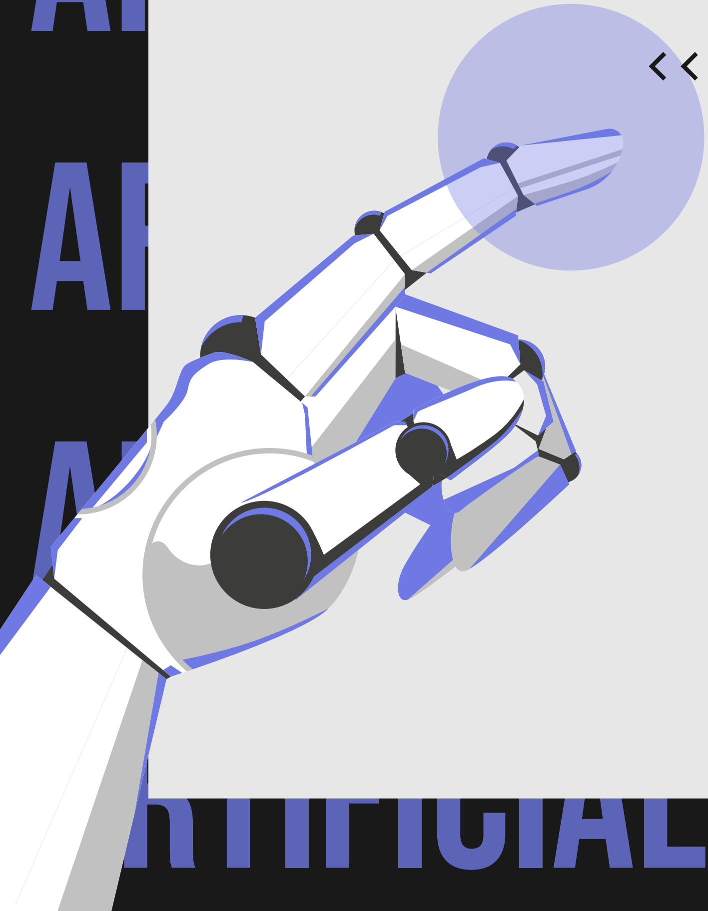


CHESS RATING PREDICTION

Final Milestone

Abdelrahman Elazab - Michael Reda

PROBLEM STATEMENT

ART
INT
(AI)

Traditional rating systems like Elo or Glicko rely on match outcomes, not the quality or timing of moves.

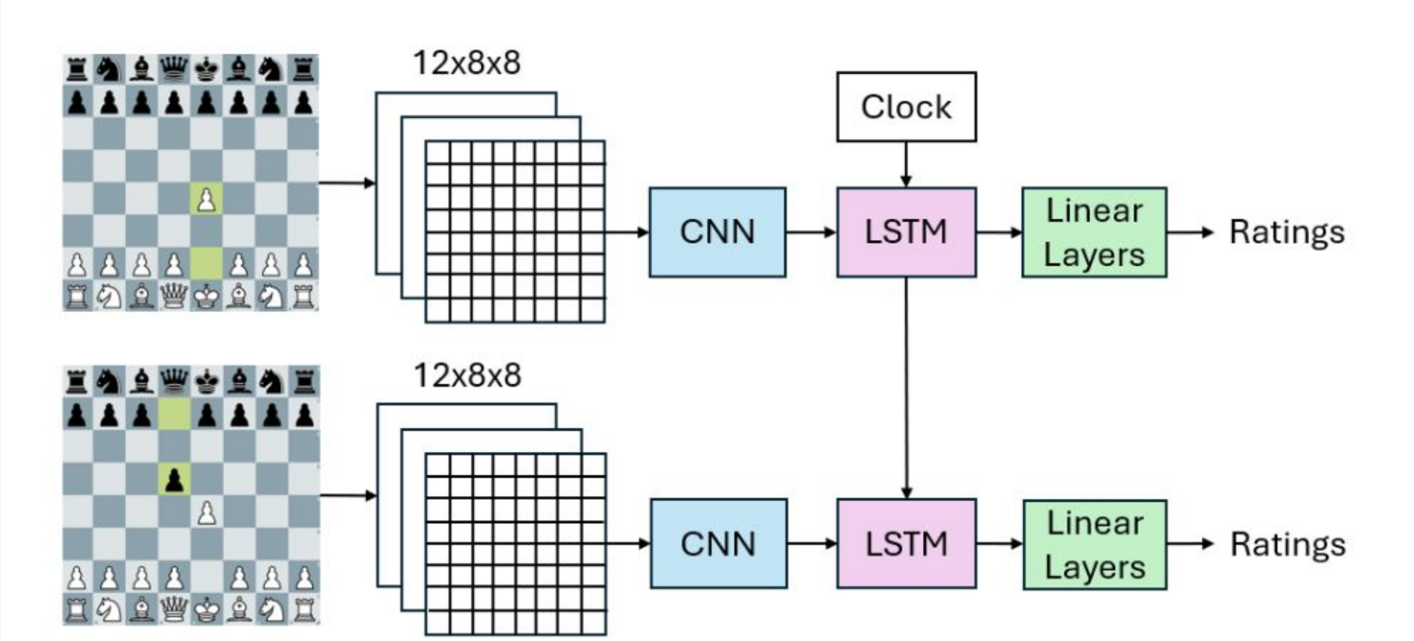
Our goal is to build a neural network that infers a player’s skill directly from gameplay behavior.

