A Foundational Deep Learning Approach to Cryptocurrency Market Forecasting

Rodolphe Lucas
MSc in Computational and Software Techniques in Engineering, Cranfield University
Supervisor: Dr. Jun Li — April–September 2025

Abstract

Cryptocurrency markets exhibit extreme volatility and regime shifts, making short-term price forecasting difficult. This study evaluates whether deep learning can deliver a sustainable edge for Bitcoin *BTC/USDT* prediction and trading. The work is structured in two parts.

Part I. The task is framed as a 4-hour ahead classification problem with three classes (Buy/Hold/Sell). A sliding-window quantile labelling scheme adapts to local volatility. LSTM, GRU, CNN–RNN and transformer architectures are benchmarked against an XGBoost baseline. Results show that while deep models modestly outperform tree-based methods, performance is unstable across temporal splits and Sell signals remain difficult to capture.

Part II. The problem is reformulated as multi-horizon regression across 1–48h horizons using separate encoders for 1h, 4h and daily sequences. We propose a novel multi-timeframe architecture inspired by the Temporal Fusion Transformer, but designed to be lighter and less computationally demanding. A weighted gating mechanism fuses representations, reflecting the James–Stein principle that joint estimation reduces variance. Models are trained on temporal splits covering 2017–2022 and ensembled using validation-weighted averaging. On the unseen test split, the ensemble achieves a *Directional Accuracy* of approximately 89% and an R^2 of 0.873 at the 24h horizon. Forecast calibration proves stronger at the 12–24h horizons than at 1–4h, highlighting the value of mid-range dynamics for robust prediction.

- **Research question:** Can deep learning models provide a robust edge for short-term Bitcoin forecasting and trading ?

Part I Directional BTC (H4) Forecasting : Benchmarking Multiple Deep Learning Models

Problem formulation :
The first stage of our study frames the forecasting task as a three-class classification problem on the **4-hour timeframe (H4)**. Given a sequence $\mathbf{X}_t \in \mathbb{R}^{w\times d}$ of $w = 24$ **past periods of 4h** with $d = 19$ **features** (OHLCV, Ichimoku, RSI, MACD, Bollinger, etc.), the goal is to predict the label at horizon $h = 4h$:

$$y_t \in \{\text{Buy}, \text{Hold}, \text{Sell}\}, \qquad \hat{y}_t = \arg\max_k f_\theta(\mathbf{X}_t)_k.$$

Labelling Strategy : We designed an adaptive labelling scheme that adjusts to market volatility and future return distributions. Future returns r_{t+h} are calculated from price changes, with a sliding window estimating local volatility to dynamically set Buy/Hold/Sell thresholds. Returns are mapped to soft probabilities using sigmoid functions, discretized into one-hot labels for training when model confidence is sufficient.
This method ensures that the labels remain realistic and responsive to changing market regimes, which is critical given the strong volatility of cryptocurrencies.

Models compared :
We benchmarked several architectures: XGBoost, LSTM, GRU, CNN–LSTM, Transformer, TCN, and Feed-forward NN.
Despite the adaptive labelling, some models reached only **Accuracy** $\approx$ **0.54** and **macro-F1** $\approx$ **0.38** (Table 1). Directional labelling induces imbalance (*Hold* dominates, *Sell* rare), so accuracy overstates performance while decision-critical tails are missed. In that way, Macro-F1 and per-class F1 provides a clearer view.

Table 1: Performance comparison of tested models on the test set (Accuracy and F1-scores by class).

Model	Accuracy	Macro F1	Buy F1	Hold F1	Sell F1
Standard LSTM	0.5051	**0.3838**	0.3685	0.6405	0.1424
LSTM + Attention	0.5566	0.3171	0.2486	**0.7028**	0.0000
GRU + LSTM	0.4600	0.3100	**0.6000**	0.3400	0.0000
GRU + LSTM + Attention	0.5075	0.3380	0.3543	0.6445	0.0151
CNN-LSTM	0.4834	0.3669	0.3787	0.6106	0.1113
Temporal Conv. Network (TCN)	0.5407	0.2977	0.1332	0.7023	0.0577
Transformer + GRU	**0.5621**	0.2985	0.1844	0.7110	0.0000
FNN	0.4798	0.3215	0.3578	0.6066	0.0000
XGBoost	0.4323	**0.3880**	0.4132	0.5392	**0.2116**

Task complexity :
The H4 horizon is challenging due to high volatility, rapid trend reversals, and ambiguous Buy/Hold/Sell boundaries, making classification difficult with noisy, overlapping signals. With only 24 historical periods, the context is limited; in Part II, we address this by incorporating multi-timeframe sequences (H1, H4, D1) to provide richer context.

