

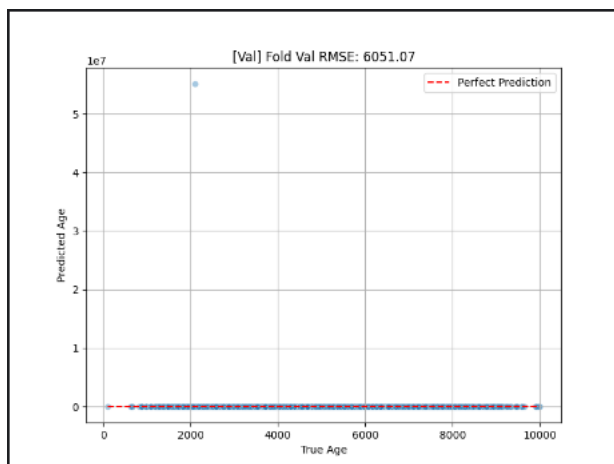
Model Performance Progression and Score Improvements

1. Initial Baseline (Custom CNN Model)

Our first attempt using a basic CNN architecture resulted in poor predictive performance:

- **RMSE ≈ 25**
- **$R^2 = -2.0221$** , indicating severe underfitting
- **Extremely high training and validation losses**

```
Found 8304 valid rows. Skipped 0 rows due to missing files.
2025-04-26 04:34:12,884 [INFO] Epoch [1/25] - Train Loss: 43860996.2571 - Val Loss: 827.5842 - Val RMSE: 28.7678 - Val  $R^2$ : -2.0221 - LR: 0.0001
2025-04-26 04:34:35,597 [INFO] Epoch [2/25] - Train Loss: 6139964.5040 - Val Loss: 626.0608 - Val RMSE: 25.0212 - Val  $R^2$ : -1.2862 - LR: 0.0001
2025-04-26 04:34:58,012 [INFO] Epoch [3/25] - Train Loss: 7343860.0280 - Val Loss: 649.6558 - Val RMSE: 25.4883 - Val  $R^2$ : -1.3724 - LR: 0.0001
```

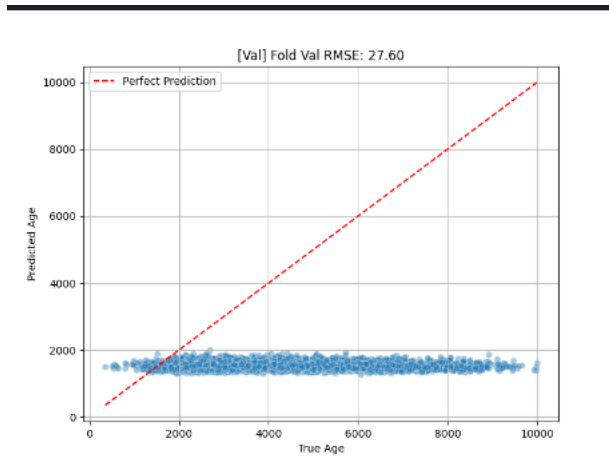


2. Gender Class Balancing

We applied partial oversampling to address the gender imbalance in the dataset.

- **RMSE improved slightly to ~ 17** , but predictions remained unstable
- **Training and validation losses were still very high**
- **$R^2 = -1.78$** , indicating continued underfitting despite balancing efforts

```
2025-04-26 04:26:03,829 [INFO] Epoch [1/25] - Train Loss: 92241360.1227 - Val Loss: 761.9436 - Val RMSE: 27.6033 - Val  $R^2$ : -1.7868 - LR: 0.0001
2025-04-26 04:26:30,601 [INFO] Epoch [2/25] - Train Loss: 10840135.0917 - Val Loss: 280.7375 - Val RMSE: 16.7552 - Val  $R^2$ : -0.0268 - LR: 0.0001
2025-04-26 04:26:57,015 [INFO] Epoch [3/25] - Train Loss: 60549044.9004 - Val Loss: 293.7312 - Val RMSE: 17.1386 - Val  $R^2$ : -0.0743 - LR: 0.0001
```

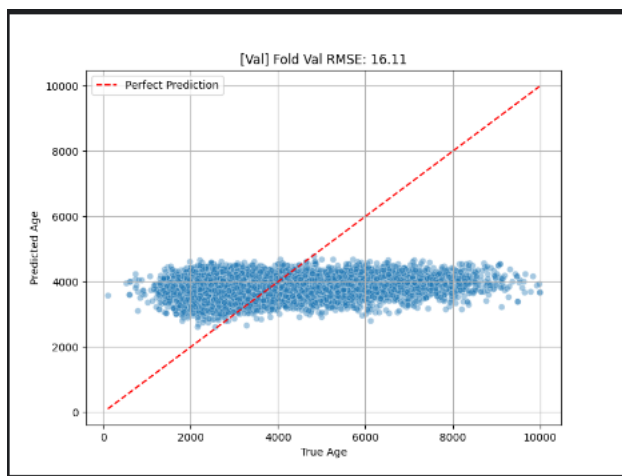


3. Data Augmentation Enhancements

We introduced several augmentations including grayscale-to-RGB conversion, image resizing, and random horizontal flipping.

