CS452 DEEP LEARNING
Assignment No.01

Name	**Vanya Shafiq**
Roll No	**22I-1953**
Section	**DS-A**

1. **Network Details and Baseline Rationale**
   - Architecture & Parameters
     - **Pre-trained Models**: MobileNetV2 and DenseNet121, both initialized with ImageNet weights.
     - **Base Layers**: Convolutional base frozen to preserve generic visual features.
     - **Parameters:** Only the added dense layers were trainable; all base parameters frozen.
     - **Top Layers Added**:
       - GlobalAveragePooling2D
       - Dense(256, ReLU) + Dropout(0.5)
       - Output layer:
         <u>Classification</u> → Dense(8, softmax), <u>Regression</u> → Dense(2, linear) for valence and arousal.
   - **Training Settings**
     - **Optimizer**: Adam (learning_rate = 1e-4).
     - **Loss Functions**: <u>Classification</u> → Sparse Categorical Crossentropy, <u>Regression</u> → Mean Squared Error (MSE)
     - **Metrics**: <u>Classification</u> → Accuracy, F1-score, Cohen's Kappa, AUC-ROC/PR, Krippendorff's Alpha, <u>Regression</u> → MAE, RMSE (Valence & Arousal), Pearson Correlation
     - **Training Data**: Train/validation/test split.
     - **Batch Size**: 16
     - **Epochs**: 10
   - Rationale for Choosing Baselines
     - **MobileNetV2**: Lightweight, efficient, and effective on small to medium-scale datasets, making it suitable as a strong, fast baseline.
     - **DenseNet121**: Deeper architecture with feature re-use through dense connectivity, offering stronger representational power at the cost of higher computation.
     - Using both allows comparison between an efficient lightweight baseline and a deeper, more expressive baseline.
   - Comparison of Baseline Performance
     - **MobileNetV2** achieved competitive results with faster training time.
     - **DenseNet121** provided higher capacity, which could lead to better accuracy/fit in some cases, but required longer training and more resources.
     - The comparison highlights the trade-off between **efficiency (MobileNetV2)** and **representational depth (DenseNet121)**.
2. **Comparison of Baseline performance**
   - MobileNetV2 and DenseNet121 were assessed as baseline models for classification and regression.
   - MobileNetV2 resulted in lower RMSE, higher correlation, and improved CCC, indicating overall better accuracy and consistency in predictions.

- DenseNet121 showed slightly better performance on SAGR for arousal but was inferior in other crucial metrics.
- MobileNetV2 emerged as the more effective baseline, offering a favorable equilibrium between accuracy, generalization, and computational efficiency.
- DenseNet121, in contrast, demanded more resources with limited improvements in performance.

3. Details of Transfer learning

- Pre-trained model used
  MobileNetV2 pre-trained on **ImageNet** (used for both classification and regression).
  DenseNet121 pre-trained on **ImageNet** (also used for both classification and regression).
- Layers Frozen vs Trainable
  In both classification and regression models, the **convolutional base (MobileNetV2 / DenseNet121)** was **frozen**, and only the newly added top layers were trainable. This is because the base layers already capture generic visual features (edges, textures, shapes) from ImageNet, while the new top layers adapt these features to the specific tasks. Freezing prevents overfitting, speeds up training, and ensures stable feature extraction, while the trainable top layers handle task-specific learning.
- Model Adaptation for the Task
  I. Classification Model:
    - Added GlobalAveragePooling2D on top of the base.
    - Added Dense(256, relu) + Dropout(0.5) for regularization.
    - Final output: Dense(8, softmax) for multi-class classification.

  II. Regression Model
    - Added GlobalAveragePooling2D.
    - Added Dense(256, relu) + Dropout(0.5).
    - Final output: Dense(2, linear) → predicting continuous values [valence, arousal].
- Training Specifics
    - **Optimizer**: Adam with learning rate 1e-4.
    - **Loss Functions**:
        - Classification → sparse_categorical_crossentropy
        - Regression → Mean Squared Error (MSE)
    - **Metrics**:
        - Classification → accuracy, F1-score, Cohen’s Kappa, AUC-ROC, AUC-PR, Krippendorff’s Alpha.
        - Regression → Mean Absolute Error (MAE), RMSE for valence & arousal, Pearson correlation.
    - **Training Parameters**:
        - Epochs: 10
        - Batch size: 16