ORIGINAL MODEL

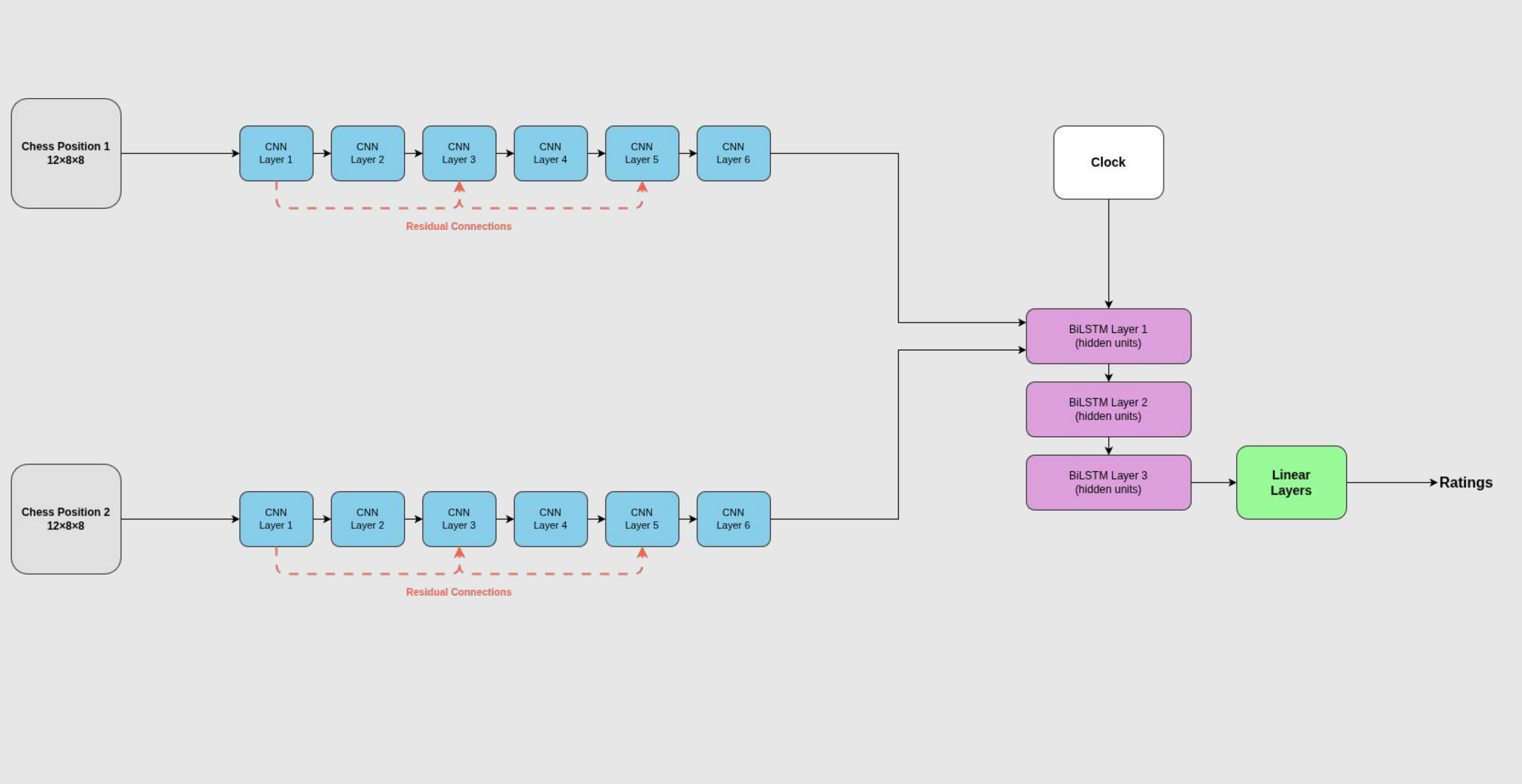
RatingNet

- Combines a 4 layer **CNN** for chess piece position features
- **Bidirectional LSTM (RNN)** that captures sequence dynamics of the moves
- Inputs: 8×8×12 board + per-move clock time.
- Achieved **MAE = 182 Elo**, with major improvement when time usage is included.
- Clock information reduced prediction error by 57 Elo on average.