- **RMSE stabilized around ~15**
- **Training and validation losses significantly decreased**
- **R^2 improved to 0.05**, showing early signs of meaningful learning

```
2025-04-26 04:22:07.089 [INFO] Epoch [1/25] - Train Loss: 313.2420 - Val Loss: 259.4757 - Val RMSE: 16.1082 - Val  $R^2$ : 0.0505 - LR: 0.00010
2025-04-26 04:22:37.618 [INFO] Epoch [2/25] - Train Loss: 251.7613 - Val Loss: 252.9206 - Val RMSE: 15.9035 - Val  $R^2$ : 0.0745 - LR: 0.00010
2025-04-26 04:23:08.369 [INFO] Epoch [3/25] - Train Loss: 244.2716 - Val Loss: 245.3958 - Val RMSE: 15.6651 - Val  $R^2$ : 0.1020 - LR: 0.00010
```

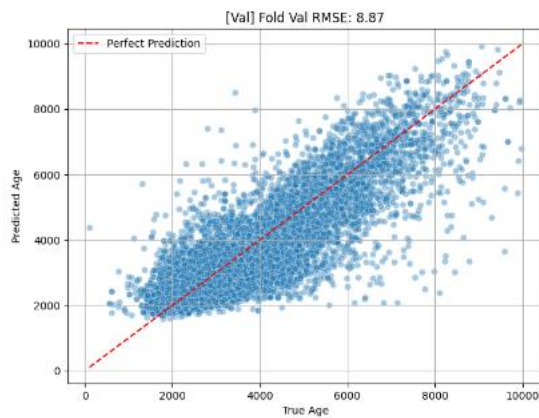


4. ResNet-50 Integration

We experimented with ResNet-18 and ResNet-34 before finalizing ResNet-50 for its deeper architecture and superior performance:

- **RMSE dropped to 9.83**, a significant improvement
- **R^2 increased to 0.60**, indicating better model fit
- **Training and validation losses consistently decreased with each epoch**

```
2025-04-26 03:25:58,849 [INFO] Epoch [1/25] - Train Loss: 200.0980 - Val Loss: 96.7354 - Val RMSE: 9.8354 - Val R2: 0.6385 - LR: 0.00010
2025-04-26 03:27:45,983 [INFO] Epoch [2/25] - Train Loss: 93.8012 - Val Loss: 89.0905 - Val RMSE: 9.4388 - Val R2: 0.6671 - LR: 0.00010
2025-04-26 03:29:33,033 [INFO] Epoch [3/25] - Train Loss: 82.4197 - Val Loss: 78.8905 - Val RMSE: 8.8820 - Val R2: 0.7052 - LR: 0.00010
```



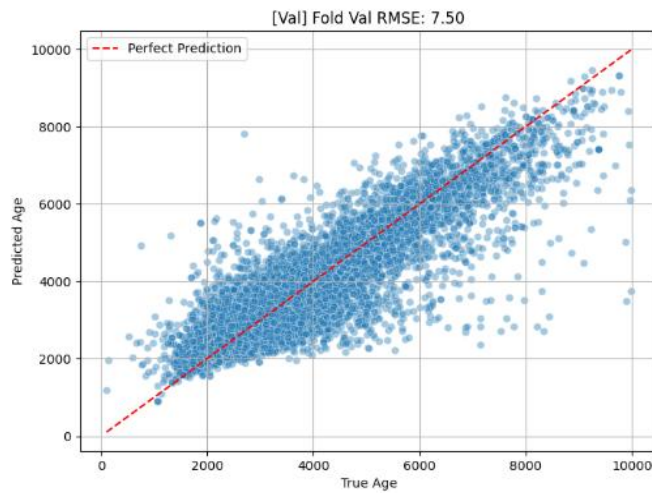
5. Final Model Optimization

We enhanced the model further by adding **dropout** and **batch normalization**, and switched from **MSE (L2 loss)** to **Smooth L1 Loss** for better robustness.

Additionally, we performed **hyperparameter tuning** to fine-tune model performance.