4. Performance Measure & discussion

a. Root Mean Squared Error (RMSE)
   - Measures the average magnitude of prediction errors.
   - Lower values indicate closer predictions to the ground truth.
   - Useful for assessing overall prediction accuracy in a regression setting.

b. Correlation (CORR)
   - Captures the linear relationship between predicted and true values.
   - Higher correlation shows that the model tracks trends and relative changes well, even if absolute values differ.

c. Sign Agreement (SAGR)
   - Evaluates whether the model correctly predicts the *direction* (positive/negative) of valence or arousal compared to the ground truth.
   - Important in emotion recognition because misclassifying the “polarity” of emotion (e.g., positive vs. negative valence) can lead to incorrect system behavior.

d. Concordance Correlation Coefficient (CCC)
   - Combines correlation with mean squared difference, measuring both accuracy and consistency.
   - More robust than CORR alone because it penalizes bias in predictions (systematic over- or under-estimation).
   - Often considered the **gold standard metric** for continuous affect prediction.

Results

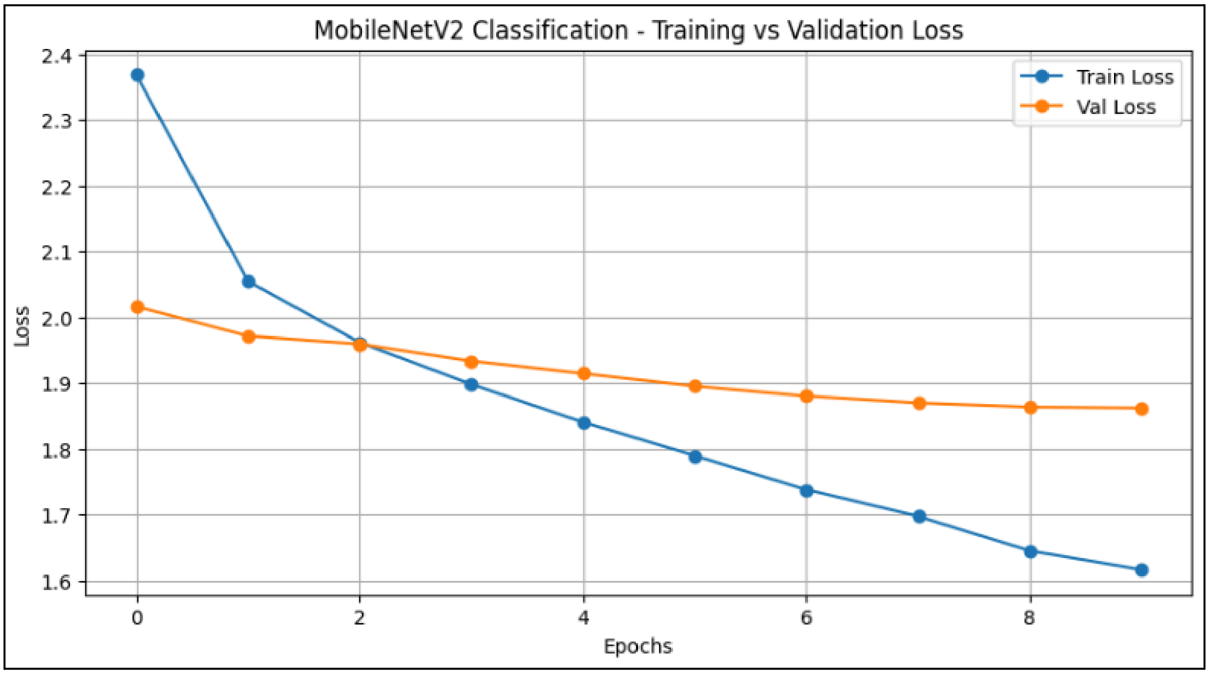
Model	RMSE (Valence)	RMSE (Arousal)	CORR (Valence)	CORR (Arousal)	SAGR (Valence)	SAGR (Arousal)	CCC (Valence)
MobileNetV2	**0.4435**	**0.3696**	**0.3061**	0.2048	**0.7183**	0.7650	**0.1876**
DenseNet121	0.4545	0.3751	0.2345	0.1845	0.6817	**0.7667**	0.1375

