ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG QUỐC TẾ
VNU-INTERNATIONAL SCHOOL

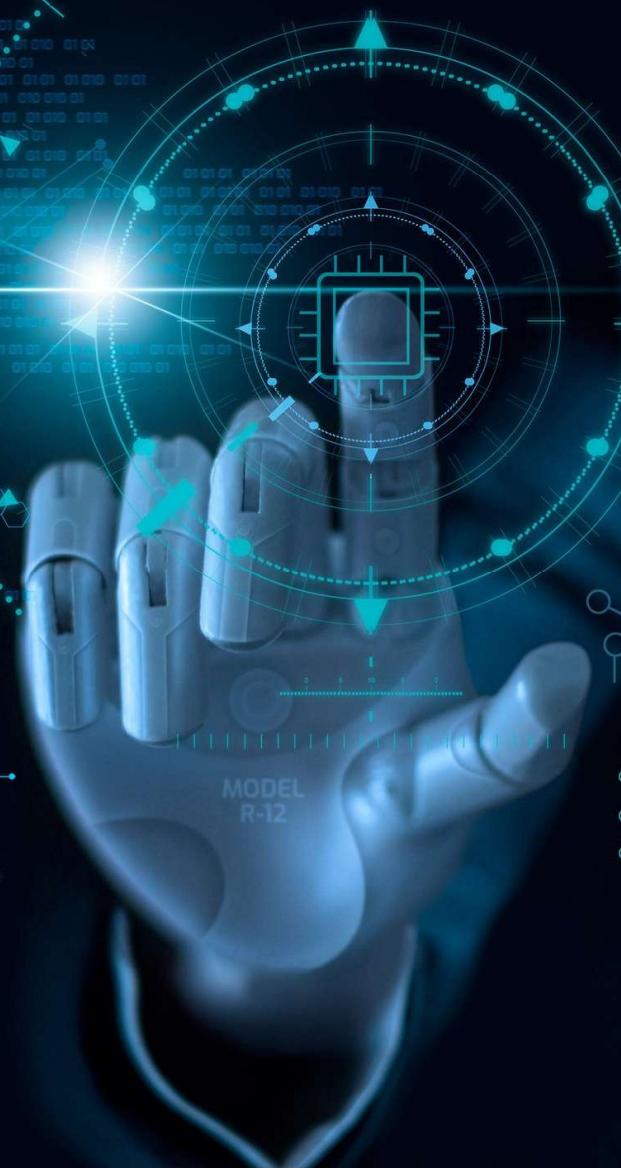
# ARTIFICIAL INTELLIGENCE

## PROJECT NAME

## Facial Image-Based Prediction System for Vietnamese Beauty Pageant Qualification

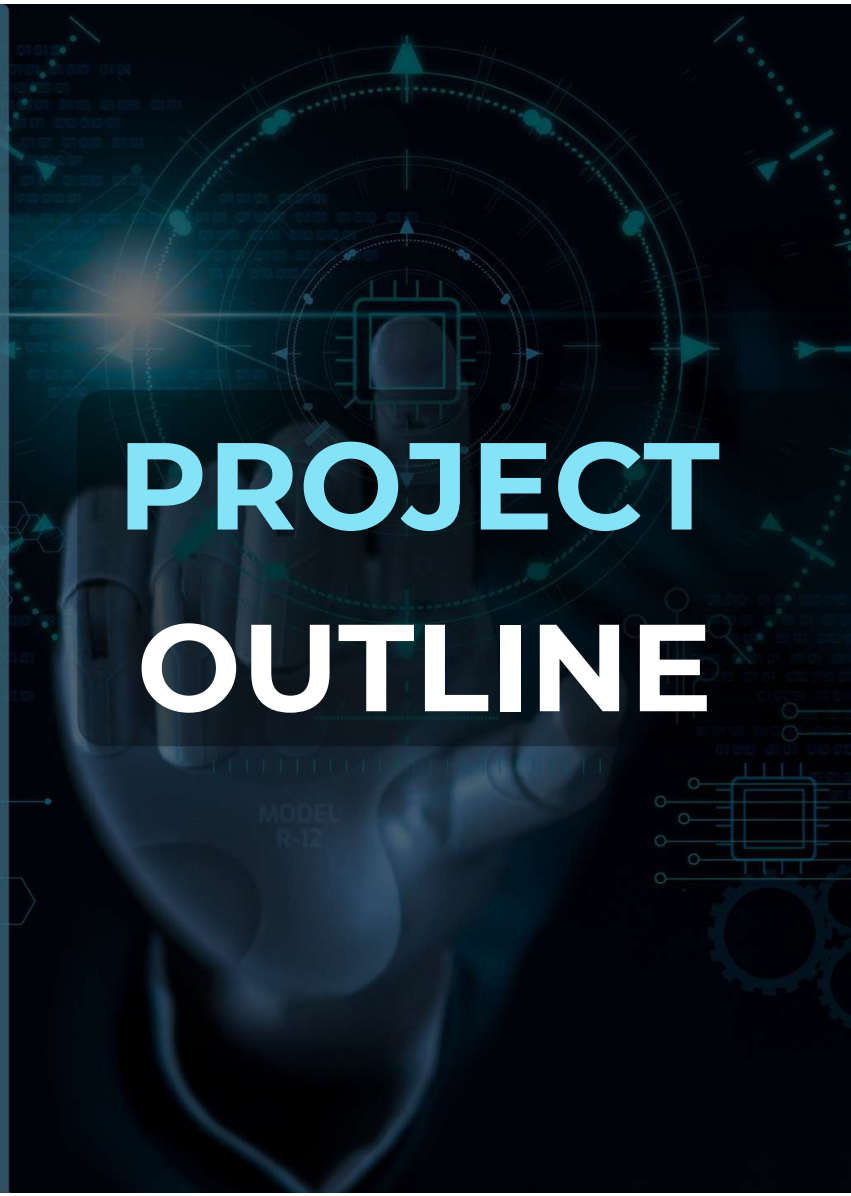

Lecturer: Ph.D Ha Manh Hung

Subject Code: INS308002

## List of Content





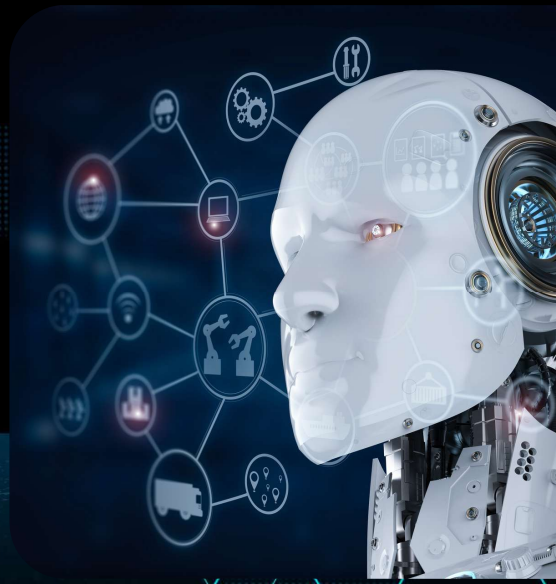
# PROJECT OUTLINE

# I. Introduction

# Introduction

## 01 Industry Background

- Facial beauty has long fascinated humanity, from ancient times to the modern day.
- Research shows that features such as symmetry, averageness, and clear skin are strongly linked to perceived attractiveness.
