Machine Learning Methods to Exploit the Predictive Power of Open, High, Low, Close (OHLC) Data

Andrew D. Mann

A thesis submitted in partial fulfilment

of the requirements for the degree of

Doctor of Philosophy

of

University College London.

Department of Computer Science

University College London

2021

I, Andrew D. Mann, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

“*An investment in knowledge pays the best interest*”

- Benjamin Franklin

Abstract

Novel machine learning techniques are developed for the prediction of financial markets, with a combination of supervised, unsupervised and Bayesian optimisation machine learning methods shown able to give a predictive power rarely previously observed. A new data mining technique named *Deep Candlestick Mining* (DCM) is proposed that is able to discover highly predictive dataset specific candlestick patterns (arrangements of open, high, low, close (OHLC) aggregated price data structures) which significantly outperform traditional candlestick patterns. The power that OHLC features can provide is further investigated, using LSTM RNNs and XGBoost trees, in the prediction of a *mid-price directional change*, defined here as the mid-point between either the open and close or high and low of an OHLC bar. This target variable has been overlooked in the literature, which is surprising given the relative ease of predicting it, significantly in excess of noisier financial quantities. However, the true value of this quantity is only known upon the period's ending – i.e. it is an after-the-fact observation. To make use of and enhance the remarkable predictability of the mid-price directional change, multi-period predictions are investigated by training many LSTM RNNs (XGBoost trees being used to identify powerful OHLC input feature combinations), over different time horizons, to construct a Bayesian optimised trend prediction ensemble. This fusion of long-, medium- and short-term information results in a model capable of predicting market trend direction to greater than 70% better than random. A trading strategy is constructed to demonstrate how this predictive power can be used by exploiting an artefact of the LSTM RNN training process which allows the trading system to size and place trades in accordance with the ensemble's predictive certainty.

Impact statement

The work presented in this thesis contributes to the development of machine learning based predictive systems for financial markets by designing new data mining, trend prediction and ensemble methodologies using Long Short Term Memory (LSTM) Recurrent Neural Networks (RNNs), a variety of clustering algorithms and advanced optimisation techniques. This research provides several specific benefits. Firstly, the development of a novel and robust data mining technique demonstrates that a very high predictive accuracy from noisy and complex environments, such as financial markets, is achievable using simple feature representations. Secondly, a new methodology coined *Deep Candlestick Mining* is developed discovering novel candlestick patterns suitable for modern day financial markets; as the use of traditional candlestick charting has been an area of extreme conflict in academia and industry the discovery here that novel variants of these patterns can be successful is of clear interest. Thirdly, an undiscovered characteristic of financial timeseries is shown to be easily predicted, and exploited through a novel technique to infuse multi-period information, producing powerful predictive models which can also generate stable systematic profits across cryptoassets. Finally, the work of this thesis demonstrates how to design financial predictive systems that generalise over a range of asset classes while maintaining robust predictive power. In conjunction with the presentation here of this work, academic and industry publications and presentations have more widely publicised the benefits of adopting the novel techniques of this thesis, which have also been implemented in real-world systematic trading systems to demonstrate the practical benefits.

Acknowledgements

I would like to thank Dr. Denise Gorse who has been my supervisor throughout this process and someone who has meaningfully shaped the way I look at the world. If it was not for her detailed eye and passion for research, I would not have achieved my goals in as rigorous a manner.

In addition, I would like to thank my family, friends, and partner for their support throughout the process allowing me to maintain focus during this arduous journey.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter presents the motivation behind this thesis and the problems for which solutions are subsequently developed. Research contributions are then summarised, detailing publications, presentations, and contributions to the scientific community. The chapter ends with an outline of how this thesis is structured.

1.1 Research motivation

The ability to predict the movement of financial markets has been a longstanding aim of industry practitioners and academics, nowadays using a variety of techniques from technical analysis (TA) to machine learning (ML) and pattern recognition methodologies. The techniques used in this work focus on extracting novel valuable information from financial markets, which is a challenging task due to the efficiency of most contemporary financial markets, the high noise content of financial data, and its non-stationarity and low signal strength. Systems which can successfully exploit any remaining inefficiencies in contemporary financial data are of obviously significant interest given the financial rewards available for doing so.

One of the first pattern recognition predictive systems for financial forecasting was developed by the Japanese in the 18th century using formations of (open, high, low, close) or OHLC data called *candlestick* patterns to forecast the price of rice. Candlestick patterns are currently at the centre of a severe controversy surrounding their statistical significance in the directional prediction of prices, with many academics and industry practitioners proposing arguments for and against their usefulness (see section 2.2.1). Another form of more traditional financial market predictive system is that of *technical analysis* (TA), of which candlestick patterns are now known to be a sub-set. TA is the practice of applying simple statistical analysis such as averages over OHLC data structures and, on occasion, volume bars to construct systems which inform a trader if a market is trending, overbought/oversold or range bound amongst other classifications. Again, there has also been substantial controversy in the academic community relating to TA's usefulness and robustness, although the technique, like candlestick charting, remains popular amongst practitioners.

Today, the largest institutions and hedge funds on the planet deploy not only these more traditional tools but also advanced prediction systems which are highly efficient, scalable, distributed and able to handle massive quantities of data. As markets have become more efficient over time the need for such advanced predictive methodologies has become progressively higher as competition increases to exploit subtle market inefficiencies. This new generation of predictive systems are often heavily based on advanced mathematics, statistical analysis and machine learning methods with the aim of gaining an advantage over other market participants.

However this advantage cannot be gained simply by looking at directional movements. In traditional markets such as futures, equities and liquid currencies, generating alpha (defined as returns outperforming a benchmark or the underlying asset's natural gains) can often be difficult. Even if a statistically significant discovery is made in terms of directional forecasting the advantage can often only result in single percentage point profits if the said discovery is used with a large capital base. Therefore, the need for robust predictive systems capable of discovering statistically significant predictive signals which also scale well with the quantity of capital placed on them is of extreme interest to banks, hedge funds and even pension funds.

