and resilience to data variation—well-suited for real-world deployment or downstream analysis.

Validation:


Fawad Khan
Sajal Ali
Unknown
Ramsha Khan
Atif Aslam
Atif Aslam
Unknown
Atif Aslam
Unknown
Unknown
Fawad Khan
Unknown
Shaheen Shah Afridi
Unknown
Sajal Ali
Unknown
Unknown
Unknown
Unknown
Unknown


Hania Amir
Hamuyun Saeed
Unknown
Hamza Ali Abbasi
Neelam Muneer