- While beauty can be subjective, there is evidence of cross-cultural agreement, suggesting some objective elements.
- Previous machine learning models have struggled to capture human-like perception of beauty.



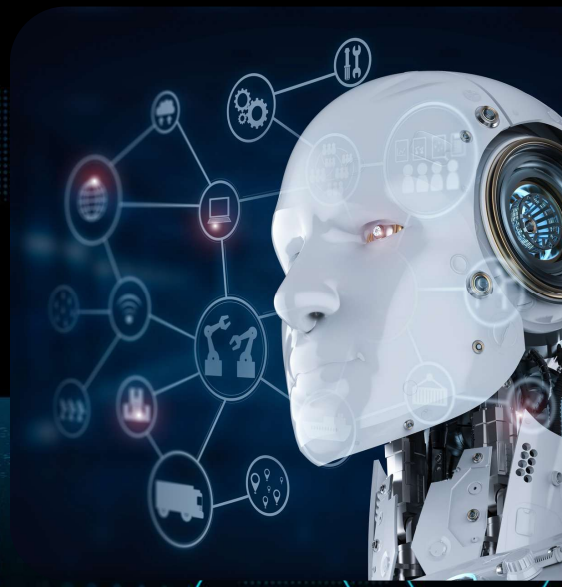
## 02 Problem Statement

- There is a lack of deep learning models trained on culturally relevant beauty standards in Vietnam.
- Existing models are limited in both feature extraction and cultural sensitivity.
- Key challenge: How to build a model that accurately predicts facial beauty based on Vietnamese aesthetic preferences?