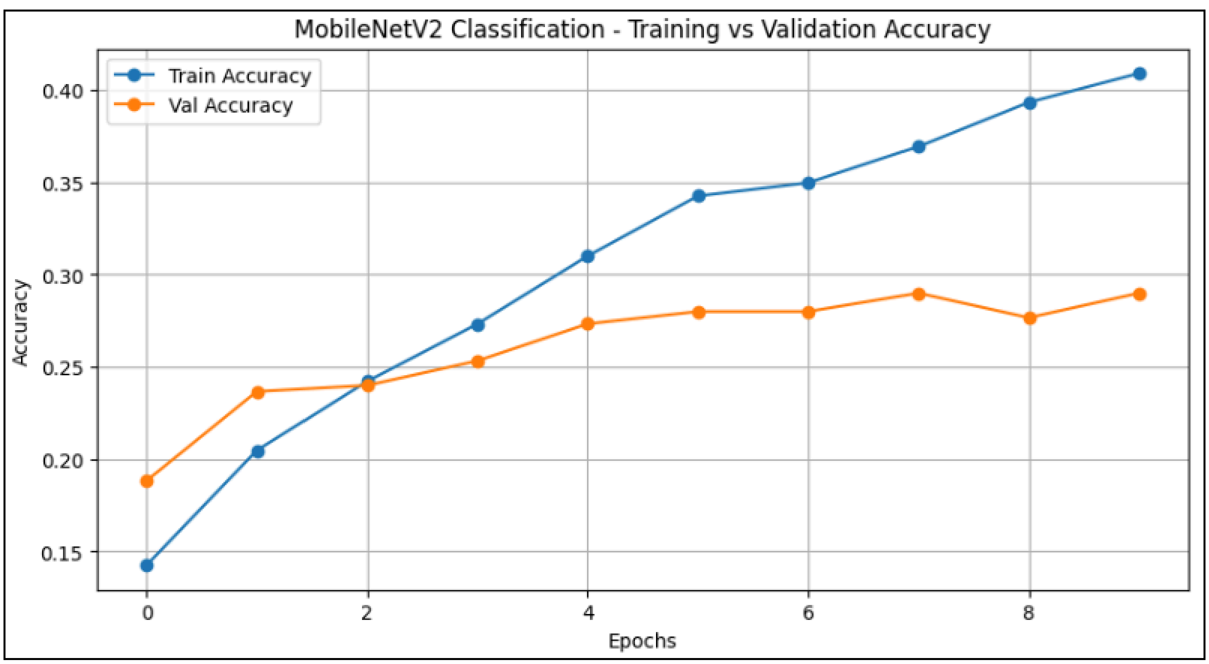
- Discussion
   - RMSE indicates MobileNet has lower prediction errors compared to DenseNet, aligning closer to the ground truth.
   - MobileNet exhibits a stronger correlation (CORR), suggesting superior trend-following abilities.
   - Both models effectively capture polarity, with MobileNet slightly better in valence and DenseNet slightly better in arousal.
   - CCC scores are low for both models, showing challenges in achieving accuracy and consistency.
   - CCC is deemed the most reliable metric for real-world deployment, as it ensures consistent and unbiased predictions alongside trend-following.

- A combination of CCC and SAGR is often preferred for applications focused on user-facing emotion awareness.
- Overall, MobileNetV2 is identified as the stronger choice for systems intended to operate effectively in real-world conditions.