Part II From Forecasting to Execution : Multi-Horizon Model for Trading Strategies

In Part II, we shift perspective by adopting a *multi-horizon regression* setup that predicts log-normalized returns jointly at $\{1h, 4h, 12h, 24h, 48h\}$. A central innovation lies in the *multi-timeframe design*: dedicated encoders process **H1**, **H4**, and **D1** sequences (lengths 100, 25, 30, respectively), whose latent representations are subsequently fused before feeding horizon-specific heads. This design explicitly mirrors the **Top-Down Analysis** method widely practiced in trading.

Problem Formulation :
We observe three timeframes simultaneously with **19 features** (OHLCV, Ichimoku, RSI, MACD, Bollinger):

$$\mathbf{X}_t^{H1} \in \mathbb{R}^{\mathbf{100}\times 19}, \qquad \mathbf{X}_t^{H4} \in \mathbb{R}^{\mathbf{25}\times 19}, \qquad \mathbf{X}_t^{D1} \in \mathbb{R}^{\mathbf{30}\times 19}.$$

P_t denotes the asset close price at time t. Predict tanh-transformed returns at horizons $h \in \{1h, 4h, 12h, 24h, 48h\}$:

$$Y_{t+h} = \tanh\left(\frac{P_{t+h} - P_t}{P_t}\right), \qquad f_\theta : (\mathbf{X}_t^{H1}, \mathbf{X}_t^{H4}, \mathbf{X}_t^{D1}) \mapsto (\hat{Y}_{t+1h}, \hat{Y}_{t+4h}, \hat{Y}_{t+12h}, \hat{Y}_{t+24h}, \hat{Y}_{t+48h}).$$

Architecture : The proposed model (Figure 1) consists of four main stages:

- **Timeframe encoders.** Each sequence (H1, H4, D1) is processed by a lightweight LSTM–GLU encoder, producing hidden states $H^{(\tau)} = [h_1, \dots, h_T]$ that summarize the temporal dynamics.
- **Temporal aggregation.** Instead of relying only on the last state h_T, we extract four complementary summaries — last, mean, max, and attention pooling — so as to preserve both extremes and global structure.
- **Learned fusion gate.** The embeddings from all timeframes are adaptively combined by a softmax-weighted gating mechanism, which learns the relative contribution of H1, H4, and D1 depending on market context.
- **Shared trunk and multi-horizon heads.** A shared representation layer captures common features, followed by horizon-specific regression heads ($g_{1h}, g_{4h}, g_{12h}, g_{24h}, g_{48h}$), enabling joint estimation across multiple forecast horizons.

Training uses the uncertainty-based multi-task loss of *Kendall*, which automatically balances the contribution of each horizon according to its noise level.

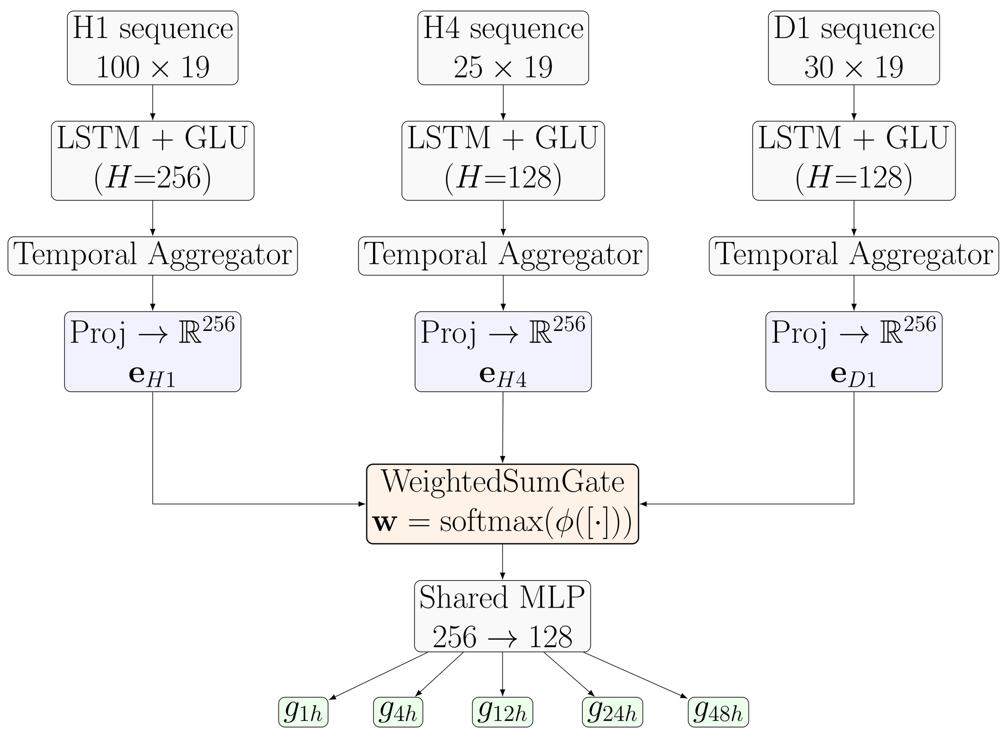

Figure 1: Multi-timeframe architecture (Part II).