# Introduction

## 03 Motivation

- Advances in AI and computer vision enable new methods for automated beauty assessment.
- This project focuses on Vietnamese standards of beauty, unlike prior works based on global datasets.
- Data is collected from national beauty pageants and contestants' social media profiles.
- Potential applications include AI beauty tech, smart content moderation, and cultural research.



## 04 Objectives

- Develop a deep learning model to predict beauty scores from labeled portrait images.
- Evaluate state-of-the-art architectures like MobileNetV2, ResNet50, and EfficientNetB3.
- Explore use cases in virtual makeup, personalized filters, and aesthetic image curation.
- Promote ethical awareness and social responsibility in applying AI to subjective traits like beauty.


II. Dataset

# Dataset Collection

## Source:

Large-scale beauty contests held in Vietnam

## Custom Dataset

1,000+ high-quality frontal portraits, Neutral lighting, minimal background distractions

## Systematic file naming convention

Format: [PageantAbbreviation][Year][Index].jpg

Example: MU20241.jpg = Miss Universe Vietnam 2024, Image #1

# Attractiveness Rating and Labeling Scheme

## Labeling based on official placement in beauty pageants:

🏅 Scoring rules:

- 5 – Winner ★★★★★
- 4 – 1st & 2nd Runner-up ★★★★☆
- 3 – Top 3–10 ★★★☆☆
- 2 – Top 11–20 ★★☆☆☆
- 1 – From top 20 onwards ★☆☆☆☆

## Structured rating file: label.csv

📄 Columns: Filename, Rating, Original Rating

⟶ Used for model training & evaluation

# Data preprocessing

## Data Loading

- download_data(): extract image
- get_labels_dict(): file name mapping  rating
- Data labeled from 1–5

## Image Processing

- Convert BGR → RGB
- Resize to 350×350 pixels
- Normalization → [0–1]
- Save as X (image) & y (label)

## Data Splitting

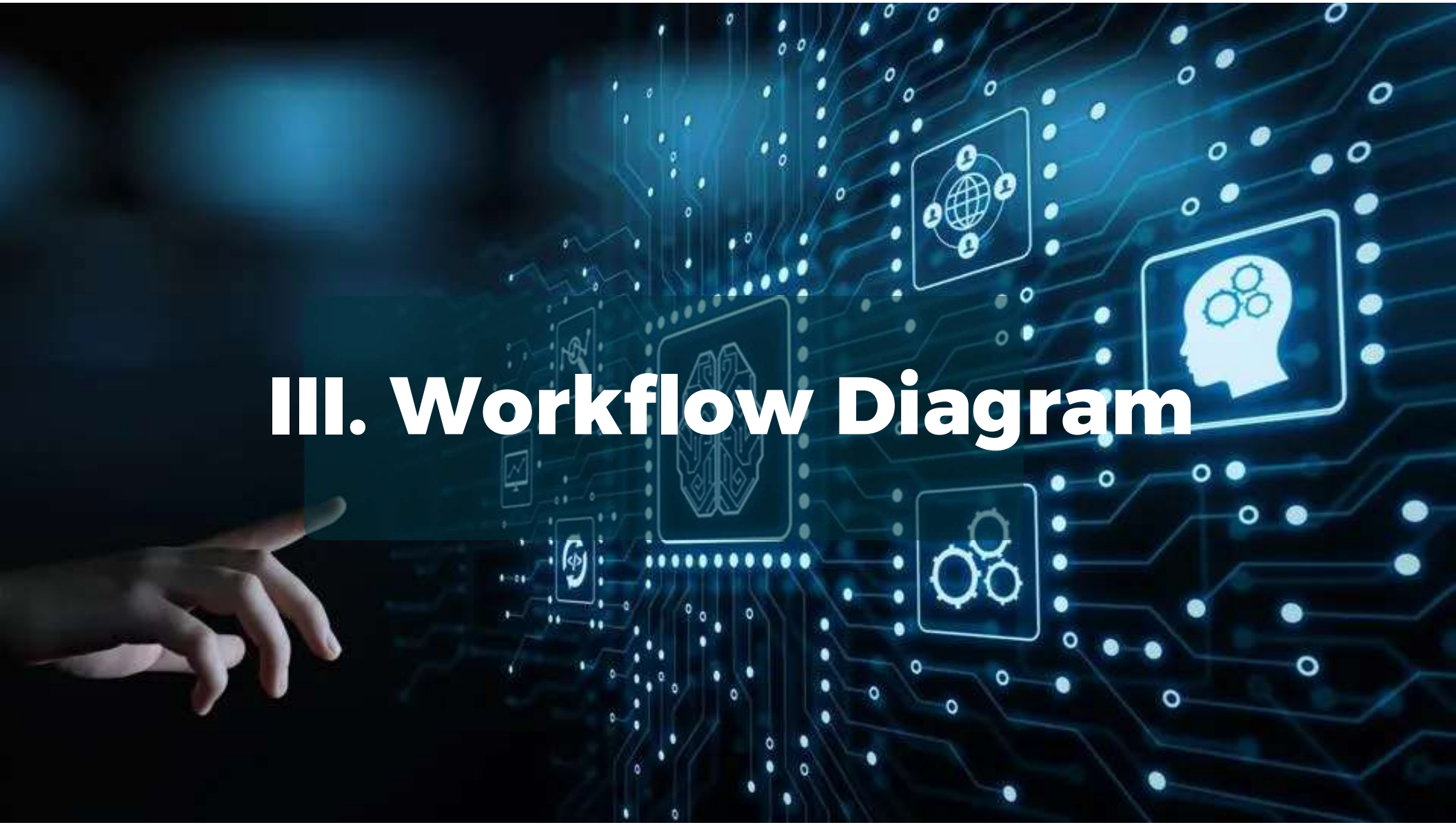
- 70% training
- 20% validation
- 10% test