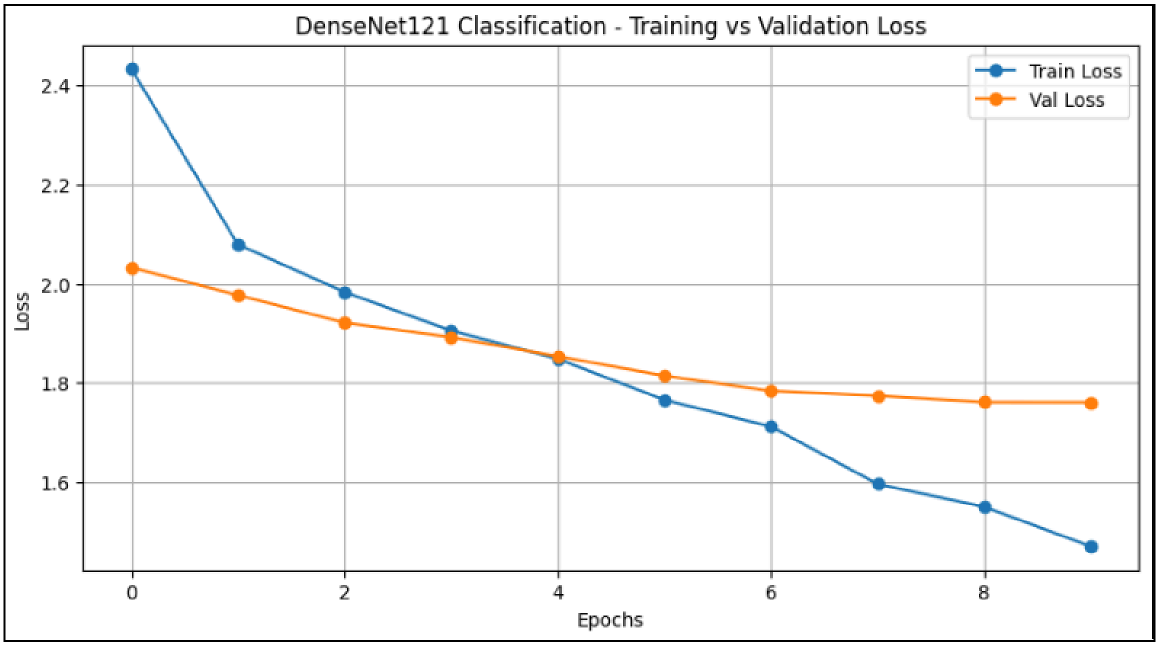
5. Training Graphs

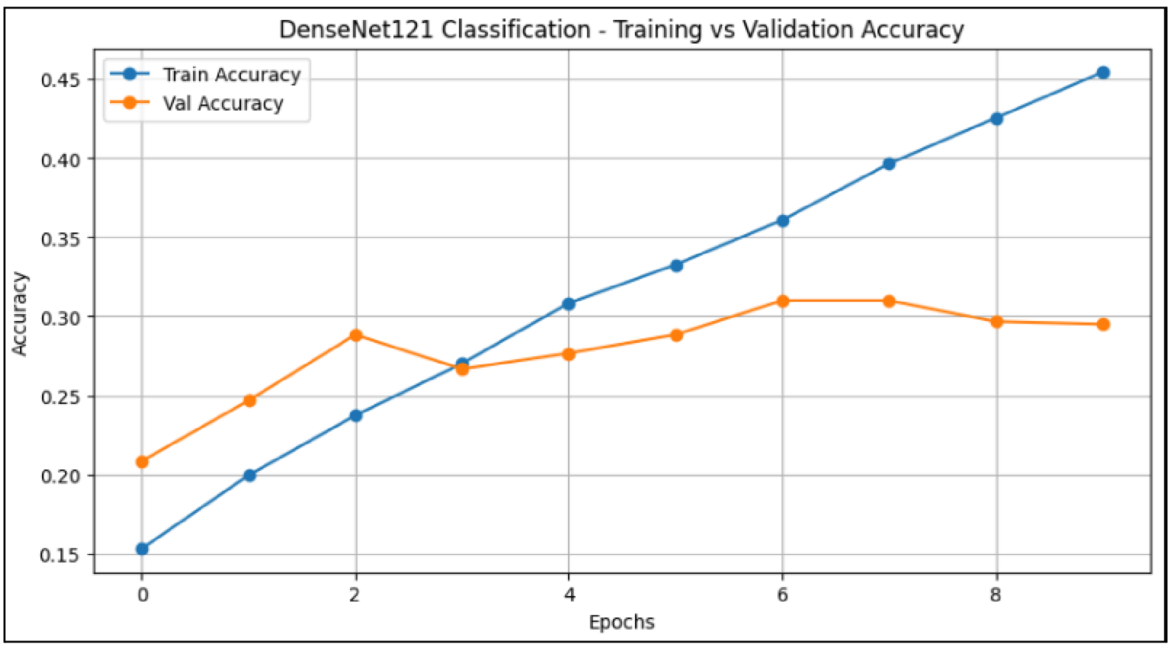
- MobileNet Model



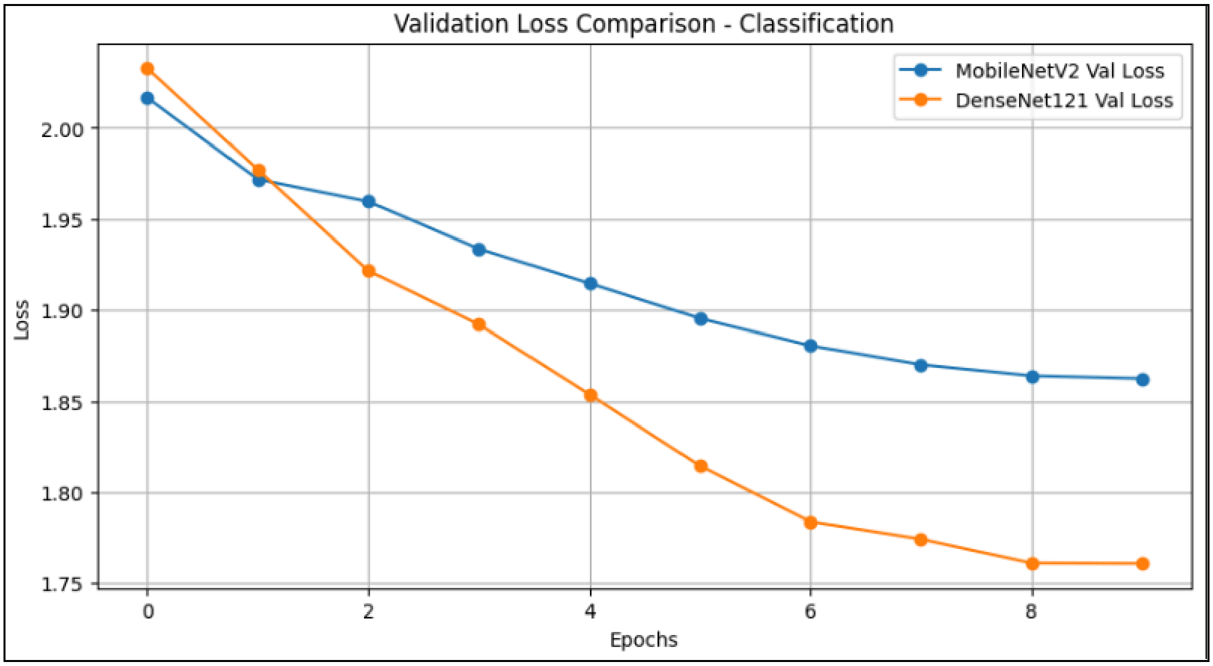


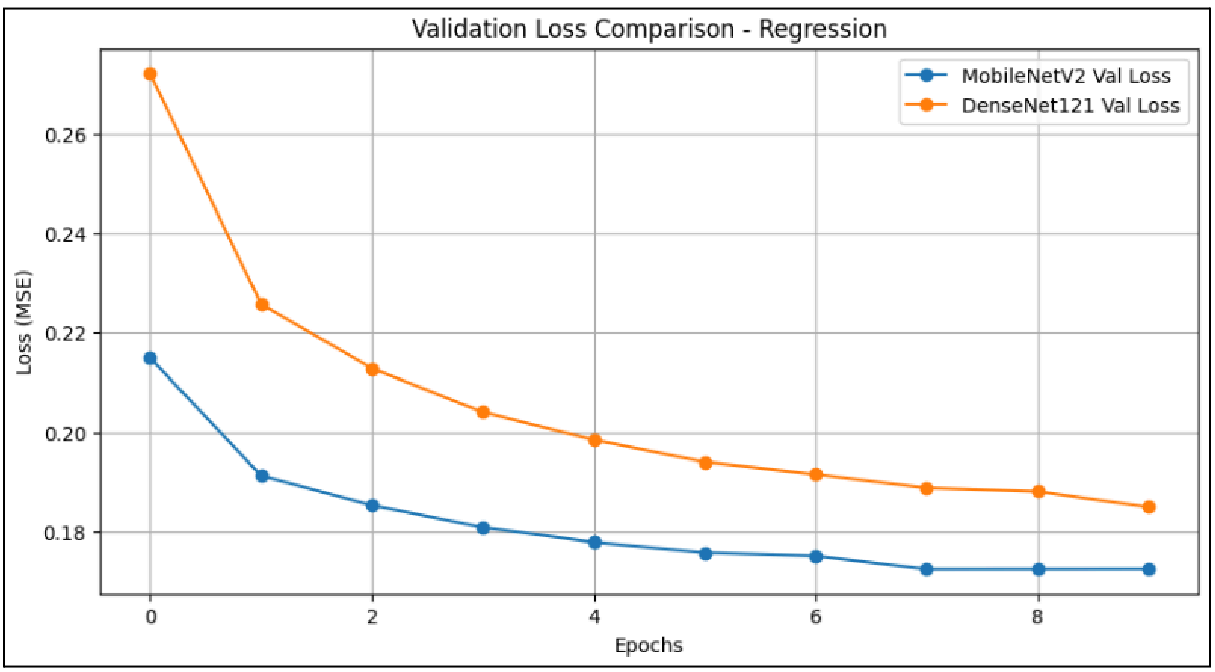
