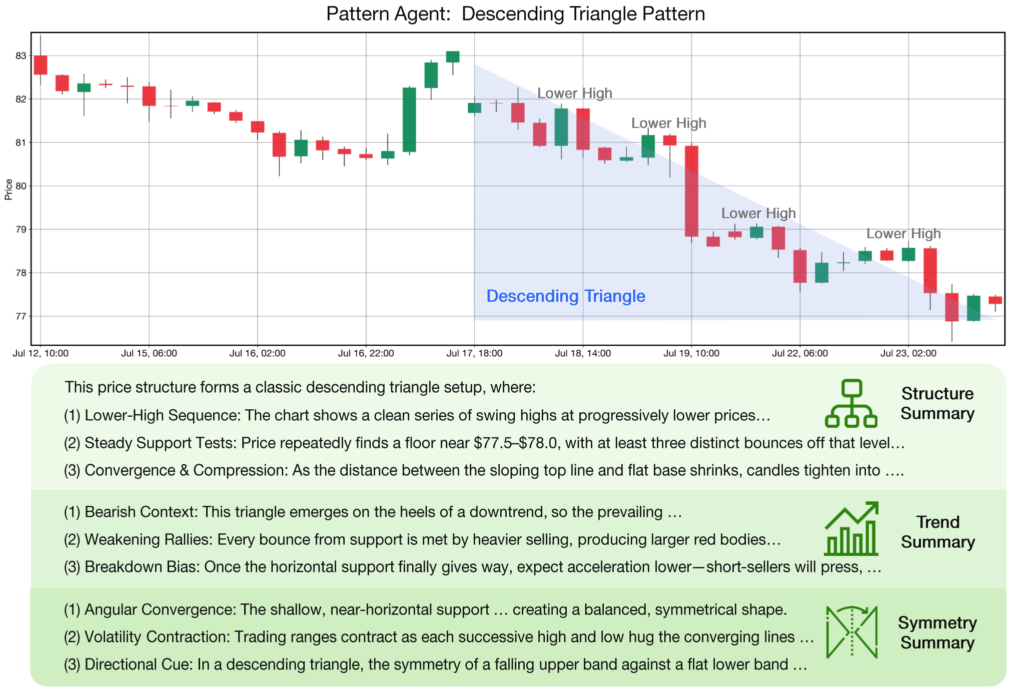

This price structure forms a classic descending triangle setup, where:

(1) Lower-High Sequence: The chart shows a clean series of swing highs at progressively lower prices…

(2) Steady Support Tests: Price repeatedly finds a floor near $77.5–$78.0, with at least three distinct bounces off that level…

(3) Convergence & Compression: As the distance between the sloping top line and flat base shrinks, candles tighten into ….

Structure Summary

(1) Bearish Context: This triangle emerges on the heels of a downtrend, so the prevailing …

(2) Weakening Rallies: Every bounce from support is met by heavier selling, producing larger red bodies…

(3) Breakdown Bias: Once the horizontal support finally gives way, expect acceleration lower—short-sellers will press, …

Trend Summary

(1) Angular Convergence: The shallow, near-horizontal support … creating a balanced, symmetrical shape.

(2) Volatility Contraction: Trading ranges contract as each successive high and low hug the converging lines …

(3) Directional Cue: In a descending triangle, the symmetry of a falling upper band against a flat lower band …

Symmetry Summary

Figure 4 Case sample of the `PatternAgent` on CL (2024). The agent extracts swing pivots, fits a declining resistance line through lower highs, and identifies flat support near 78. As the gap narrows, it classifies the formation as a descending triangle and generates three structured summaries: Structure ("lower highs" vs. "flat support"), Trend (bearish breakdown bias), and Symmetry (triangular convergence). Dashed edges and EMA overlays are visual aids only; the classification is derived solely from bar geometry.

$\mathcal{R}_{cc}$, $\mathcal{R}_{\max}$, and $\mathcal{R}_{\min}$, incorporate risk-constrained execution, simulating realistic stop-loss and take-profit behavior. Specifically, a trade is exited at the first price among the next three candlesticks that hits either the stop-loss or reward threshold. We adopt a fixed stop-loss threshold $\rho = 0.0005$ (i.e., 0.05%), selected to reflect the relatively small fluctuations typical within a short three-candlestick forecast horizon, consistent with prior work (Kissell, 2013). The corresponding reward threshold is determined using the LLM-generated risk–reward ratio $r = \frac{\mathcal{R}}{\rho}$, where $\mathcal{R} = r * \rho$ is the maximum allowed gain and $-r$ is the maximum allowed loss.