Results (multi-horizon) : The proposed architecture was trained on four temporal folds (Splits 1–4) and evaluated strictly on the unseen out-of-sample fold (Split 5). In addition to the four stand-alone models, we constructed three ensembles (mean, median, validation-weighted) from the per-horizon predictions $\hat{\mathbf{Y}}_t^{(m)}$. This ensemble strategy improves stability and reduces variance compared to single-split models.
Overall, performance is strongest at intermediate horizons (**12–24h**), where noise is partly smoothed yet trends remain actionable. Very short horizons (1–4h) remain noise-dominated, while longer ones (48h) suffer from regime shifts. At 24h, for instance, the model achieves $R^2 \approx 0.87$ and Directional Accuracy ≈ 0.89 in the tanh-transformed space (Figures 2–3).

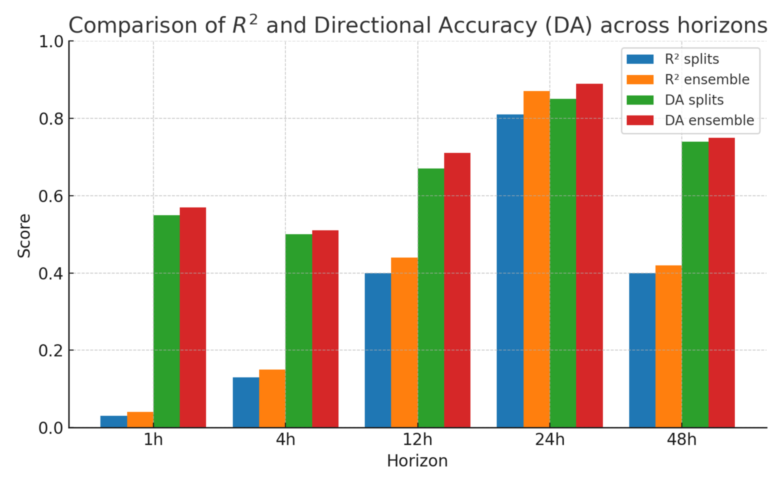

Figure 2: Predicted vs. actual returns at 24h ($R^2 \approx 0.87$).

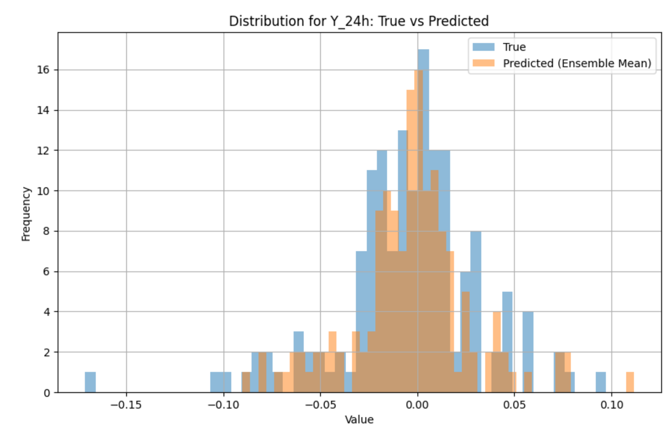

Figure 3: Distribution of predicted returns at 24h, showing calibration.

Table 2: Metrics averaged across all forecast horizons on the held-out test fold (Split 5). Best per column in **bold**.

Model	MSE	MAE	RMSE	R^2	DA
split_1	0.0005	0.0135	0.0216	0.4984	0.6684
split_2	0.0004	0.0132	0.0212	0.5194	0.6434
split_3	0.0005	0.0139	0.0220	0.4818	0.6539
split_4	0.0004	0.0129	**0.0210**	0.5278	0.6868
ensemble_mean	0.0004	**0.0127**	0.0207	0.5404	0.6868
ensemble_median	0.0004	0.0128	0.0208	0.5369	0.6763
ensemble_val_weighted	**0.0004**	**0.0127**	**0.0207**	**0.5411**	**0.6882**

Backtesting of Trading Strategy

Setup : Backtest on **unseen data** (post 2022-09-20), BTC/USDT. Three market regimes are tested, diagnosed with Forecastable Component Analysis (Ω_1):

- Downtrend (2025-01-20 to 2025-04-09, $\Omega_1 = 0.077$)
- Uptrend (2025-04-09 to 2025-08-08, $\Omega_1 = 0.1285$)
- High Volatility (2024-03-12 to 2024-11-06, $\Omega_1 = 0.1009$)