- DenseNet model





- Loss comparison btw models


Validation Loss Comparison - Classification
MobileNetV2 Val Loss
DenseNet121 Val Loss
Loss
2.00
1.95
1.90
1.85
1.80
1.75
0
2
4
6
8
Epochs


Validation Loss Comparison - Regression
MobileNetV2 Val Loss
DenseNet121 Val Loss
Loss (MSE)
0.26
0.24
0.22
0.20
0.18
0
2
4
6
8
Epochs

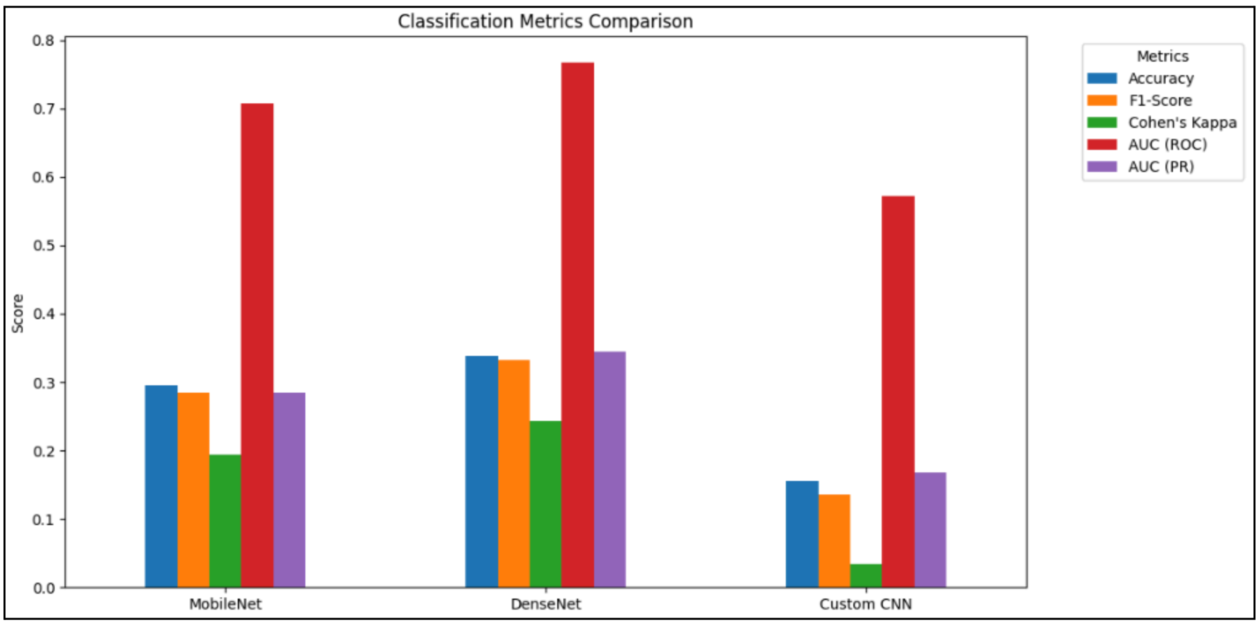
- Comparison between DenseNet, MobileNet & Custom CNN