Adding to this challenge and complexity, there is a newly emerging asset class, that of *cryptoassets*, of which a subset are known as *cryptocurrencies*, the best-known example of which is Bitcoin. This new asset class holds promise as it has many inefficiencies compared to that of traditional markets, although these inefficiencies can be short lived, capacity limited and come with an increased risk when trading through a cryptoasset exchange, which frequently come under attack from hackers. This new market offers up new challenges as well as new opportunities for financial predictive systems given the extreme volatility of cryptocurrencies, market manipulations and a fragmented market microstructure.

This thesis aims to develop predictive systems with the ability to generalise effectively across both traditional and cryptoasset markets while providing robust and statistically significant predictive power. Furthermore, the effective use of this predictive power in wealth generation systems is investigated. This is a challenging task given the economics of trading (fees, low signal to noise rations, signal capacity, changing market dynamics, high competition for each inefficiency etc.). In summary, the main aims of this thesis are:

- To develop novel and powerful pattern mining techniques for the prediction of financial markets.
- To engage with the controversies surrounding the predictive power of candlestick patterns and to obtain at least a partial resolution.

- To exploit the predictive power of a previously unconsidered characteristic of financial time series data.
- To explore the use of multi-period information in the construction of predictive models.

1.2 Research contribution

The main goal of this research is to investigate the effectiveness of machine learning methods in the construction of predictive systems for directional forecasting. To achieve this goal will involve the fusion of ideas which are centuries old with cutting-edge machine learning techniques, enabling these ideas to be investigated, improved and reinvented. The research contributions made are:

- Development of a feature mining process capable of providing superior predictive power without the cost of excessive complexities in the feature space.
- Demonstration that while virtually none of the traditional candlestick patterns were predictive, novel candlestick patterns can be highly predictive.
- Discovery of a characteristic of a financial price series which can be exploited to give risk-adjusted returns that even in a cryptocurrency market are considerably in excess of those obtained from a ‘Buy & Hold’ strategy.
- Provision of evidence of the added value that multi-period information can bring to financial predictive ensembles.

1.3 Publications and presentations

Parts of this thesis have been published in the International Conference on Artificial Neural Networks:

Mann, A.D. & Gorse, D. (2017) A new methodology to exploit predictive power in (open, high, low, close) data. *Proceedings of the 26th International Conference on Artificial Neural Networks.* Alghero, Italy, 11-14 September 2017. Heidelberg: Springer. 495-502,

and the International Conference on Neural Information Processing:

Mann, A.D. & Gorse, D. (2017) Deep Candlestick Mining. *Proceedings of the 24th International Conference on Neural Information Processing.* Guangzhou, China, 14-18 November 2017. Heidelberg: Springer. 913-921.

Aside from presentations at these conferences parts of this thesis were also presented at the industrial conference QuantMinds, Vienna, May 2019. Further presentations with extracts of this work were given at these guest talks:

- Algorithmic Trading and Machine Learning for Cryptocurrencies, University of Glasgow, September 2017,
- Machine Learning for Cryptocurrency Trading, University of Newcastle, September 2017,
- Artificial Intelligence for Algorithmic Trading, Sussex, March 2017,
- Reinforcement Learning for Quantitative Finance, MathWorks, 2020.

Further benefits of this work have been realised in the “real-world” by developing scalable distributed systematic trading and research systems to deploy each research discovery and assess its resilience to live market data.

1.4 Thesis outline

This thesis is divided into six chapters inclusive of the introduction.

Chapter 2 provides necessary background to this work and a survey of related literature. The background section includes technical details of the machine learning models used throughout this work and the performance metrics used to ascertain the effectiveness of the predictions made. The literature survey reviews relevant literature in relation to candlesticks, technical analysis and the use of machine learning methods in financial prediction and trading systems.

Chapter 3 presents research pertaining to *Deep Candlestick Mining* (DCM), a method for identifying statistically significant novel candlestick patterns. A feature mining process using extreme gradient boosted trees in combination with correlation filtering is presented. The rich feature sets provided by this methodology are then used in a large-scale Bayesian and Hyperband hyperparameter optimisation process to tune an LSTM RNN's predictive power. Using the optimised LSTM RNN, a novel clustering process is proposed for the identification of new dataset specific candlestick patterns. These new patterns are statistically significant and able to provide predictive power significantly outperforming traditional ones. The feature mining and hyperparameter optimisation methodologies are used throughout the remainder of the work.

Chapter 4 extends the use of OHLC data structures as they have been shown to yield significant predictive power in Chapter 3. A new financial timeseries characteristic is proposed, derived from the structure of an OHLC bar. This is shown to be an easily predicted metric although since it relates to an after-the-fact observation it can be difficult to use for wealth generation.

Chapter 5 solves this last-mentioned problem, allowing the new metric to be used effectively for wealth generation. It does so by training many LSTM RNNs to predict this metric over differing horizons, using differing feature aggregation periods. The models are then used to construct a Bayesian optimised ensemble predictor achieving directional forecasts greater than 70% better than random, which is remarkable when considering the noise content within a financial time series. A market neutral trading strategy using the ensemble's predictive power is then presented, demonstrating an ability to generate exceptional risk-adjusted returns in the notoriously volatile cryptoasset market.

Lastly, **Chapter 6** discusses the research presented in this thesis, reflecting on the aims proposed in Chapter 1 and the solutions found to address them. Future research directions are provided along with concluding remarks.