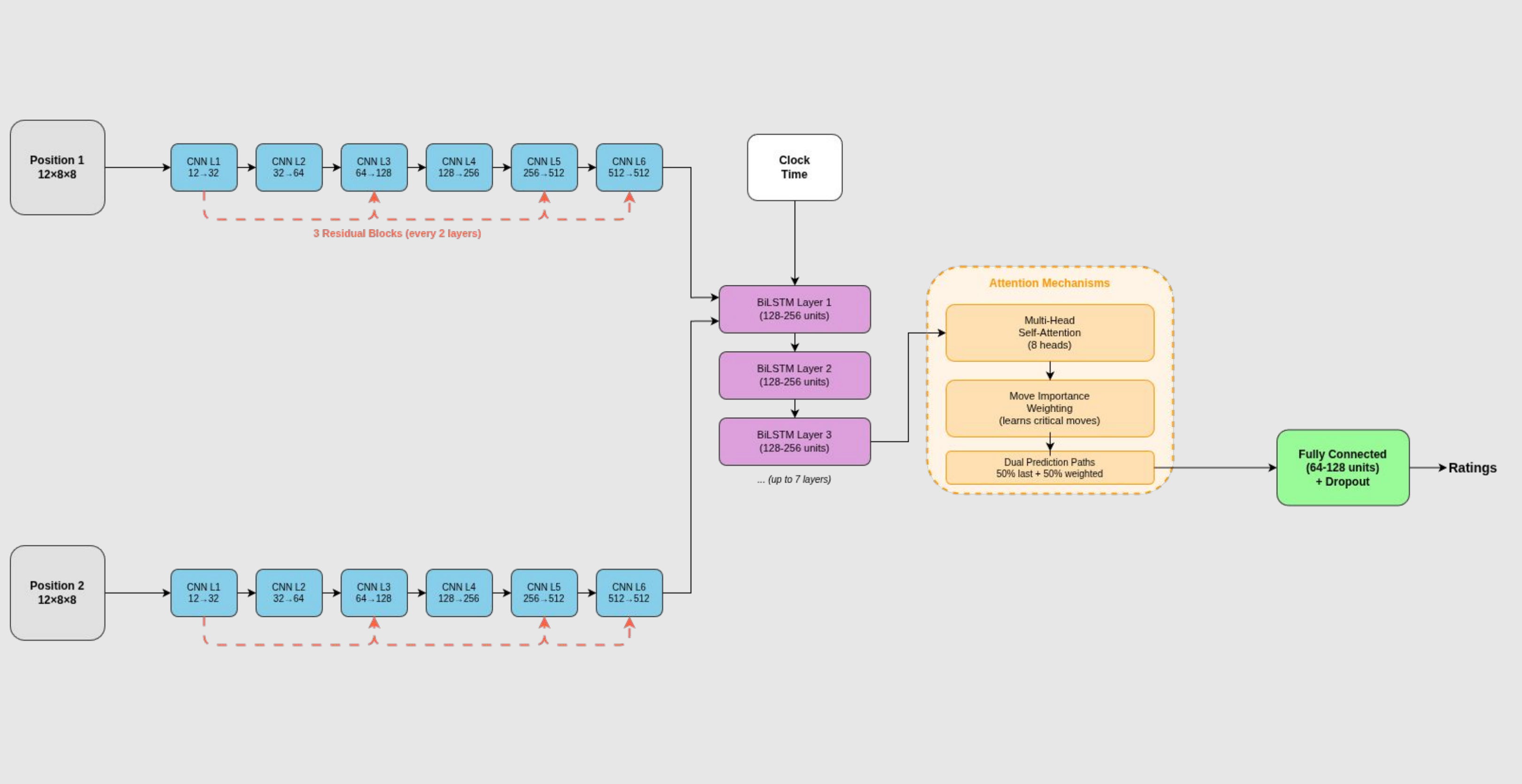
ORIGINAL MODEL



PROPOSED ARCHITECTURE



FINAL ARCHITECTURE

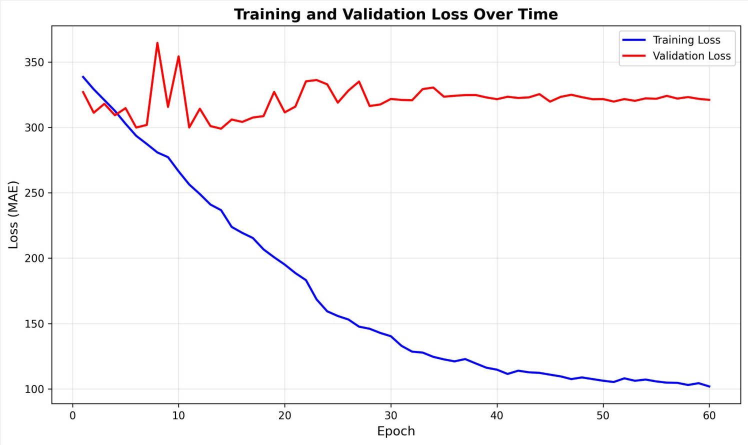


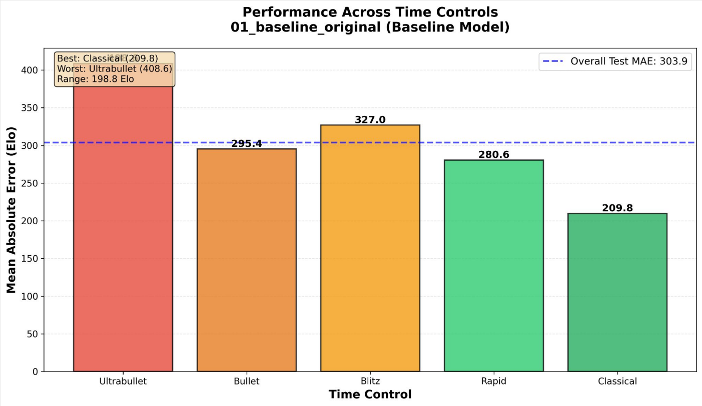
TRAINING ENHANCEMENT

- Data augmentation (horizontal flip, color swap with vertical flip)
- Cosine annealing scheduler to adjust the learning rate
- Mixed precision training
- Gradient clipping to prevent gradient explosion
- **Enhanced the early stopping techniques.**