$\mathcal{R}_{\max}$ represents the best-case rate of return (RoR) achievable over the next three candlesticks under the current LLM-issued trading decision (either LONG or SHORT). It assumes the optimal exit occurs at the most favorable intra-candle price point, i.e., the maximum high for a long position or the minimum low for a short position. Conversely, $\mathcal{R}_{\min}$ captures the most adverse price movement during the same interval. These two metrics represent a bounded range of maximum profit or loss outcomes under realistic, risk-managed execution (Lo, 2001).

Asset	Method	Acc α ↑	$\Delta\alpha\%$ ↑	$\mathcal{R}_{cc}$ ↑	$\mathcal{R}_{\max}$ ↑	$\mathcal{R}_{\min}$ ↑
BTC	Baseline	45.0	–	-0.009	1.220	-1.245
	LR	46.0	+2.2%	-0.066	**1.245**	**-1.210**
	XGBoost	45.3	+0.7%	-0.050	1.218	-1.331
	Our	**50.7**	**+12.7%**	**0.089**	1.232	-1.212
CL	Baseline	41.0	–	-0.373	0.970	-1.348
	LR	54.3	+32.4%	-0.114	1.178	-1.141
	XGBoost	40.0	-2.4%	-0.056	0.958	-1.151
	Our	**55.0**	**+34.1%**	**-0.008**	**1.200**	**-1.119**
DJI	Baseline	47.0	–	0.048	0.755	-0.793
	LR	52.0	+10.6%	0.149	0.790	-0.725
	XGBoost	47.3	+0.6%	-0.020	0.874	-0.660
	Our	**52.3**	**+11.3%**	**0.163**	**0.891**	**-0.649**
ES	Baseline	51.0	–	-0.048	0.538	-0.552
	LR	43.0	-15.7%	0.032	0.553	-0.546
	XGBoost	52.0	+2.0%	-0.182	0.440	-0.644
	Our	**55.0**	**+7.8%**	**0.179**	**0.613**	**-0.485**
VIX	Baseline	46.3	–	0.059	3.259	-3.157
	LR	48.7	+5.2%	-0.140	3.407	-3.099
	XGBoost	53.3	+15.1%	0.161	3.325	-3.110
	Our	**54.7**	**+18.1%**	**0.458**	**3.872**	**-2.851**
NQ	Baseline	43.7	–	-0.140	0.646	-0.793
	LR	48.7	+11.4%	0.147	0.782	-0.670
	XGBoost	47.3	+8.2%	-0.007	0.706	-0.753
	Our	**55.3**	**+26.5%**	**0.216**	**0.814**	**-0.639**
QQQ	Baseline	47.3	–	-0.048	0.930	-1.017
	LR	56.0	+18.4%	0.175	1.113	**-0.849**
	XGBoost	52.7	+11.4%	0.210	**1.206**	-0.973
	Our	**59.7**	**+26.2%**	**0.211**	1.052	-0.881
SPX	Baseline	47.3	–	-0.162	0.719	-0.862
	LR	59.7	+26.2%	**0.377**	0.960	-0.648
	XGBoost	60.0	+26.8%	0.050	0.782	-0.712
	Our	**63.7**	**+34.6%**	0.341	**0.965**	**-0.641**

Table 1 Performance comparison across trading symbols. Results are shown for random(Baseline), Linear Regression(LR), XGBoost, and our QuantAgent. Bold values indicate the best performance for each metric across methods. Upward arrows (↑) denote metrics where higher values are better.

5 Results

5.1 Main Results

In Table 1, we compare our agent-based LLM trader to the three baselines, Random Baseline, Linear Regression(LR) and XGBoost, across eight widely traded markets.

From the table, we draw several key observations: *(i)* Our accuracy outperforms all methods across the eight evaluated markets especially on NQ, where we achieve a 26.5% increase over the random baseline and a clear margin over both LR and XGBoost. *(ii)* Despite the presence of risk caps, our $\mathcal{R}_{\text{cc}}$ still achieves the best performance in 7 out of 8 assets, suggesting that our model can consistently capture profitable short-term trends under realistic trading constraints. *(iii)* We obtain the highest $\mathcal{R}_{\max}$ in 6 out of the 8 markets and are nearly tied with the best performer in the remaining two, indicating our system effectively captures potential upside while respecting risk bounds. *(iv)* Similarly, our $\mathcal{R}_{\min}$ shows strong robustness, ranking among the least negative values across most assets. This implies that our method not only captures gains but also limits downside risk effectively.

Overall, the results highlight that our approach generalizes well across diverse asset classes in 4-hour time frame, balancing accuracy and return while maintaining robust risk control.