## Data Augmentation

- Horizontal flip
- Rotation ±40°
- Width/Height shift ±10%
- Shear & Zoom 20%

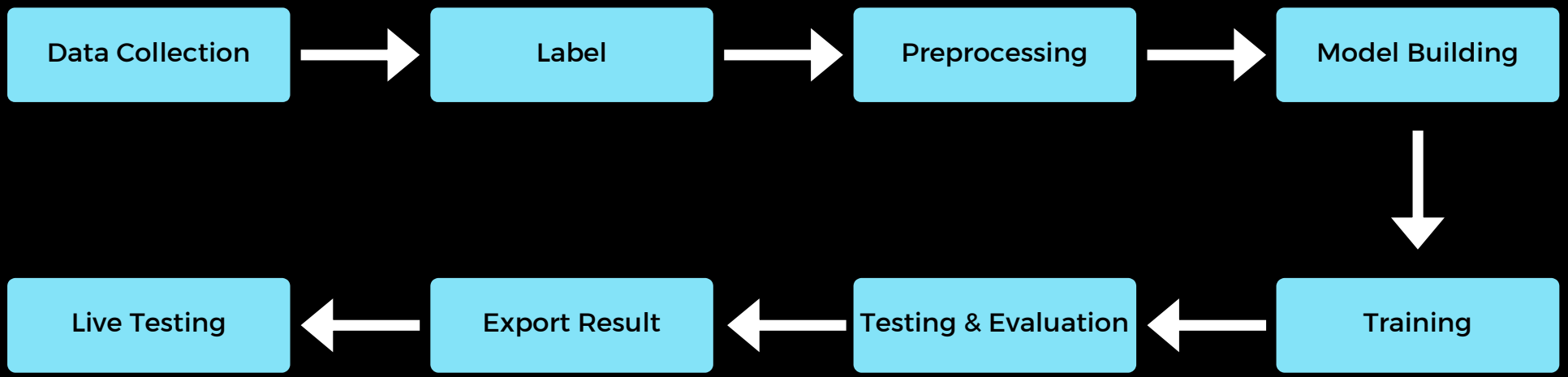
## Data Generator

- Create train/val/test generator
- Batch size = 32
- shuffle=False with test → keep original order



# III. Workflow Diagram

## Workflow Diagram Of Facial Image-Based Prediction System for Vietnamese Beauty Pageant Qualification




IV. Training Models

# Build the model

Build three model: MobileNetV2, ResNet50 and EfficientNetB3 from Keras and Sequential in Keras to easily combine the network layers in a linear order

```
model = Sequential()
model.add(basemodel)
model.add(Dense(1))
```

```
basemodel = MobileNetV2(include_top=False, pooling='avg', weights='imagenet')
```

```
1 basemodel = ResNet50(include_top=False, pooling='avg', weights='imagenet')
```

```
1 basemodel = EfficientNetB3(include_top=False, pooling='avg', weights='imagenet')
```

# Multi Stage Training

- Determine the training parameters
- Base model freezing
- Compile Model
- Setup Callbacks
- Train Model