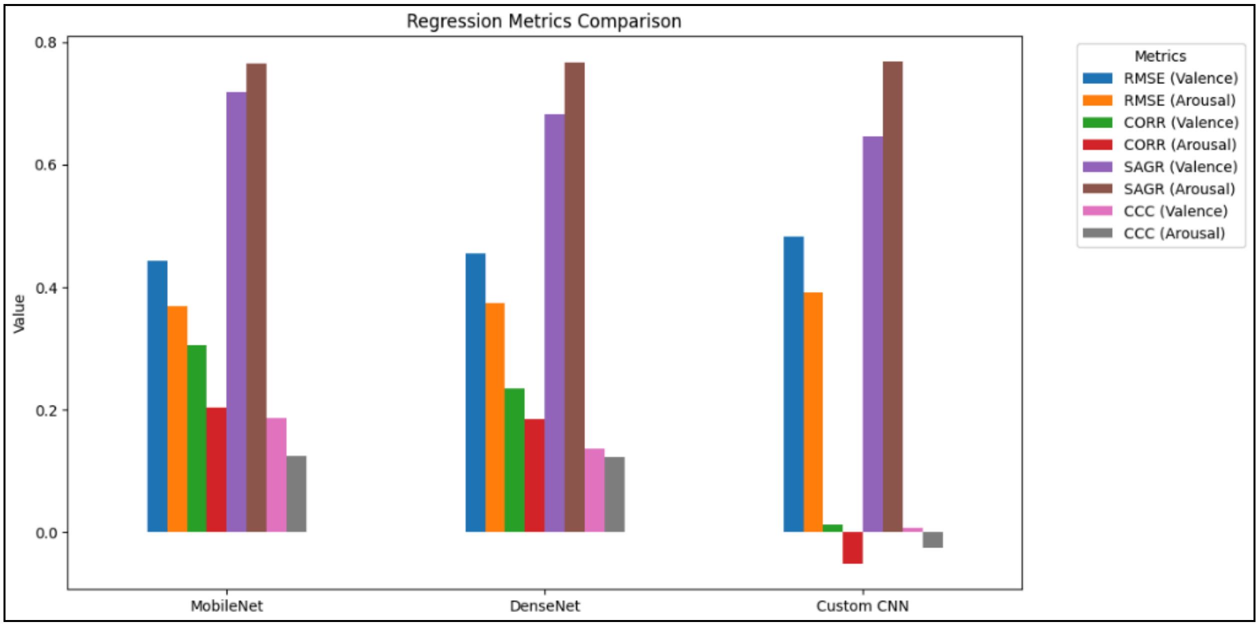
Classification Metrics

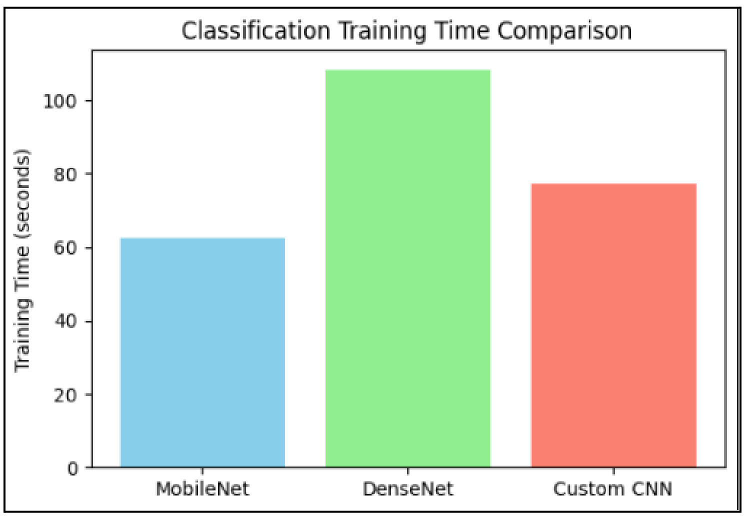
	Accuracy	F1-Score	Cohen's Kappa	AUC (ROC)	AUC (PR)
MobileNet	0.295000	0.284377	0.194286	0.707575	0.284489
DenseNet	0.338333	0.331868	0.243810	0.766829	0.345196
Custom CNN	0.155000	0.136317	0.034286	0.571657	0.168540

