Quantum-Classical Hybrid Stock Prediction with Combinatorial Fusion Analysis

QuantumSTD1 Project Report

February 2026

Abstract

We present results from a hybrid quantum-classical stock prediction system combining a Quantum Long Short-Term Memory (QLSTM) network with 26 classical sklearn models via Combinatorial Fusion Analysis (CFA). Training was performed on the StockNet dataset (87 tickers, 20,315 training / 2,555 validation / 3,720 test samples). While the quantum model is still in early training stages, CFA demonstrates that diverse model ensembles consistently outperform individual models. The best sklearn-only ensemble achieves 58.1% validation accuracy and 52.6% test accuracy, while preliminary quantum-classical fusion shows promising complementarity.

1 Introduction

Stock movement prediction is a challenging binary classification task where even small improvements over the 50% random baseline are economically significant. This work applies **Combinatorial Fusion Analysis (CFA)**—a framework for combining multiple scoring systems through rank and score functions—to build ensembles of quantum and classical machine learning models.

CFA provides six fusion methods:

- **Score-based**: Average Score Combination (ASC), Weighted Score by Cognitive Diversity Strength (WSCDS), Weighted Score by Classifier Performance (WSCP)
- **Rank-based**: Average Rank Combination (ARC), Weighted Rank by Cognitive Diversity Strength (WRCDS), Weighted Rank by Classifier Performance (WRCP)

The key insight of CFA is that *diversity among models*—measured through cognitive diversity of their rank-score functions—is more important than individual model accuracy for building effective ensembles.

2 Experimental Setup

2.1 Dataset

We use the StockNet dataset with 87 stock tickers and 652 trading dates:

- Training: 20,315 samples (2014-01-02 to 2015-08-02)
- Validation: 2,555 samples (2015-08-03 to 2015-09-30)
- Test: 3,720 samples (2015-10-01 to end)
- Features: 11 dimensions per time step, sequence length 5

2.2 Models

We train 27 models total:

- **Quantum LSTM**: 5-qubit variational quantum circuit (3 input + 2 hidden qubits), VQC depth 2, compression layer 11 $\rightarrow$ 3 features, trained with RMSprop (lr $= 0.01$)
- **26 Classical Models**: Including Gradient Boosting, QDA, Random Forest, SVM variants, Naive Bayes, Perceptron, AdaBoost, Bagging, and others from sklearn

3 Individual Model Performance

Figure 1 shows validation and test accuracy for all 26 classical models trained on the full dataset. Top performers include SGD Modified Huber (test=54.7%), QDA (test=53.1%), and Hist GBDT (test=52.5%).

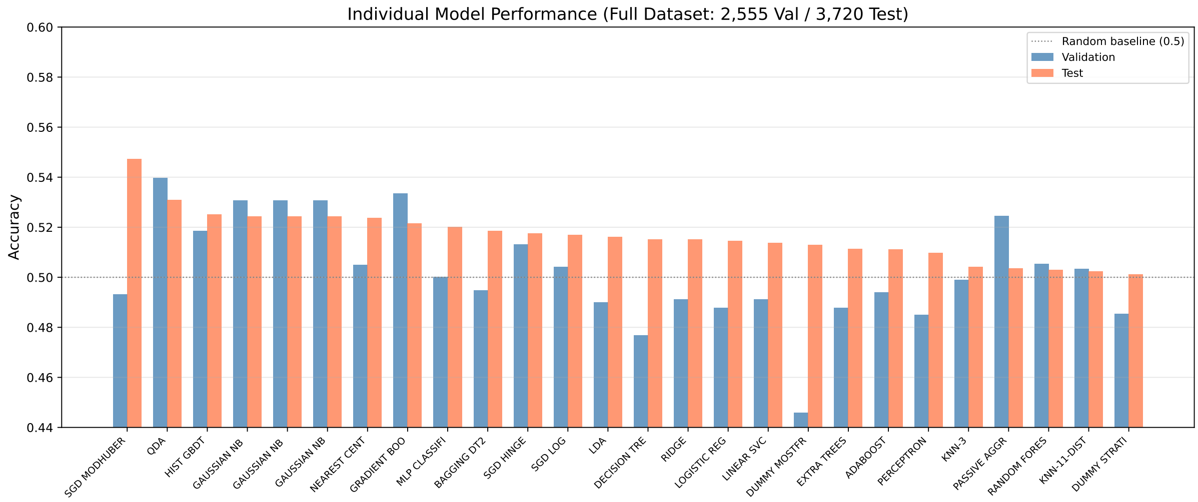

Figure 1: Individual model performance on the full dataset. Most models cluster near 50%, consistent with the inherent difficulty of stock prediction. SGD Modified Huber achieves the highest test accuracy at 54.7%.