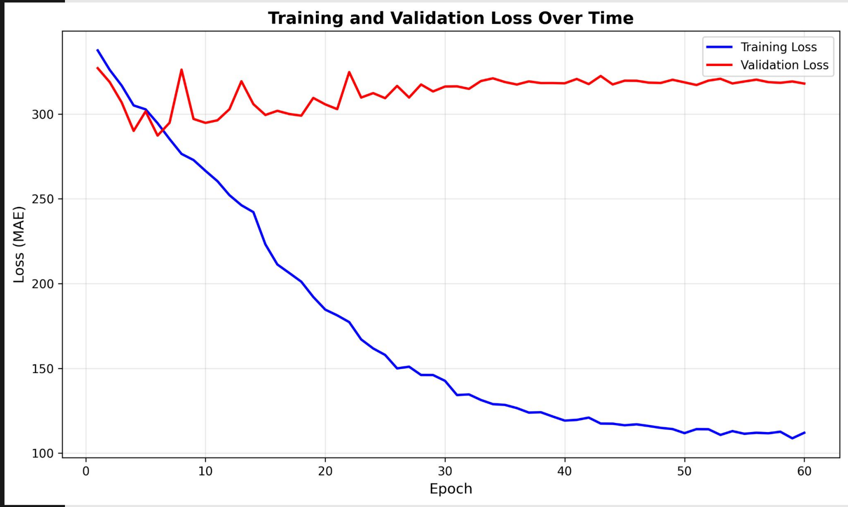
RESULTS

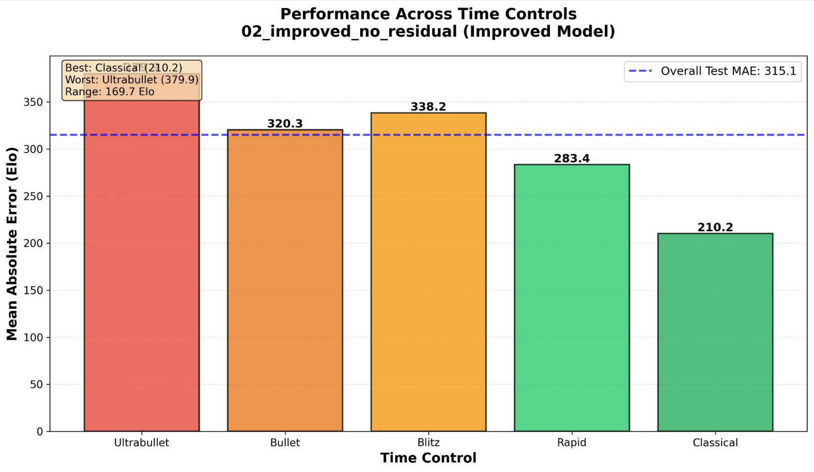
INITIAL RESULTS-BASELINE



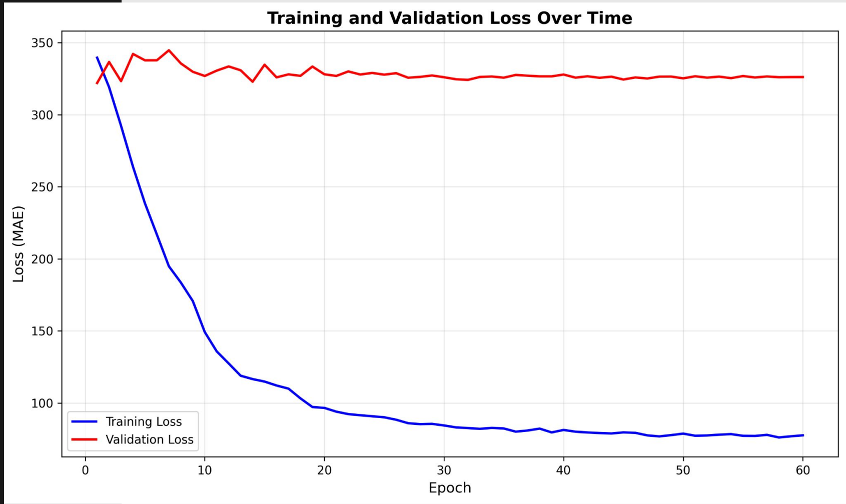


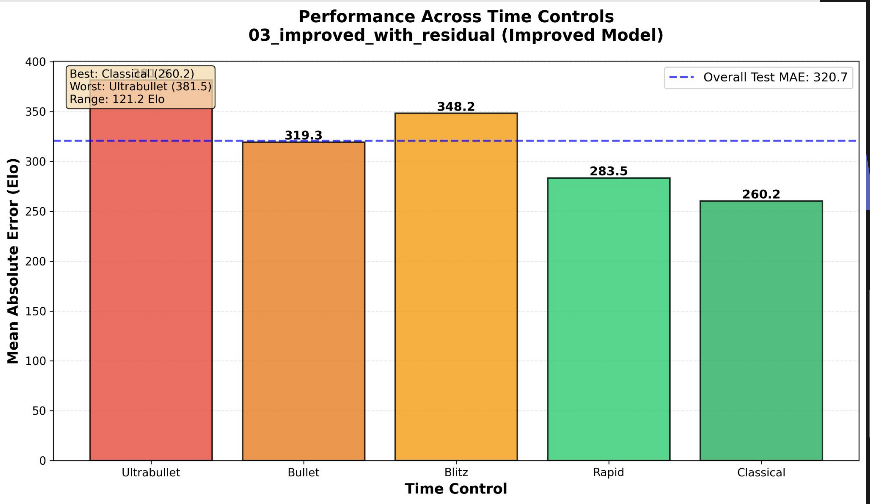