Chapter 2

Background and literature survey

Technical aspects of this thesis and the surrounding literature relating to financial market prediction systems are first discussed; in this section data representations, machine learning models and evaluation metrics are introduced. The literature review that follows focuses on research relating to the design of predictive systems using similar datasets, methods or machine learning algorithms to predict financial quantities or as part of a trading strategy. This survey highlights a gap in the literature both regarding the design of feature mining and predictive systems that extract maximal predictive power from relatively simple, yet intuitive feature sets, and regarding the prediction of target metrics not considered before. The surveyed work demonstrates there is no need to over engineer predictive systems, which results in model complexity explosions, if effective methodologies are developed around feature engineering and signal generation.

2.1 Technical background

Tools, definitions and data used throughout this thesis are introduced here.

2.1.1 OHLC data structures

The use of open, high, low, close (OHLC) data structures is central to this work as it forms the basis of the feature mining methodology introduced in Chapter 3 which is utilised throughout the remaining research. OHLC data is aggregated to represent specific time periods, e.g., 15 minutes. Traditional representations of OHLC data use clock time, although, more recently, OHLC bars have been constructed in volume, trade and quote time. The data structure is also referred to as a *candlestick* (or, if referring to multiple OHLC data structures clustered together, a *candlestick pattern*) as the body and shadow of the data structure look like a candle:

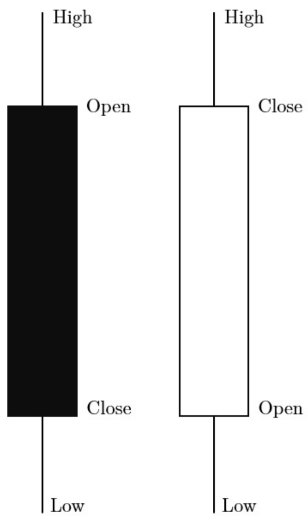


Figure 2.1: OHLC data structures

The region between the open and close prices is called the body. The shadow represents the regions between the high or low in relation to the top or bottom of the body. As shown in this example a black bodied candlestick represents a downward movement whereas a white bodied candlestick represents an upward movement.

2.1.2 Definition of the mid-price

Traditionally the mid-price is defined as the mid-point between the bid and ask prices; this is not the case in this work. Here, two definitions of mid-price are used, and illustrated in Figure 2.2. Mid-price-1 is defined as the mid-point between the high and low of an OHLC structure defined as:

$$mid\_price\_1 = \frac{high + low}{2} \tag{2.1}$$

whereas mid-price-2 focuses on the real body of the OHLC data structure (the region of the candlestick between the open and close prices) and is defined as

$$mid\_price\_2 = \frac{open + close}{2}. \tag{2.2}$$

The predominant reason for investigating the use of a mid-price as a prediction target (as introduced in Chapter 4) is the observation that mid-price time series are far less noisy than the traditional close price series (see Section 4.2.1), making for a more stable trend prediction target.

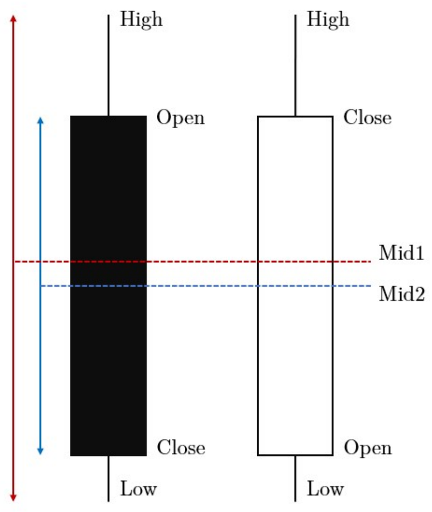


Figure 2.2: Candlestick bars showing the difference between mid-price-1 and mid-price-2 on an OHLC data structure.

2.1.3 Cryptoassets

A key aspect of the research in this thesis is the use of cryptoassets (Bank of England, 2019), of which the best-known example is Bitcoin (Nakamoto, 2008), as the underlying timeseries data. A cryptoasset can be defined in a general sense as a digital representation of value or contractual rights which is cryptographically secured. The asset can be freely traded, transferred or stored. These processes are enabled by distributed ledger and digital wallet technology, the details of which are beyond the scope of this thesis. This work will consider a sub-class of cryptoassets called *cryptocurrencies*. To trade a cryptocurrency such as Bitcoin or Ethereum is much the same as trading a stock, future or traditional currency. Cryptoasset exchanges such as Coinbase, Bitfinex or OKEx provide a professional trading interface from which a user can interact with the market as one would do so in any other market, with the major difference being the increased risk to capital held on an exchange due to the prevalence of hackers.

It is hypothesised here that there may be more inefficiencies in cryptocurrency markets than in traditional markets (such as the FGBL German Bund Futures contract or the EURUSD market) given the prevalence in the cryptocurrency markets of retail traders, fragmented liquidity across global trading venues, ease of manipulation and the relative immaturity of cryptocurrencies as an emerging market. This of course would imply trading in cryptocurrency markets might be easier and more lucrative. However, given the relative immaturity of cryptocurrency markets, it is also hypothesised that statistical artefacts which can easily be exploited can cease to persist more quickly than in traditional markets. This, conversely, would tend of erode the profitability of cryptocurrency trading. Hence, cryptocurrency timeseries datasets provide an interesting timeseries prediction problem, which this thesis will explore further.

It should be noted that as this work is focused on predicting and trading cryptocurrencies on cryptoasset exchanges pegged against the USD the underlying datasets are of the same format and style as traditional financial timeseries data. This makes it easier to design predictive systems which generalise across asset classes (rather than being cryptocurrency specific), as it is hoped that the novel methods introduced in this work will have a wider application.

2.1.4 Machine learning methods

The machine learning models and training techniques used in the work of this thesis will be described from a practical point of view below. For a review of machine learning methods as used in finance see section 2.2.3.