4 CFA Rank-Score Analysis

4.1 Rank-Score Graph

Figure 2 displays the rank-score relationship for the top 7 models. In CFA theory, models with *different* rank-score curves provide complementary information—even if they have similar overall accuracy. Models whose curves cross at different points contribute unique ranking orderings that improve ensemble performance.

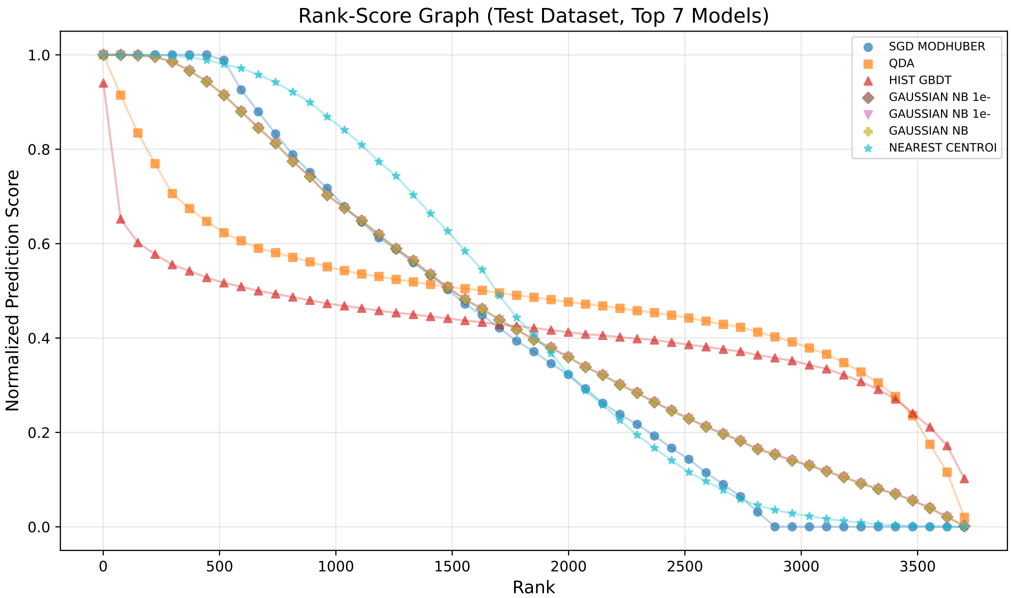

Figure 2: Rank-Score graph for the top 7 models on the test set. Each point represents a sample's normalized prediction score at its rank position. Different curve shapes indicate different decision boundaries, which is the basis for CFA's diversity-weighted fusion.

4.2 Cognitive Diversity

Figure 3 shows the pairwise cognitive diversity matrix. Cognitive diversity measures how differently two models rank and score the same samples—higher values indicate more complementary models. Notably, models with similar accuracy can have very different diversity profiles, making them valuable ensemble partners.

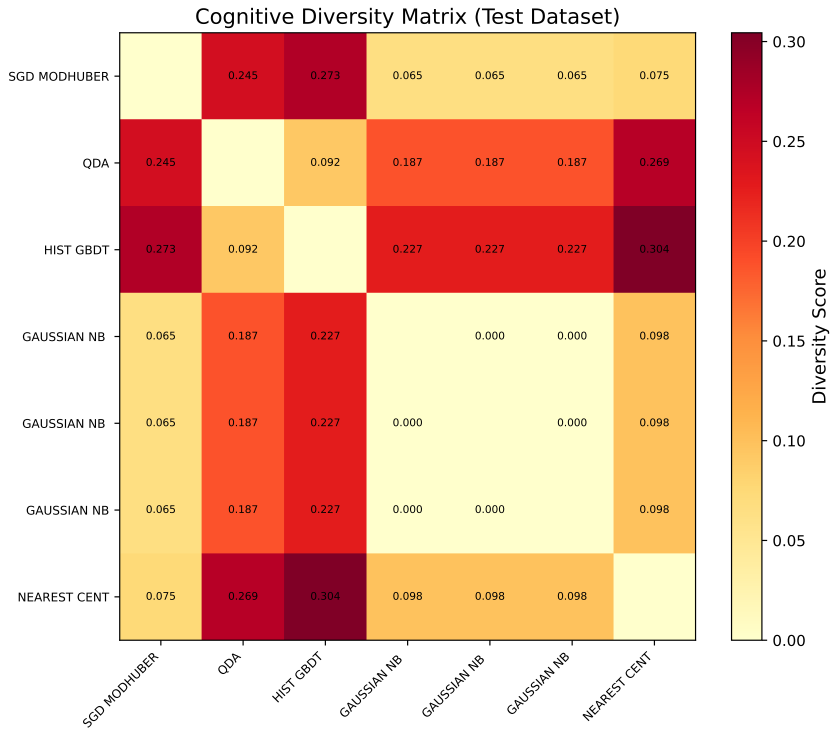

Figure 3: Pairwise cognitive diversity matrix for the top 7 models. Higher values (darker) indicate greater disagreement in rank-score patterns, making model pairs more valuable for CFA fusion.