BASELINE WITH 6 CONVOLUTIONAL LAYERS



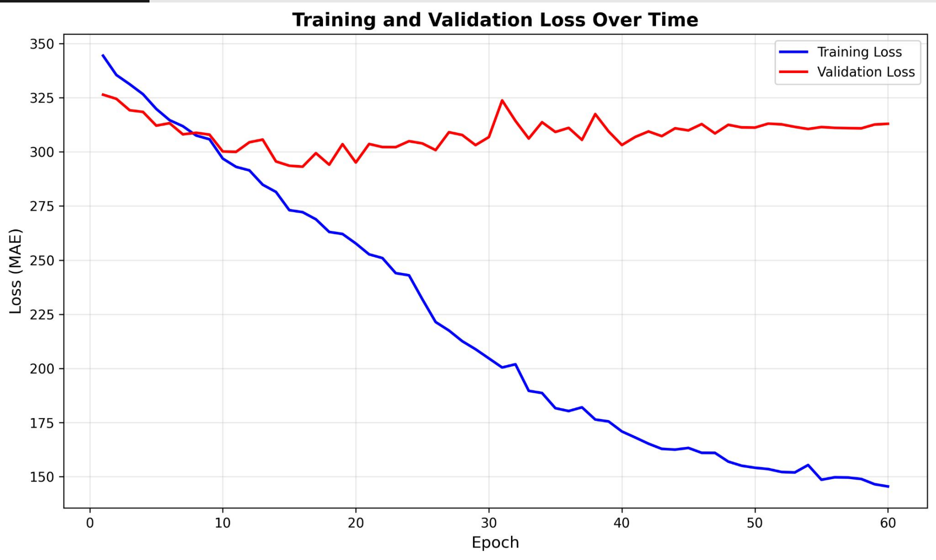


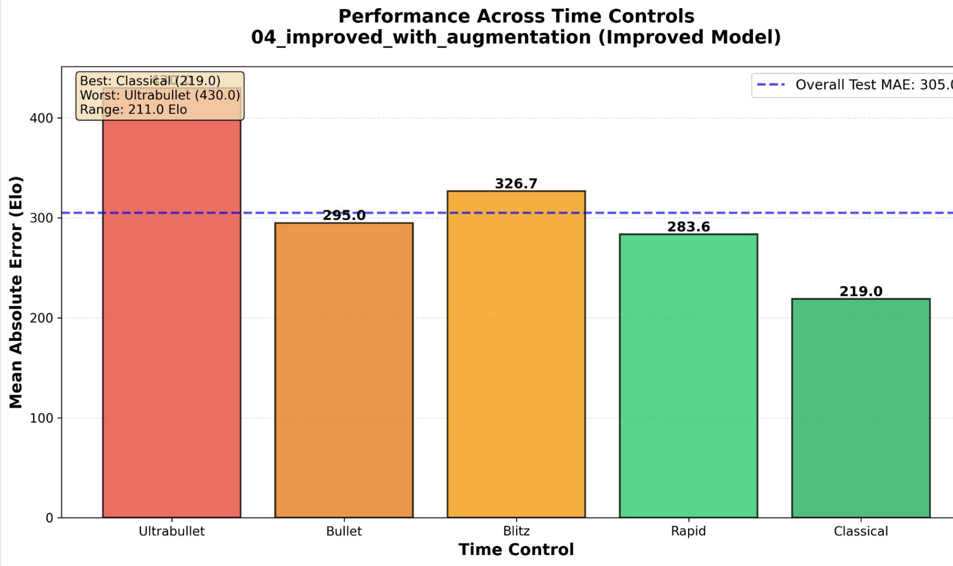
RESNET-LSTM: CNN WITH RESIDUAL CONNECTIONS