Long Short Term Memory (LSTM) Recurrent Neural Networks (RNNs)

LSTM RNNs (Hochreiter and Schmidhuber, 1997) are a class of neural network designed for sequential timeseries data. The primary issue with vanilla RNNs when applied to such data is the vanishing gradient problem, which makes it difficult to respond to the information from further back in time, as when 'unfolded in time' a vanilla RNN is effectively a deep architecture in which neurons in the layers closest to the 'input' (furthest back in time) fail to receive sufficient weight updates as the partial derivative of the error function is propagated backwards in training. To solve this problem Hochreiter and Schmidhuber (1997) proposed gated cell and memory blocks (which have a series of activation functions encapsulated in these cell architectures) which solved the vanishing gradient problem by retaining both long- and short-term information. There are many subsequent variants of LSTM cells (Gers & Schmidhuber (2000); Cho, et al. (2014); Yao, et al. (2015); Koutnik, et al. (2014)). However Greff, et al. (2015) do a comparison of many different LSTM architectures and find minimal differences in performance. Hence, in this work, the vanilla LSTM cell is used. This cell can be characterised by the following equations:

$$f_t = \sigma\left(W_{xf}x_t + W_{hf}h_{t-1} + b_f\right), \tag{2.3}$$

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i), \tag{2.4}$$

$$\tilde{c}_t = tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c), \tag{2.5}$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t, \tag{2.6}$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o), \tag{2.7}$$

$$h_t = o_t tanh(c_t) \tag{2.8}$$

where $\sigma(\_)$ represents a sigmoid function, W is the weight matrix, x_t is the input, h_{t-1} is the previous input state, b is the bias term and $tanh(\_)$ is a hyperbolic tangent function. To compute the LSTM cells outputs the input variables x_t and h_{t-1} go through a *forget gate* (f_t) first and then an *input gate* (i_t) to decide what information to update. New candidate values $(\tilde{c}_t)$ are then computed which can potentially be added to the cell's internal state. To update the current state (c_t) the old state (c_{t-1}) is multiplied into f_t to forget selected information which is subsequently updated with the new candidate values $(i_t \cdot \tilde{c}_t)$. A sigmoid function is then applied over the cell state to decide what information will be output from the internal state. The cell state is then put through a $tanh$ function to squash values in the range -1 and 1. The result of the $tanh$ operation is multiplied into the sigmoid output producing the output from this gate as represented in equation 2.8. This occurs for each LSTM cell in the network. The weights of these cells are commonly optimised in the usual way using gradient based optimisers, of which a popular choice is ADAM, described below.

Adaptive momentum estimation (ADAM)

ADAM (Kingma and Ba, 2015) is a first order gradient-based adaptive optimisation algorithm popular for training neural networks and designed to combine the benefits of both RMSProp (Hinton, 2014) and AdaGrad (Duchi et al, 2011). The algorithm uses squared gradients to scale the learning rate, like in RMSProp, and uses moving averages of the gradient rather than the raw gradient, like AdaGrad. The weight update rule for ADAM is

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\widehat{v_t}} + \epsilon} \cdot \widehat{m_t} \tag{2.9}$$

where θ represents the model parameters, η is the learning rate, $\widehat{m_t}$ is an estimate of the exponential decaying moving average of past gradients and $\widehat{v_t}$ is an estimate of the exponential decaying moving average of past squared gradients. $\widehat{m_t}$ and $\widehat{v_t}$ are computed as

$$\widehat{m_t} = \frac{m_t}{1 - \beta_1^t}, \tag{2.10}$$

$$\widehat{v_t} = \frac{v_t}{1 - \beta_2^t} \tag{2.11}$$

where the exponential decaying moving average of past squared gradients (v_t) and gradients (m_t) is computed as

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t, \tag{2.12}$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \tag{2.13}$$

with g_t representing the gradient at period t and β the decay rate. Interestingly, as m_t and v_t are initialised with zeros, they suffer from a bias towards zero in the initial epochs, and when decay rates are small. Hence equations 2.12 and 2.13 are used in the update formula (equation 2.9), rather than equations 2.9 and 2.10, to counteract this effect. In this thesis ADAM was used after first conducting empirical tests (as in the example of Figure 2.3) to establish this was, indeed, the best performing optimiser. It should be noted that while Figure 2.3 uses BTCUSD the superiority of ADAM was also demonstrated for EURUSD, ETHUSD, and FGBL.

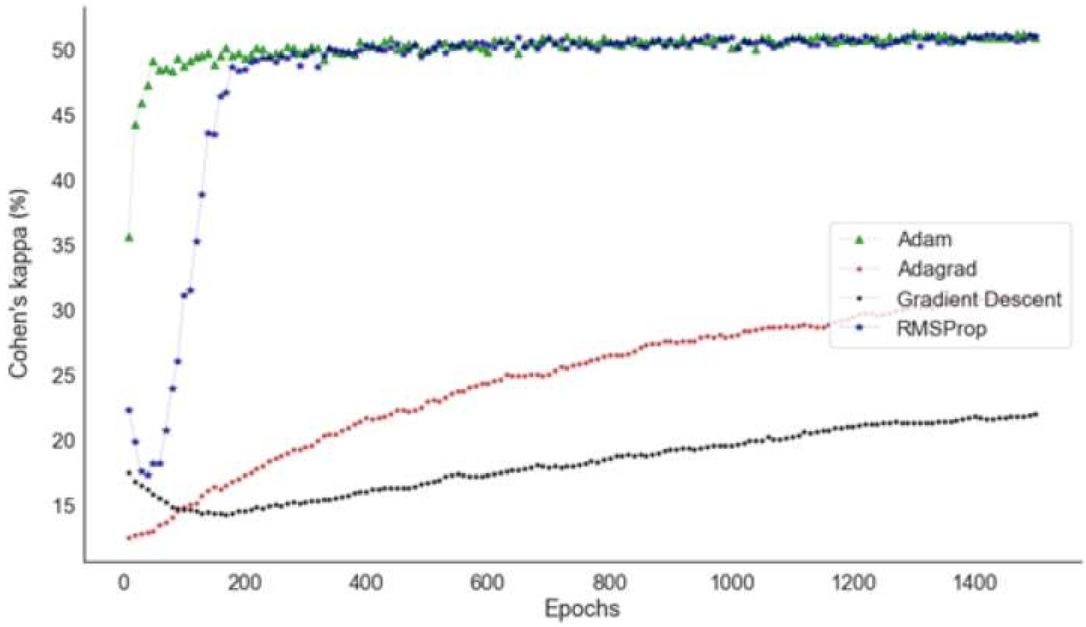


Figure 2.3: An example empirical test of different optimisers using BTCUSD OHLC features.