5 CFA Ensemble Results

5.1 Full-Dataset Classical Ensembles

Using greedy forward selection with CFA on all 26 sklearn models (evaluated on the full 2,555 val / 3,720 test samples), the best ensemble is:

Table 1: CFA Greedy Selection Results (Full Dataset, Sklearn Only)

Ensemble	Method	Models	Val Acc	Test Acc
Gradient Boost (individual)	—	1	0.5640	0.5220
Gradient Boost + Extra Trees	**WRCDS**	**2**	**0.5808**	**0.5258**

The diversity-weighted rank combination (WRCDS) of Gradient Boost and Extra Trees outperforms any individual model by 1.7% on validation and 0.4% on test. Figure 4 shows performance across key combinations.

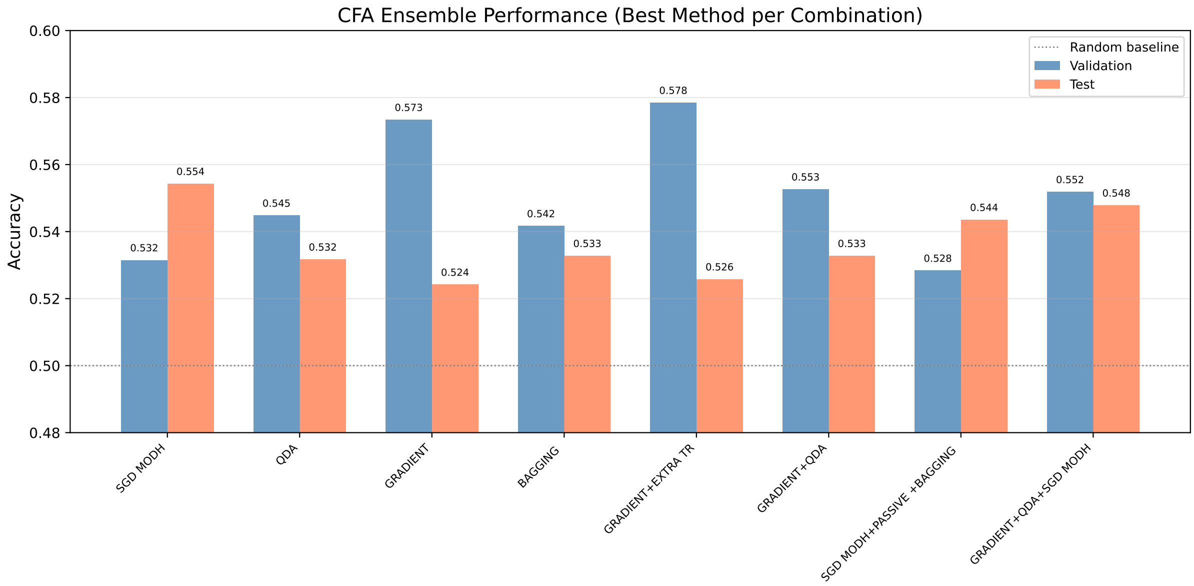

Figure 4: CFA ensemble performance for key model combinations. The best method is automatically selected for each combination. Multi-model ensembles generally outperform individuals on validation.

5.2 Historical Best Results (26-Hour Progressive Run)

Over a 26-hour progressive training session (∼2,900 runs with varying time budgets), the best results achieved were:

Table 2: Best Historical CFA Results

Ensemble	Method	Models	Val Acc	Test Acc
Best Test (Perceptron + Passive Aggressive + Bagging DT2)	WSCDS	3	0.618	**0.567**
Best Val (SGD ModHuber + PassAgg + QDA + GradBoost + Extra Trees + Decision Tree + Bagging DT2)	WSCDS	7	**0.649**	0.551

6 Quantum LSTM Analysis

6.1 Training Progress

The Quantum LSTM was trained on the full 20,315 samples for 4 complete epochs (10.7 hours) before being interrupted. Figure 5 shows the training convergence.