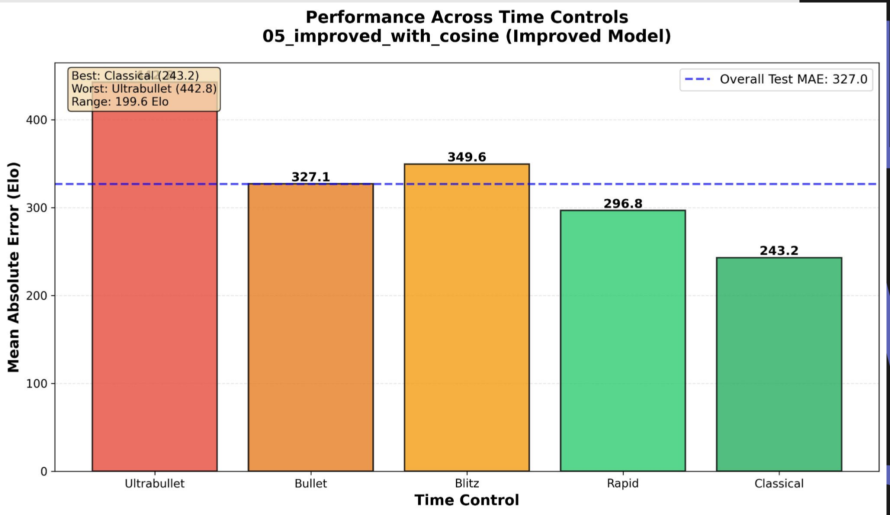


RESNET-LSTM WITH BOARD AUGMENTATION

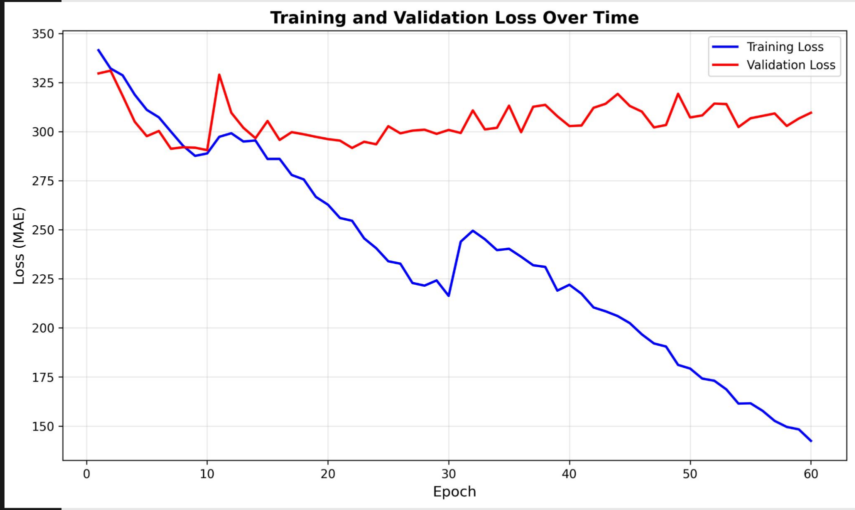


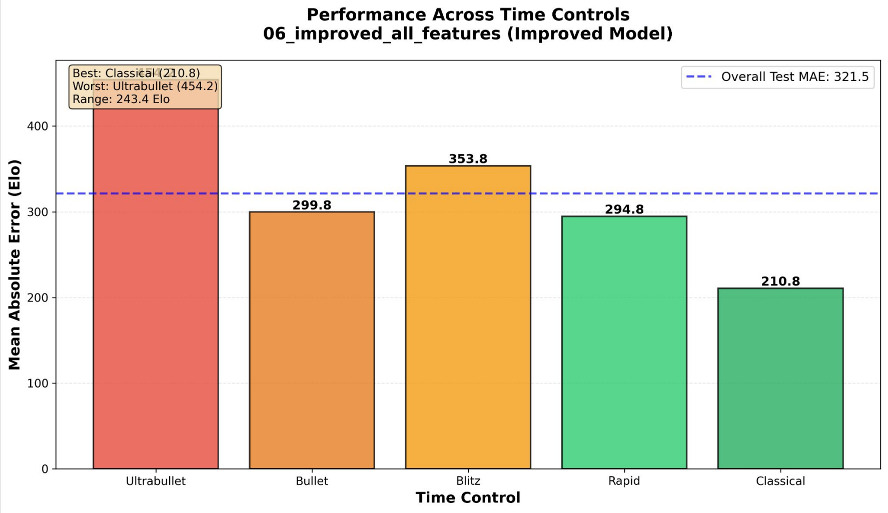


RESNET-LSTM WITH COSINE ANNEALING SCHEDULER

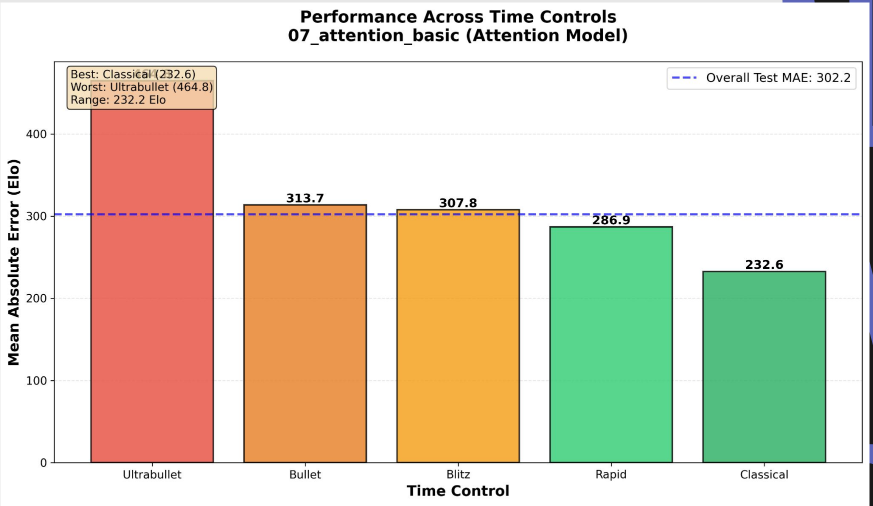


COMBINED ARCHITECTURE AND TRAINING ENHANCEMENTS

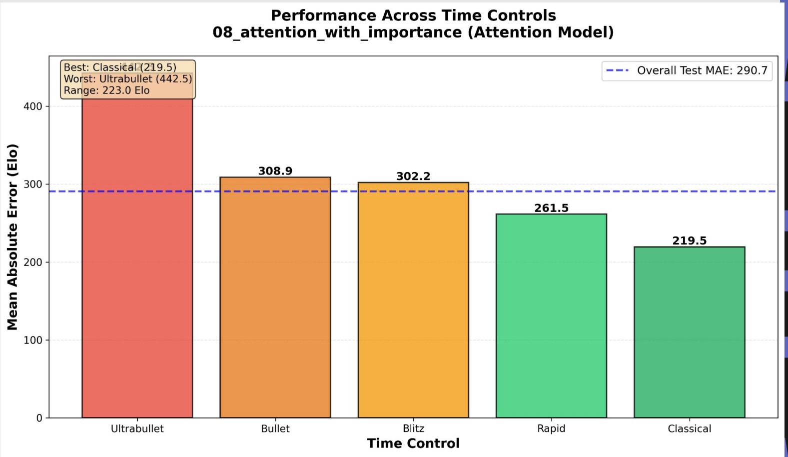




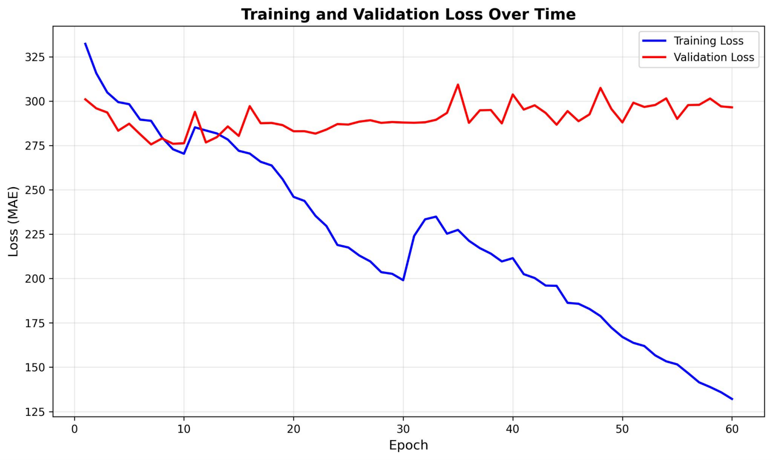
MULTI-HEAD SELF-ATTENTION NETWORK

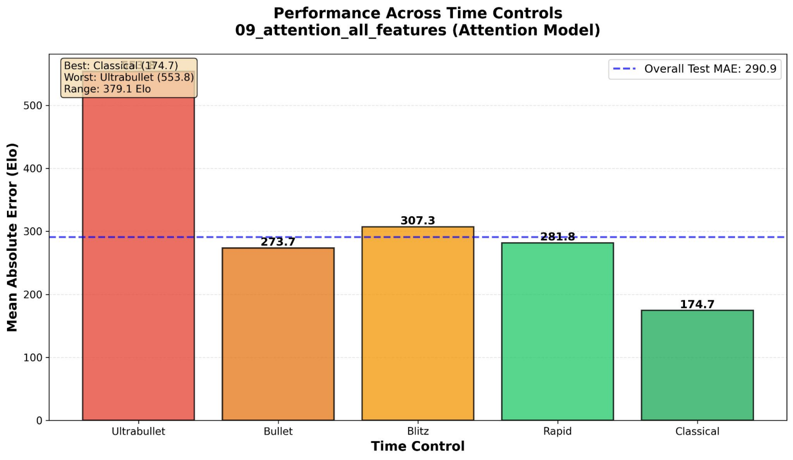


ATTENTION NETWORK WITH MOVE IMPORTANCE WEIGHTING



FULL ATTENTION ARCHITECTURE WITH ALL TRAINING ENHANCEMENTS

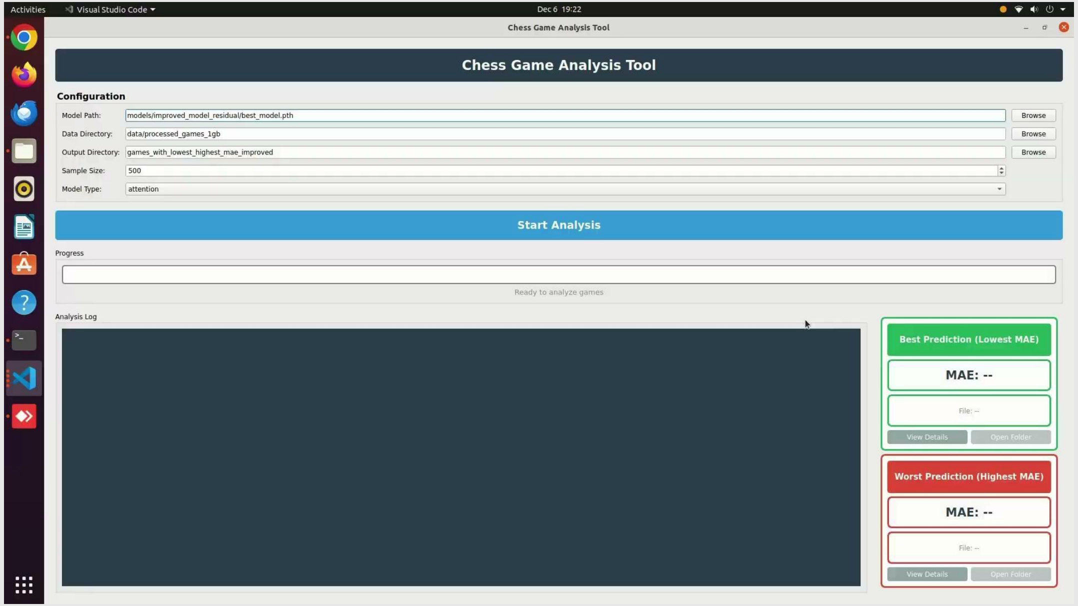




RESULTS SUMMARY

Modela	Test MAE	Improvement (vs. 308.6)	Parameters
Baseline (Exp 01)	308.6	-	2.4M
Improved No Residual (02)	303.8	-4.7 ✅	3.6M
With Residual (03)	324.1	+15.5 ❌	3.8M
+ Augmentation (04)	303.0	-5.5 ✅	3.8M
+ Cosine Scheduler (05)	335.0	+26.4 ❌	3.8M
All Features (Improved) (06)	303.3	-5.2 ✅	3.8M
Attention Basic (07)	304.4	-4.2 ✅	4.1M
+ Move Importance (08)	304.8	-3.8 ✅	4.1M
+ All Features (Attn) (09)	286.5	-22.1 ✅	4.1M

PROJECT DEMO



INTERESTING FINDINGS

Challenge	Solution
Residual connections increased overfitting	Combined it with data augmentation
Learning rate scheduling alone insufficient	Combined scheduler with augmentation
Time control variability where ultrabullet consistently 2x harder than other time controls	May need time-control-specific models for production

LESSONS LEARNED

- Attention mechanisms (especially move importance weighting) were more effective than simply increasing model depth
- Residual connections need to be balanced with regularization techniques like data augmentation