Batch Normalization (BN)

To accelerate the rate of training and enhance regularization when training neural networks BN (Ioffe and Szegedy, 2015) can be used in the construction of network layers. BN allows for higher learning rates and less attention on the initialisation of parameters in a network by reducing the internal *covariate shift*. The covariate shift refers to the change in the distribution of input values into a learning algorithm which can cause serious issues for machine learning algorithms as they are sensitive to input distribution causing the intermediate layer to continuously adapt to new input distributions. From a theoretical standpoint, BN aims to limit the covariate shift by normalizing the activations of each layer through transforming the inputs to be of mean zero and unit variance. This allows each layer to learn from a more stable distribution of inputs. However, in practice the restriction of zero mean and unit variance can limit the power of the network. To avoid this, BN allows the network to learn a pair of parameters to scale (γ) and shift (β) the activation using mini-batch statistics. This process can be described more formally with the following equations:

$$\mu_{\mathcal{B}} = \frac{1}{m}\sum_{i=1}^{m} x_i \tag{2.14}$$

$$\sigma_{\mathcal{B}}^2 = \frac{1}{m}\sum_{i=1}^{m} (x_i - \mu_{\mathcal{B}})^2 \tag{2.15}$$

$$\hat{x}_i = \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \varepsilon}} \tag{2.16}$$

$$y_i = \gamma \hat{x}_i + \beta \tag{2.17}$$

Equations (2.14) and (2.15) compute the mean and variance of the minibatch, (2.16) normalises the activation x and (2.17) scales and shifts the resulting value. BN is particularly useful for modelling financial datasets, especially cryptocurrency data, where the distribution of inputs from a minibatch can vary significantly as a result of extreme volatility dynamics.

Dropout

To further enhance regularisation and improve generalisation in the training of a neural network *Dropout* (Srivastava et al, 2014) is used. The basic mechanism for Dropout is to temporarily remove, with some probability, a neuron and its incoming and outgoing connections from the network. Dropout can be applied to any neuron in the network, including the inputs, but not the outputs. This mechanism forces nodes in each layer to probabilistically deal with an increase or decrease in information processing responsibility which has the effect of stopping network layers co-adapting. The result is a network which is more robust and, in many cases, one with a lower training error. There has been some discussion of the wisdom of using both BN and Dropout together, as detailed by Li et al. (2018), who suggests that implementing Dropout before BN would cause numerical instability. It is suggested in this work Dropout should be implemented if needed after the BN layer when constructing a network, which is the approach taken here.

Mini-batching

Mini-batching is a halfway house between stochastic gradient descent (pattern-by-pattern weight updating, which can be overly erratic) and batched gradient descent (weight updates performed so as to improve performance over the entire pattern set, which can be slow and prone to trapping in suboptimal solutions). Mini-batch size is an important hyperparameter in the design of modern neural networks. Keskar et al. (2017) present findings to suggest large batch sizes tend to converge to sharp minimizers of training and test functions leading to poor generalisation whereas small batch sizes tend to converge to flat minimizers due to the inherent noise in the gradient estimation. Methods to select optimal batch sizes have focused on dynamic sampling (Byrd et al., 2012; Friedlander & Schmidt, 2012) with little uptake. The default size is often a value of the power of two to fit memory requirements of the GPU or CPU hardware. Masters and Luschi (2018) suggest the minibatch size value should be tuned after hyperparameter optimisation to avoid adverse interactions with the learning rate of a model optimisation. Here, as the optimisation of this value is outside of the scope of this thesis, which focuses on novel input representations and targets rather than the internal working of the predictors, a default value of 128 is used.

Early stopping

Early stopping is when a learning algorithm stops before the final specified epoch, because an error threshold has been achieved or because it has been observed that further training may damage generalisation. Hinton (2015) suggested that early stopping should always be implemented by monitoring the error on a validation dataset with some patience (i.e. after a certain period, if the validation error does not improve training should be terminated). Other methods for early stopping include examining the convergence of the average weight update within the neural network. This work monitors the average weight update and activates a patience when the weight updates tend to zero.

Bayesian Optimisation (BO)

BO uses a probabilistic model, commonly Gaussian Processes (GPs), for modelling an objective function based on previously observed data points (Snoek et al., 2012) which can be defined as

$$\mathrm{P}(f(x)|x) = \aleph\big(\mu(x), \sigma^2(x)\big) \tag{2.18}$$

where $\aleph$ denotes the standard Normal distribution. To estimate $\mu(x)$ and $\sigma^2(x)$ the GPs are fitted to the data using a kernel function (for a rigorous review of GPs and the process of fitting them to data see Rasmussen and Williams, 2006). An acquisition function is required to be optimised over the model. The job of the acquisition function is to control the trade-off between exploration – improving the models understanding of less well explored parts of the search space - and exploitation – favouring parts of the search space which appear to give promising results. This does not require the objective function to be evaluated. In general, exact equations can be derived for this optimisation to find a solution for the optimisation problem using gradient-based optimisation methods. A popular acquisition function is the Upper Confidence Bound (UCB) (Brochu et al., 2010), defined as

$$x_{t+1} = arg\max_{x}\left(\mu_t(x) + \kappa\sigma_t(x)\right) \tag{2.19}$$

where κ is a pre-specified parameter that tunes the trade-off between exploration and exploitation. New, promising, parameter configurations are identified using an acquisition function based on the current model.

