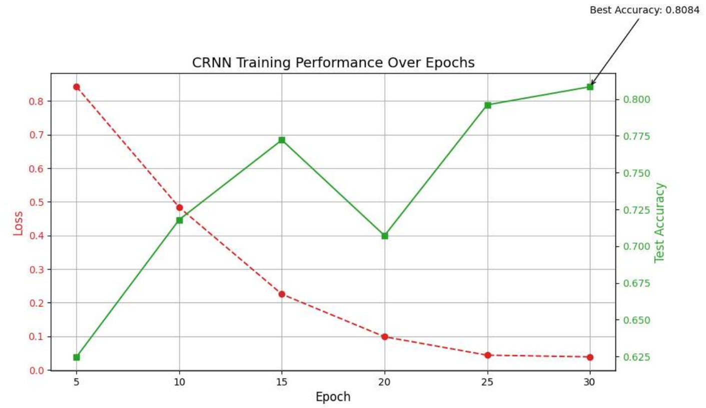
Model Evaluation

Performance Trends

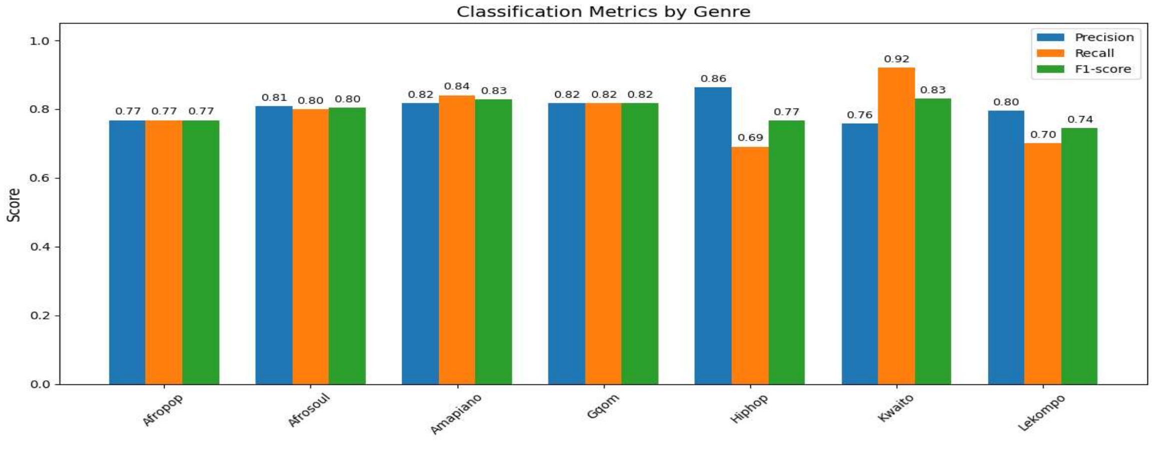
As shown in the evaluation graph below, training loss decreased steadily from **0.84** at epoch 5 to **0.038** by epoch 30, indicating successful convergence. Simultaneously, test accuracy improved from **62%** to **81%**, with slight fluctuations around epoch 20 suggesting temporary overfitting, which was corrected in later stages.



Classification Metrics by Genre

Per-genre Precision, Recall, and F1-scores revealed further insights:

- *Kwaito* achieved the highest Recall (0.92), indicating the model frequently recognized its distinct traits.
- *Hiphop* showed high Precision (0.86) but lower Recall, suggesting fewer false positives but potential underrepresentation.
- Overall, genres like *Amapiano* and *Gqom* demonstrated balanced scores, validating the feature extraction and class balancing strategies.


Classification Metrics by Genre
Precision
Recall
F1-score
Score
Afropop
Afrosoul
Amapiano
Gqom
Hiphop
Kwaito
Lekompo