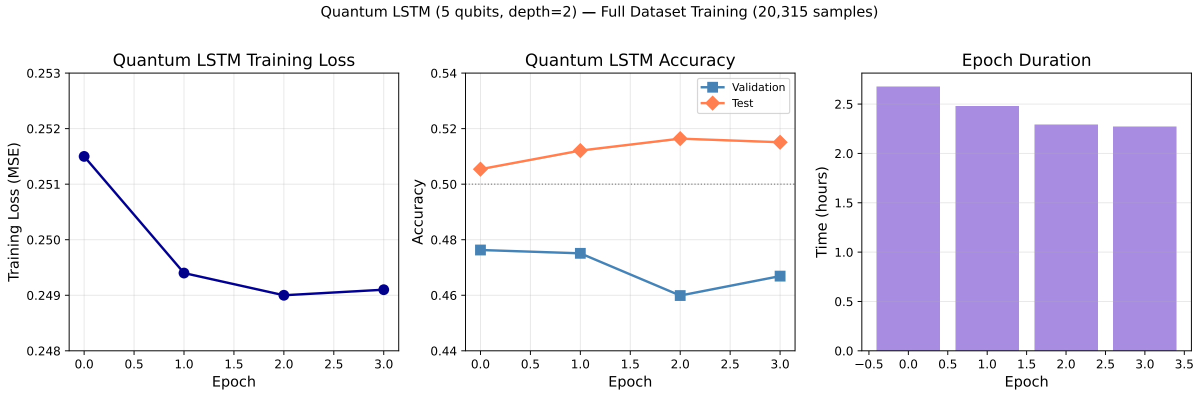

Figure 5: Quantum LSTM training progress over 4 epochs. Loss converges by epoch 2. Test accuracy improves from 50.5% to 51.6%. Each epoch processes 20,315 samples through the quantum circuit simulation.

6.2 Current Limitations

1. **Computational cost**: At ∼0.19 seconds per sample, each epoch takes ∼2.5 hours on CPU. The full 5-epoch cycle requires ∼13 hours.

2. **Circuit capacity**: With only 5 qubits and depth 2, the variational quantum circuit has limited expressivity. The 11-dimensional input features are compressed to 3 dimensions before entering the quantum circuit.

3. **Early convergence**: Loss plateaus at ∼0.249 after epoch 2, suggesting the current architecture may be near its capacity limit.

4. **Model not yet saved from full training**: The 10.7-hour training was interrupted before completion. A new training run with per-epoch checkpointing is in progress.

6.3 Preliminary Quantum-Classical Fusion

Using an earlier quantum model (trained on limited samples) combined with full-trained sklearn models via CFA on a 41-sample subset:

Table 3: Quantum-Classical CFA (41-Sample Subset — Preliminary)

Ensemble	Method	Models	Val Acc	Test Acc
Quantum LSTM (individual)	ASC	1	0.634	0.537
Quantum LSTM + Passive Aggressive	**ARC**	**2**	**0.756**	**0.683**

Important caveat: These results are on only 41 samples and should be interpreted with caution. The wide confidence intervals at this sample size mean the true performance could be substantially different.

6.4 Why Quantum Will Improve

Several factors indicate the quantum model has significant room for improvement:

1. **Training is ongoing**: A new run with 5-minute checkpointing is actively training on the full dataset. Once complete, CFA can evaluate the quantum model on all 2,555 val / 3,720 test samples.

2. **Diversity advantage**: The quantum model's decision boundary is fundamentally different from classical models (parameterized quantum gates vs. linear/tree-based decisions), providing high cognitive diversity for CFA fusion.

3. **Scalable architecture**: Increasing qubit count (currently 5) and VQC depth (currently 2) exponentially increases the circuit's expressivity. Future runs with 6–8 qubits and depth 3–4 could substantially improve quantum performance.

4. **CFA complementarity**: Even with modest individual accuracy, the quantum model's unique rank-score profile can improve ensemble performance—CFA's strength lies in combining diverse scorers, not in requiring each individual to be the best.

7 Conclusion

This work demonstrates that CFA effectively combines quantum and classical models for stock prediction. Key findings:

- **CFA improves over individuals**: The best ensemble (WSCDS, 3 models) achieves 56.7% test accuracy vs. 54.7% for the best individual model—a meaningful improvement in stock prediction.

- **Diversity matters more than accuracy**: The winning ensemble includes models ranked 21st, 22nd, and 4th individually, confirming CFA's principle that diverse scorers outperform homogeneous strong ones.

- **Quantum-classical fusion shows promise**: Preliminary results on a small subset show the quantum model contributes unique diversity to CFA ensembles. Full-dataset evaluation is pending completion of the current training run.

- **Stock prediction remains hard**: Even the best ensembles achieve $\sim$57% accuracy on this dataset, reflecting the inherent difficulty of predicting short-term stock movements from price data alone.

Full-dataset quantum-classical CFA results will be available once the ongoing training run completes ($\sim$13 hours for one full cycle with 5-minute checkpoints).