The vanilla BO process iterates through the following steps until the maximum budget is reached:

1. Select the point which corresponds to the maximisation of the acquisition function.
2. Evaluate the objective function.
3. Add the observation to the posterior distribution and refit the model.
4. Repeat steps 1-3 until the budget runs out.

As BO requires experience to effectively fit a model it behaves much like random search in early iterations. However with sufficient budget the BO algorithm offers significant speed ups to other methodologies such as random search and grid search.

Bayesian Optimisation with Hyper-Bands (BOHB)

BOHB (Falkner et al., 2018) is a hyperparameter optimisation process capable of efficiently optimising large parameter spaces. BOHB combines the advantages of Bayesian optimisation (BO) (Shahriari et al., 2016) and HyperBand (HB) (Li et al., 2016) into one algorithm. First, HB is used to determine the number of parameter configurations to evaluate and with what budget (a budget represents the total number of evaluations permitted on a function which is usually expensive to compute) and replaces the randomised selection of parameter configurations at the start of each HB iteration using a model-based search which is trained on configurations which were previously evaluated. Once a selection of promising parameter configurations has been identified BO is used to guide the search process in the latter part of the optimisation process allowing the algorithm to refine the parameter configurations selected in the first stage of the search using HBs. The BO part is implemented as a variant of the Tree Parzen Estimator (Bergstra et al., 2011) with a product kernel. BOHB has shown 55x speed ups in hyperparameter searches when compared to vanilla hyperparameter selection processes such as random search.

Extreme Gradient Boosted Trees (XGBoost)

In this thesis XGBoost is used to reduce the dimensionality of a large OHLC based feature universe by removing features which are not important in the prediction of a specific target. XGBoost (Chen, 2016) is a supervised learning tree-based algorithm in which the objective function (loss function and regularization) to minimise at iteration t is specified as

$$\mathcal{L}^t = \sum_{i=1}^{n} \iota\left(y_i, \hat{y}_i^{t-1} + f_t(x_i)\right) + \Omega(f_t) \tag{2.20}$$

where $\mathcal{L}$ is a standard function of Classification and Regression Tree (CART) learners unable to be optimised using optimisation methods in a Euclidean space (Chen, 2016), y_i is the target variable, $\hat{y}_i^t$ represents the prediction, $f_t(x_i)$ is the function with input features that most improves the model and Ω is a penalization function.

To optimise the objective function of (2.20) second order approximation using Taylor's theorem (see Morris (1972) for details) can be used to transform it into the Euclidean domain for optimisation. The quality of a tree structure q can be computed as

$$\tilde{\mathcal{L}}^t(q) = -\frac{1}{2}\sum_{j=1}^{T}\frac{\left(\sum_{i\in I_j} g_i\right)^2}{\sum_{i\in I_j} h_j + \lambda} + \gamma T \tag{2.21}$$

where $g_i = \partial_{\hat{y}^{(t-1)}} \iota(y_i, \hat{y}^{t-1})$ and $h_i = \partial^2_{\hat{y}^{(t-1)}} \iota(y_i, \hat{y}^{t-1})$ result from the second order approximation, q represents the tree structure in question and $I_j = \{i \mid q(x_i) = j\}$ is the instance set of leaf j. This formula returns the minimal loss value for a given tree structure allowing the original loss function to be evaluated using the optimal weight values. In practice a learner is built by: (1) initialising a single root with all training examples; (2) iterating over the features and values per feature while evaluating each split by computing the gain = loss(father instances) – (loss(left branch) + loss(right branch)); and (3) testing if the gain for the next split is negative, in which case stop growing that tree branch, otherwise continue. This "exact greedy algorithm" runs in $O(n * m)$ where n represents the number of training samples and m represents the number of features. Notably, XGBoost can be used to assess the importance of each feature to its target using the gain described above, an advantage of this method that will be used in chapters 3, 4 and 5.

k-means++

In this work k-means++ is used to cluster important OHLC based features for validating predictive properties. k-means++ (Arthur and Vassilvitskii, 2007) is an unsupervised learning algorithm with the objective of grouping similar datapoints based on some similarity metric. The algorithm is an enhancement of k-means with the main difference being an additional process to set the initial seedings. The algorithm can be described as follows:

1. Select an arbitrary set of cluster centroids $C = \{c_1, \ldots, c_k\}$.
2. Uniformly at random select an initial center c_1 from the feature set χ.

3. Select the next center $c_i = x' \in \chi$ with a probability $\frac{D(x')^2}{\sum_{x\in\chi} D(x)^2}$ where $D(x)$ denotes the shortest distance from a data point x to its closest center.
4. Repeat step 3 until all k centers have been selected.
5. For each $i \in \{1, \dots, k\}$ set c_i to be the set of points in χ which are closer to c_i than $c_j \ \forall\, j \neq i$.
6. For each $i \in \{1, \dots, k\}$ set c_i to be the center of all points in C_i: $c_i = \frac{1}{|C_i|}\sum_{x\in C_i} x$.
7. Repeat stages 5 and 6 until C does not change.