- Time control variability presents a significant challenge - UltraBullet games are inherently harder to predict
- The combination of architectural and training improvements works better than individual changes

CONCLUSION & FUTURE WORK

ARTI
CIAL
INTE
IGEN
(AI)

/(AI)/(AI)/

CONCLUSION

We successfully improved upon the baseline RatingNet model by implementing attention mechanisms combined with architectural enhancements, achieving a 22.1 Elo reduction in MAE (from 308.6 to 286.5). While this is still higher than the state-of-the-art 182 Elo reported in the original RatingNet paper, our attention-based approach demonstrates the value of focusing on critical moves in chess games.

FUTURE WORK

- Implement **engine-derived** features (centipawn loss, blunder probability) from Stockfish
- Explore time-control-specific models to handle the variability between bullet and classical games
- Further hyperparameter tuning to reduce overfitting and improve model stability
- Explore **transformer-based** architectures as an alternative to LSTM
- Investigate why cosine scheduler alone decreased performance, potentially exploring different learning rate schedules.

EACH TEAM MEMBER'S CONTRIBUTION

Michael Reda	Abdelrahman Elazab
Successfully deployed the baseline locally by **resolving CUDA driver conflicts** and initializing the NVIDIA GPU for high-speed training	Focused on **initial code comprehension, library setup**, and troubleshooting the limited GPU environment on Google Colab