**Stage 1: train the custom model, which is the only fully connected layer.**

```
Số GPU đang sử dụng: 1
Model: "Mobinet"
```

| Layer (type) | Output Shape | Param # |
|---|---|---|
| mobilenetv2_1.00_224 (Functional) | (None, 1280) | 2,257,984 |
| dense_3 (Dense) | (None, 1) | 1,281 |

```
Total params: 2,259,265 (8.62 MB)
Trainable params: 1,281 (5.00 KB)
Non-trainable params: 2,257,984 (8.61 MB)
```

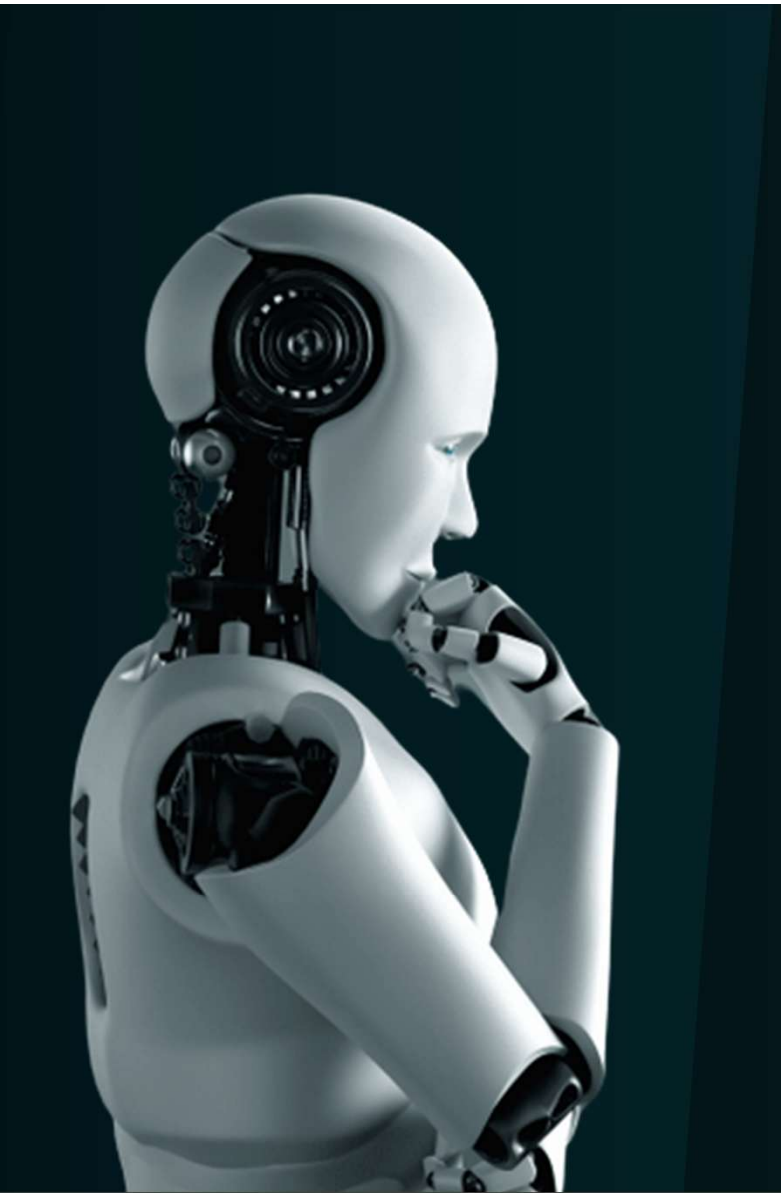
# Multi Stage Training

This phase is done similarly to stage 1 but the learning rate is reduced to 0.0001 and allows the entire model to be trained including the basemodel

**Stage 2: Retrain the entire model with a reduced learning rate.**

```
lr=0.0001
model.trainable = True
```


V. Testing & Evaluation



# Training result

## Library used

matplotlib.pyplot
used to visualize the loss graph.

## Method

Visualization of trend fluctuations over time:

- loss: MSE value on the training set after each epoch.
- val_loss: MSE value on the testing set after each epoch.