Steps 1-3 are the additional steps implemented by k-means++; vanilla k-means arbitrarily selects k centers.

t-Distributed Stochastic Neighbour Embedding (t-SNE)

In this thesis t-SNE is used to verify that clear structural partitions are discovered when applying the k-means++ algorithm to OHLC features as described above. t-SNE (Maaten and Hinton, 2008) is a non-linear dimensionality reduction technique which is well suited for visualisation of high dimensional datasets. The process for reducing dimensionality can be summarised at a high level as:

1. Create a Gaussian probability distribution in the high-dimensionality space which dictates the relationships between neighbouring points.
2. Create a Student t-distribution with a single degree of freedom (also known as the Cauchy distribution) to recreate the probability distribution of the high-dimensional space in the low-dimensional space. This prevents the "crowding problem" which is common in low-dimensional space where points get too crowded due to the curse of dimensionality. The Student t-distribution is selected due to its longer tails compared to a Gaussian distribution which helps alleviate this problem.

3. The low-dimensional probability mapping is then optimised by applying gradient descent with momentum on the KL-divergence (a measure of how two probability distributions are different, sometimes referred to as relative entropy) between the high and low probability density functions.

Following the above process results in a low-dimensional visualisation of a high dimensional space.

2.1.5 Performance evaluation metrics

The assessment of predictive classification power will be central to this thesis. It will be especially important to account for imbalanced datasets. Many of the metrics used are constructed from a confusion matrix structure. For a binary classification problem, a confusion matrix is constructed as in Table 2.1:

Table 2.1: Confusion matrix

	Predicted Up Price Movement	Predicted Down Price Movement
Actual Up Price Movement	True Positives (TP)	False Negatives (FN) [Type II Error]
Actual Down Price Movement	False Positives (FP) [Type I Error]	True Negatives (TN)

where

- TP: The number of predictions where an up prediction was made and the resulting price movement was up,
- TN: The number of predictions where a down prediction was made and the resulting price movement was down,

- FN: The number of predictions where a down prediction was made and the resulting price movement was up,
- FP: The number of predictions where an up prediction was made, and the resulting price movement was down.

(The confusion matrix can be generalised to a multi-class classification problem by extending the columns and rows to equal the number of classes in the classification problem, but this will not be needed for the work of this thesis.)

A number of performance metrics that can be computed from the confusion matrix will now be listed and defined, before moving to other metrics with different, statistical, bases for their definition.

Simple Accuracy is computed as the proportion of correctly predicted directional movements. It can be formally defined as:

$$simple\_accuracy = \frac{TP + TN}{TP + FP + FN + TN}. \qquad (2.22)$$

It has the advantage of simplicity but the weakness of being an unreliable indicator of performance in a strongly trending market, where there may be a tendency to overpredict the majority class.

Cohen's kappa, sometimes referred to as the *normalised percentage better than random* (NPBR) is a more robust performance metric for imbalanced datasets when compared to a simple accuracy measure, with a range of -100% to 100%, a score of 0% being equivalent to chance. The metric is defined below:

$$t = TP + FP + FN + TN, \qquad (2.23)$$

$$R_{total} = \frac{(TP + FP)(TP + FN) + (TN + FP)(TN + FN)}{t}, \qquad (2.24)$$

$$Cohen's\ Kappa = \frac{(TP\ +\ TN) - R_{total}}{t - R_{total}}. \tag{2.25}$$

This measure allows a comparison against random, which is a valuable metric to state as it gives relative contextual information.

Matthews Correlation Coefficient (MCC) is a complementary metric to record alongside Cohen's kappa. The metric is a correlation coefficient between the observed and predicted classifications which is bounded between -1 to 1, with a perfect classification receiving a score of 1. Notable, a result of 0—the same as for a random prediction—is obtained if all examples are assigned to the same class, making the MCC a tough metric for the prediction of imbalanced datasets. The MCC metric can be specified for a binary classification problem as

$$mcc = \frac{TP \cdot T\mathrm{N} - FP \cdot FN}{[(TP + FP)(TP + FN)(TN + FP)(TN + FN)]^{0.5}} \tag{2.26}$$

and is often presented as a percentage value by multiplying the value by 100. The MCC can further be generalised for multi-class classification problems. However, as binary classification (such as market up/down) is the focus of this thesis the multi-classification details are emitted.

Precision, sometimes also referred to as the *positive predictive value*, is the number of true positives divided by the sum of true positives and false positives. It is the measure of a classifier's exactness and is often low if the model's predictions resulted in many false positives. The Precision metric is given by:

$$precision = \frac{TP}{TP + FP}. \tag{2.27}$$

Recall, which is sometimes referred to as *model sensitivity*, is defined as the ratio of correct predicted positive labels (TPs) to all observations of that class. It can be defined as:

$$recall = \frac{TP}{TP + FN}. \tag{2.28}$$

The **F1 Score** is the weighted average of recall and precision meaning the metric considers both false positives and negatives in its computation. Although the metric is less transparent than simple accuracy it is often more useful to understand classification power when datasets are imbalanced (such as the ones in finance markets). The F1 score can be defined as:

$$F1 = 2 \cdot \frac{Recall \cdot Precision}{Recall \; + \; Precision}. \tag{2.29}$$

Finally in this group of metrics derived from the confusion matrix, the **Support** is often recorded when presenting the F1 score, Precision and Recall metrics and is defined as the number of true responses per class.

Aside from the above classification metrics, based on the confusion matrix, performance may also be measured using an exact binomial test.

A **Binomial test** is formulated as

$$P(X \mid n, p) = \frac{n!}{(n \; - \; X)! \; X!} \; \cdot \; (p)^{X} \; \cdot \; (q)^{n \, - \, X} \tag{2.30}$$

where

- p is the hypothetical probability of a correct directional classification;
- q is equal to $(1 \; - \; p)$;
- n is the total number of trials;
- X is the number of correctly classified predictions.

The result of plugging in the above variables is a p-value which can be rejected at different confidence intervals depending on the problem. Usually a 5% confidence interval is appropriate. This statistical test indicates if a pattern's predictive power significantly differs from an expected predictive ability. In this thesis the greater one-sided tail is used following the approach presented in Lu et al. (2012) who also tests the significance of several candlestick patterns predictive value.