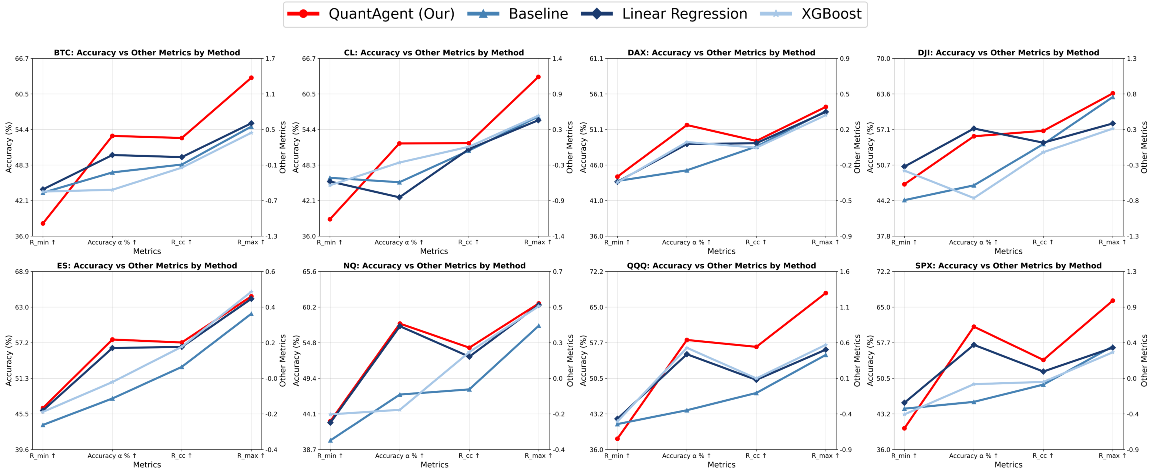

Figure 5 1-hour performance comparison across eight assets. Results are shown for random (Baseline), Linear Regression, XGBoost, and our QuantAgent. Arrows indicate higher values are better.

Furthermore, Figure 5 shows a comparative performance trend with same metrics for each asset in the 1-hour time frame. Our method (QuantAgent) consistently outperforms all baselines across most metrics and markets, especially in SPX, QQQ, and BTC where our method shows the most pronounced performance gap. The red line (QuantAgent) dominates across most plots, indicating both higher directional accuracy and better risk-aware returns, though it shows less satisfactory $\mathcal{R}_{\min}$ in some assets. Baseline, Linear Regression, and XGBoost exhibit weaker and less stable patterns, often lagging across most metrics. This visualization highlights the robustness and generalization capability of our approach under both profit-seeking and risk-constrained conditions in 1-hour time horizon.

5.2 Case Study on Continuous Short-Term Prediction

To evaluate short-horizon prediction consistency, the LLM's directional accuracy was further tested on a randomly selected 100-bar SPX segment using 10 overlapping windows, each offset by 5 bars (Qin et al., 2017).

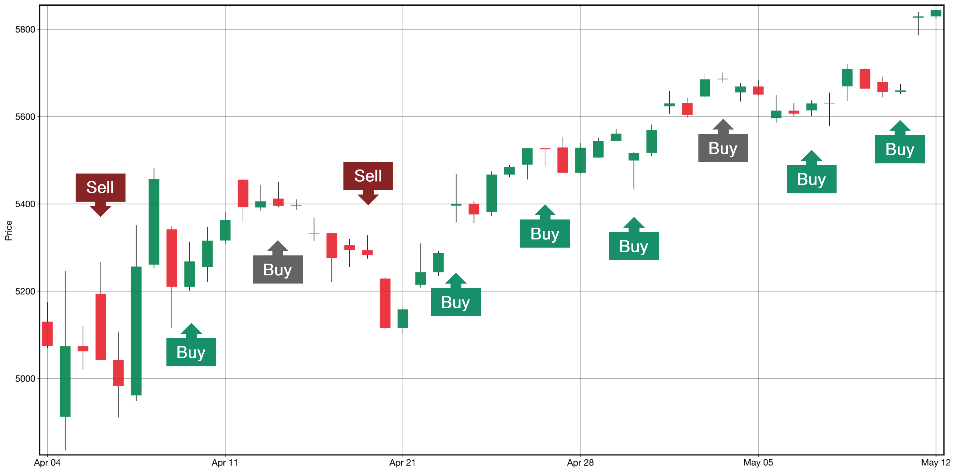

Figure 6 **Case study of high-frequency prediction on SPX (2025).** Correct forecasts (8/10) are marked with green Buy or red Sell badges, while mispredictions are shown in grey (2/10).