Implemented a **key architectural update** by successfully increasing the LSTM layer count from 3 to 4, satisfying a proposed update	Explored **alternative evaluation metrics** by running experiments using Mean Squared Error (MSE) in addition to Mean Absolute Error (MAE)
Managed and **optimized resource usage** by solving the persistent CUDA Out of Memory (OOM) error through reducing the `train_batch_size` and `val_batch_size`	Conducted trials on varying large dataset sizes (6.4 GB vs. 10.7 GB) and performed general hyperparameter tuning (epochs, batch sizes)
Designed and implemented the **attention** mechanisms	Implemented the **deeper CNN architectures** and residual connection framework
Improved training strategy by using the **Cosine Learning** rate schedules	Implemented the **data augmentation** pipeline (horizontal flip, color swap)

THANKS

ARTI
CIAL
INTE
IGEN
(AI)

/(AI)/(AI)/

RESOURCES

- Chowdhary, S., Iacopini, I., & Battiston, F. (2023). Quantifying human performance in chess. Scientific Reports, 13(1), 2113. https://doi.org/10.1038/s41598-023-29272-6
IEEE. (2024). Big Data Cup 2024: Predicting chess puzzle difficulty. KnowledgePit. https://knowledgepit.ml/predicting-chess-puzzle-difficulty/

- Omori, M., & Tadepalli, P. (2024). Chess rating estimation from moves and clock times using a CNN-LSTM. arXiv preprint arXiv:2409.11506. https://arxiv.org/abs/2409.11506

- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830. https://www.researchgate.net/publication/51969319_Scikit-learn_Machine_Learning_in_Python

- Tijhuis, T., Blom, P. M., & Spronck, P. (2023). Predicting chess player rating based on a single game. In 2023 IEEE Conference on Games (CoG) (pp. 1–8). IEEE. https://doi.org/10.1109/CoG58078.2023.10324584