Decision Pipeline & Strategy : Multi-horizon predictions from the model are first aggregated through an ensemble (mean, median, or validation-weighted) and compressed into a single consensus score via cross-horizon gating. Trade entries are then triggered only when this score exceeds percentile-based thresholds (e.g. **P90** = top 10% most confident signals), ensuring strong selectivity but lowering trade frequency. The execution layer was designed as a very rigid stress test: trades are capped at 24h holding time, with a fixed 2% stop-loss and symmetric take-profit rules. Under these constraints, downtrend regimes yield systematic losses (Sharpe < 0), while uptrends produce modestly positive results (Hit Ratio $\approx$ 50%, Profit Factor $\approx$ 1.1). In high-volatility regimes, however, the framework demonstrates genuine predictive power: at P90, the strategy reaches a Sharpe ratio above 1.0 and **Profit Factor around 1.6** despite only 21 trades over 239 days. This confirms that the predictive model contains exploitable signal, but profitability is bottlenecked by rigid trade management rules. The strategy can be tuned to increase exposure or risk in order to maximize profit, while adaptive mechanisms such as volatility-scaled stops, dynamic profit-taking, or flexible sizing are needed to unlock sustainable long-term performance.

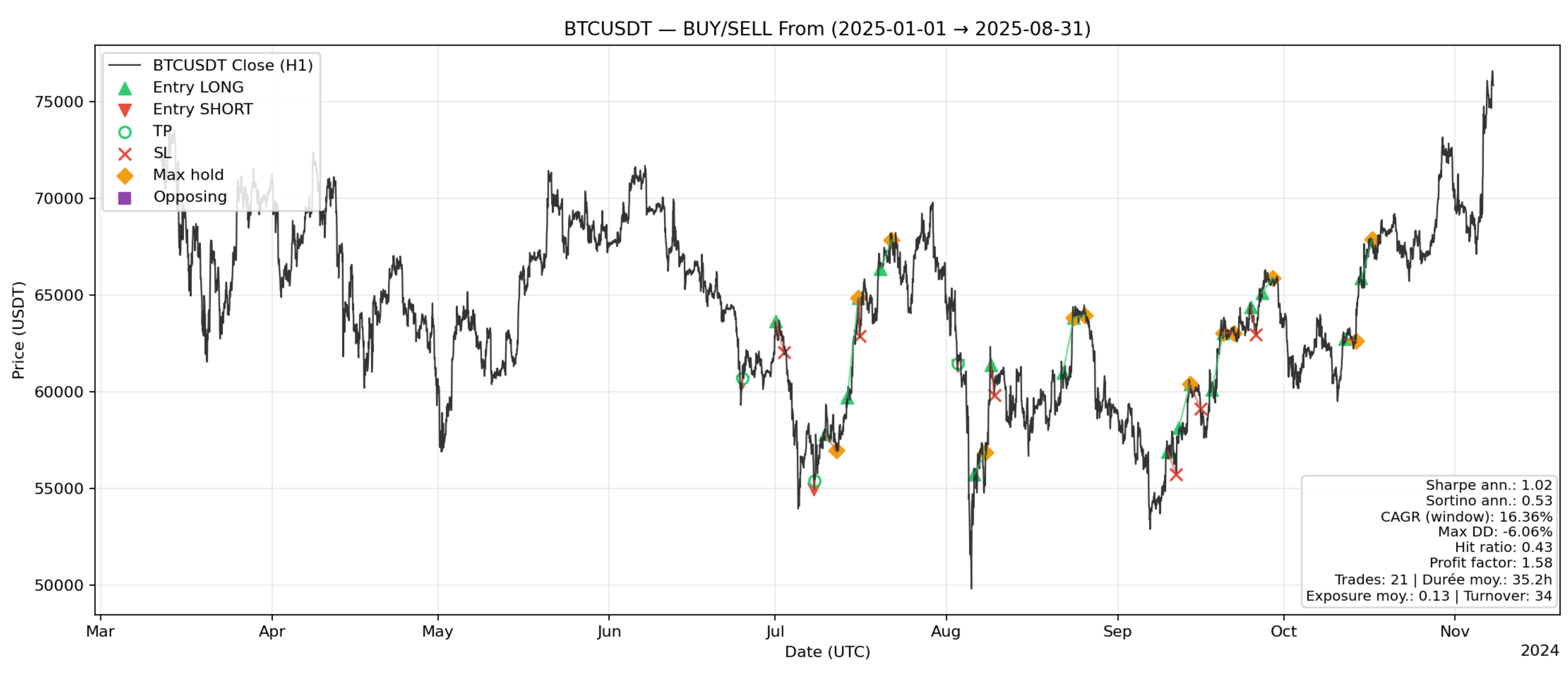

Figure 4: Example trades in Volatility regime ($P90$).