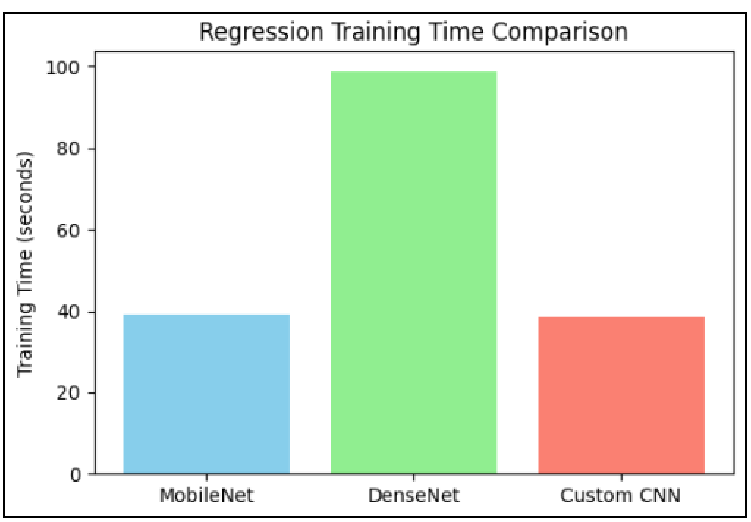
Regression Metrics

	RMSE (Valence)	RMSE (Arousal)	CORR (Valence)	CORR (Arousal)	SAGR (Valence)	SAGR (Arousal)	CCC (Valence)	CCC (Arousal)
MobileNet	0.443455	0.369550	0.306145	0.204772	0.718333	0.765000	0.187579	0.125273
DenseNet	0.454506	0.375126	0.234541	0.184539	0.681667	0.766667	0.137473	0.123704
Custom CNN	0.482136	0.392019	0.013585	-0.051618	0.646667	0.768333	0.007123	-0.025344



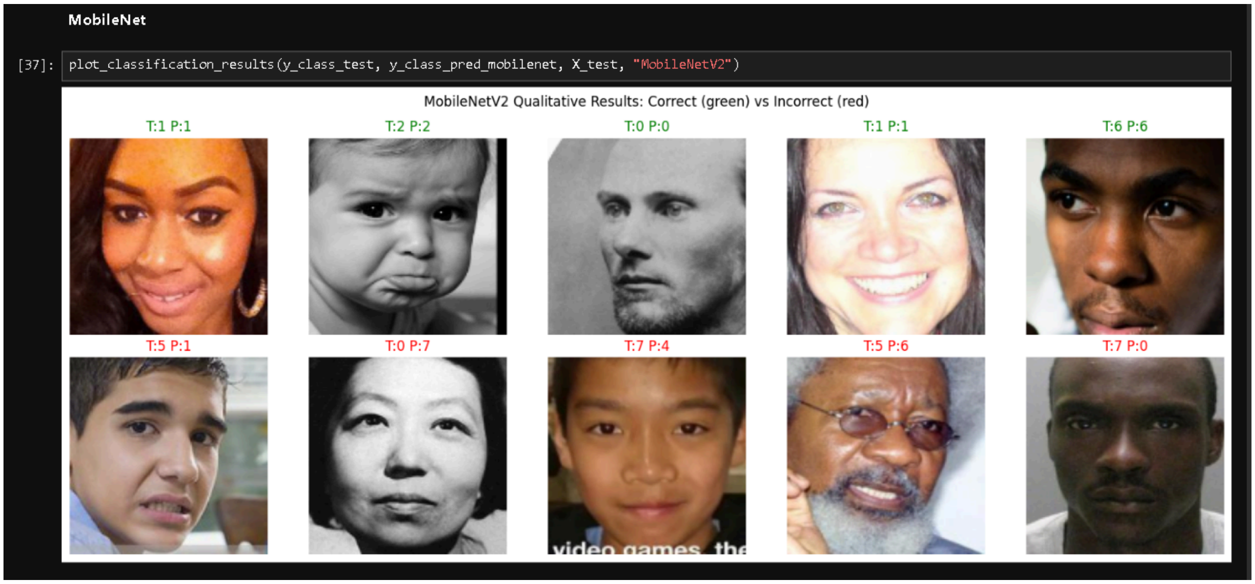

Regression Metrics Comparison
Value
0.8
0.6
0.4
0.2
0.0
MobileNet
DenseNet
Custom CNN
Metrics
RMSE (Valence)
RMSE (Arousal)
CORR (Valence)
CORR (Arousal)
SAGR (Valence)
SAGR (Arousal)
CCC (Valence)
CCC (Arousal)


Classification Training Time Comparison
Training Time (seconds)
100
80
60
40
20
0
MobileNet
DenseNet
Custom CNN


Regression Training Time Comparison
Training Time (seconds)
100
80
60
40
20
0
MobileNet
DenseNet
Custom CNN

6. Correctly and Incorrectly classified images

- MobileNet Model
  Green shows correct predictions whereas red shows incorrect.



- DenseNet Model
  Green shows correct predictions whereas red shows incorrect.