# Stage 1

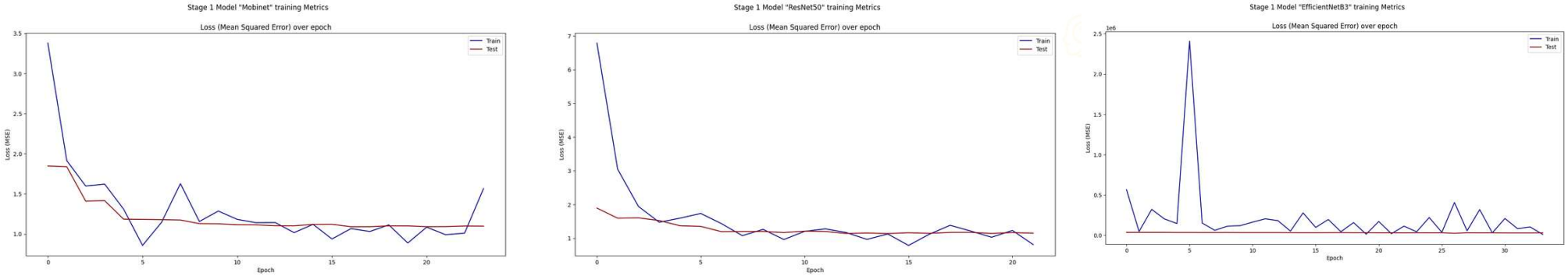



**Figure: Compare Stage training results 1**

**MobileNetV2** exhibited stable convergence and low overfitting, suitable for real-time use. However, **ResNet50**, although slightly overfitting in Stage 1, showed the best performance in the final evaluation. **EfficientNetB3's** unstable training behavior highlighted significant limitations in both convergence and reliability, which informed our model selection process for Stage 2.

# Stage 2

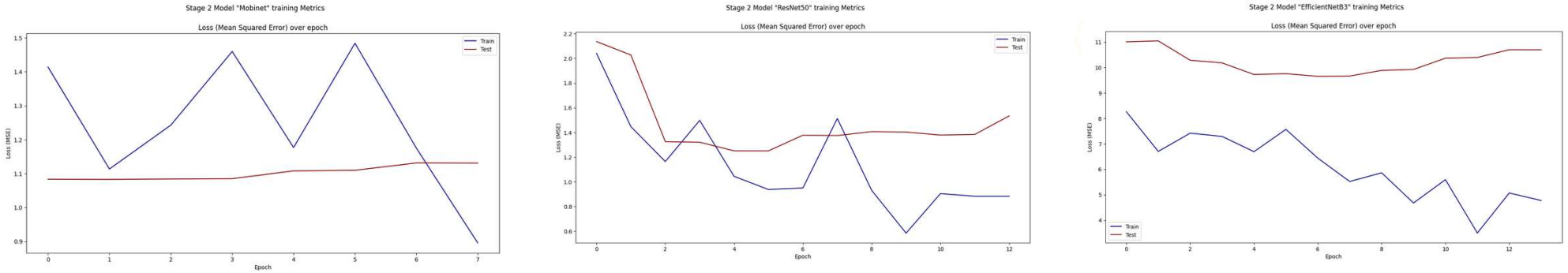



**Figure: Compare Stage training results 2**

**ResNet50** achieved the best prediction performance despite minor overfitting signs, as validated by test-set MAE and RMSE. **MobileNetV2** maintained stability and is ideal for fast deployment. In contrast, **EfficientNetB3** showed poor generalization, making it the least suitable model under the current setup.
=> Only **the head of the network** was trained, as full fine-tuning led to degraded performance on the validation set.

# performance evaluation

| Models | RMSE | MAE |
|---|---|---|
| MobileNetV2 | 0.8897 | 0.8396 |
| ResNet50 | 0.5285 | 0.5211 |
| EfficientNetB3 | 8.6934 | 2.6947 |

**Figure: performance evaluation**

**ResNet50** is the model that performs the best in this experiment for predicting face attractiveness based on both RMSE and MAE. Because of its consistent performance and lightweight design, **MobileNetV2** is still a formidable competitor. **EfficientNetB3** performs much worse, though, indicating that it would not be appropriate in the absence of bigger datasets or more fine-tuning. This highlights how crucial it is to select architectures for subjective prediction tasks like face beauty that are in line with task complexity and dataset scale.

# Conclusion

The study successfully developed a deep learning model to predict the facial attractiveness of Vietnamese beauty queens, with ResNet50 achieving the best performance, followed by MobileNetV 2. The system was implemented in real-time via webcam, showing strong application potential. However, limitations include a small dataset, subjective labels, and a lack of testing under diverse conditions. Future work should focus on expanding the dataset, incorporating contextual features, and ensuring fairness and transparency in AI deployment.