- **Best RMSE achieved: ~7.5**
- **$R^2 = 0.79$** , indicating strong predictive accuracy
- **Training loss ≈ 3 , Validation loss ≈ 4**

```
2025-04-26 05:42:37,484 [INFO] Epoch [23/25] - Train Loss: 3.0899 - Val Loss: 4.7210 - Val RMSE: 7.5741 - Val R2: 0.7883 - LR: 0.00005
2025-04-26 05:44:24,131 [INFO] Epoch [24/25] - Train Loss: 3.0000 - Val Loss: 4.7226 - Val RMSE: 7.6510 - Val R2: 0.7839 - LR: 0.00005
2025-04-26 05:46:10,916 [INFO] Epoch [25/25] - Train Loss: 2.9245 - Val Loss: 4.6212 - Val RMSE: 7.4962 - Val R2: 0.7926 - LR: 0.00005
2025-04-26 05:46:11,486 [INFO] Best model (Val RMSE=7.4962) saved to /home/ubuntu/ML_Project/output/checkpoints/fold_0.pth
```



Final Averaged Results-

```
✅ Best Hyperparameter Config:
{'lr': 0.0001, 'batch_size': 32, 'dropout': 0.5, 'epochs': 25, 'early_stop_patience': 2}
Best Avg RMSE: 7.82531228972835
```

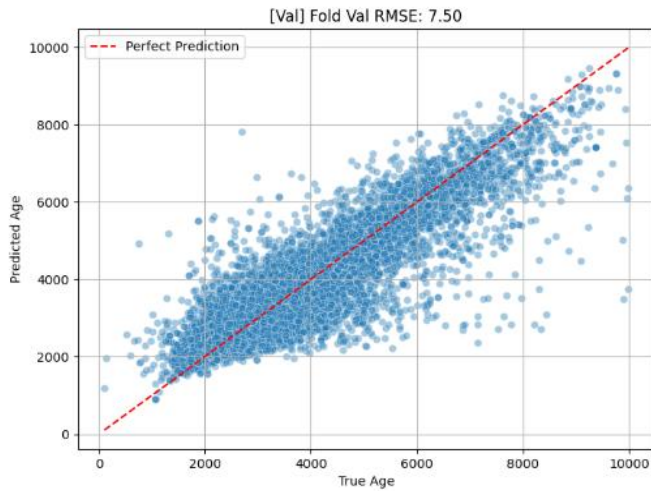
Impact of Cross-Validation on Model Performance

With Cross-Validation:

We implemented 5-fold cross-validation to ensure consistent evaluation across different data splits.

- **RMSE stabilized around ~7.49**
- **R^2 improved to ~0.79**, indicating strong model generalization
- **Training loss \approx 2.92–3.09, Validation loss \approx 4.62–4.72**, showing convergence

```
2025-04-26 05:42:37,484 [INFO] Epoch [23/25] - Train Loss: 3.0899 - Val Loss: 4.7210 - Val RMSE: 7.5741 - Val R2: 0.7883 - LR: 0.00005
2025-04-26 05:44:24,131 [INFO] Epoch [24/25] - Train Loss: 3.0000 - Val Loss: 4.7226 - Val RMSE: 7.6510 - Val R2: 0.7839 - LR: 0.00005
2025-04-26 05:46:10,916 [INFO] Epoch [25/25] - Train Loss: 2.9245 - Val Loss: 4.6212 - Val RMSE: 7.4962 - Val R2: 0.7926 - LR: 0.00005
2025-04-26 05:46:11,486 [INFO] Best model (Val RMSE=7.4962) saved to /home/ubuntu/ML_Project/output/checkpoints/fold_0.pth
```



Without Cross-Validation:

A simple 80-20 train-validation split was used.

- **RMSE increased**, indicating less stable performance
- **R^2 decreased**, suggesting weaker model fit
- **Training and validation losses were slightly higher**

```
2025-04-26 09:00:33,891 [INFO] Epoch [24/25] - Train Loss: 9.6636 - Val Loss: 9.1604 - Val RMSE: 13.1023 - Val R2: 0.3687 - LR: 0.00050
2025-04-26 09:02:50,901 [INFO] Epoch [25/25] - Train Loss: 9.5391 - Val Loss: 9.3485 - Val RMSE: 13.3565 - Val R2: 0.3440 - LR: 0.00050
2025-04-26 09:02:51,268 [INFO] Best model (Val RMSE=13.0687) saved to /home/ubuntu/ML_Project/output/checkpoints/resnet_final_model.pth
✅ Final model RMSE: 13.06869873402812
```