In this work the assessment of a clustering algorithm's effectiveness, for the purpose of selecting the optimal number of clusters k, is done using the **Silhouette coefficient** (SC) (Rousseeuw, 1987). The SC represents how similar a datapoint is to its assigned cluster (measuring cohesion) compared to other clusters (measuring separation). Usually the mean SC is reported representing how well the data is clustered for k clusters. The optimal value of k is the one that maximises SC. The metric can be formulised for each data point $i \in C_i$ as

$$a(i) = \frac{1}{|C_i| - 1} \sum_{j \in C_i, i \neq j} d(i,j) \tag{2.31}$$

$$b(i) = \min_{k \neq i} \frac{1}{|C_k|} \sum_{j \in C_k} d(i,j) \tag{2.32}$$

$$s(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}}, if\ |C_i| > 1 \tag{2.33}$$

where $a(i)$ is the mean intra-cluster distance, $b(i)$ is the mean nearest-cluster distance, C_i is the cluster in question and C_k is the total number of clusters equal to k. SC is bounded between -1 and 1 with higher values representing more effectively clustered datapoints (i.e. the cluster is denser as a result of data points being very similar).

As is evident from the discussion of above performance evaluation metrics, assumptions are made in many of them which can potentially distort the interpretation of performance, making it important to use a variety of statistics to confirm the validity of a conclusion, especially in financial data where signal to noise ratios are low and false discoveries are easily recorded.

2.2 Related Work

This section presents a review of research related to the work of this thesis, covering both traditional and advanced methodologies for financial prediction. Special attention is given to the use of OHLC data structures (candlesticks) in designing predictive systems as there has been conflict in the literature about their usefulness.

2.2.1 Candlestick patterns

Japanese candlestick patterns are one of the oldest forms of pattern recognition techniques used in the attempt to predict a market's directional movement and are a controversial topic in academia. They were first proposed by Munehisa Homma around 1750 for charting the price behaviour of rice markets. Since then a debate as to whether candlestick patterns do indeed provide predictive power has been ongoing for decades amongst industry practitioners and academics, with support from the financial industry opposed by academic scepticism. The adoption of candlestick patterns has been widespread among both professional and retail traders with books dedicated to the study of candlestick patterns (Morris, 1992;1995; Bigalow, 2011). Often these patterns are constructed over multiple lags of OHLC bars to create a pattern which considers historical information to predict future trends.

However, as mentioned above, academic studies have been predominantly critical of the reputed predictive power of candlestick patterns. Marshall et al. (2006; 2008) find that the relationships between OHLC levels have no useful information when applied to stocks in the Dow Jones Industrial Average (DJIA). Horton (2009) confirms there is little to no value in candlestick charting. Fock et al. (2005) present negative results for both the DAX stock index and the FGBL

German Bund futures contract (FGBL). The latter result is noteworthy on the context of the work of this thesis, for which predictive power is demonstrated for the FGBL, though it should be noted that the OHLC-derived patterns of the thesis are not traditional candlesticks but data-mined constructions. A further analysis on the DJIA by Duvinage et al. (2013) assesses the profitability of candlestick patterns on an intra-day 5-minute aggregation across the 30 constituents of the index. They too report a negative result, finding that after transaction costs the predictive power encapsulated in the 83 patterns they study is too weak to be profitable.

An interesting approach is taken by Detollenaere and Mazza (2014) who demonstrate that execution costs can be significantly reduced by timing the market correctly using candlestick patterns as a signal as to when to execute trades, agreeing with Mazza (2015), where it was demonstrated the occurrence of a Doji pattern was correlated with higher liquidity. However, Detollenaere and Mazza overall concur with the negative results of Marshall et al. (2006; 2008) and Duvinage et al (2013), showing that candlestick patterns are not able to predict future returns of 81 European stocks.

However not all academic studies are critical of the use of candlestick charting. On the positive side, Xie et al. (2012) find that candlestick patterns have significant predictive power to forecast US equity returns. Lu et al. (2015) find predictive power in several patterns, but these are rare and the research in addition did not sufficiently address the distinction between candlestick patterns being able to yield profit and their being able to predict trends. They did however make the observation that when considering more volatile markets evidence in favour of candlestick trading strategies strengthened. Chen et al. (2016) assesses the predictive power of four pairs of popular candlestick patterns (bullish and bearish versions of the same patterns), aggregated on a daily frequency for the

Chinese markets, finding five of the eight candlestick patterns to have significant predictive power over medium and large market capitalisation stocks, although predictive power decayed rapidly as the predictive horizon is increased. One study, that of Lu (2014), finds that traditional patterns have little value but that novel patterns may do so when applied to the Taiwan stock market. This result is in line with that of Prado (2013), where it is shown that though traditional candlestick patterns on Brazilian markets do not show statistically significant power, significant results can be obtained when the patterns are finetuned for a specific market; this finding is in line with discoveries, to be presented later, of this thesis, though it should again be emphasised that the application of a novel feature mining process in the work of the thesis creates a very different context.

Overall the evidence in the literature favours the dominant academic belief that candlestick patterns have little predictive value. However, along with those results that suggest that novel, or asset-tuned, patterns may be predictive, there is also evidence that combining OHLC data structure with other more advanced techniques such as fuzzy logic (Lan et al. 2011; Tsai, 2014; Naranjo, 2018) and machine learning (see section 2.2.3) may achieve statistically significant predictive systems. Thus the results presented in this thesis, albeit in the context of a mid-price prediction and/or utilising advanced machine learning methods, while they may surprise academic researchers, are at least to some degree presaged by these few hints of the possible value of modified and/or augmented, non-traditional, forms of candlestick charting.

2.2.2 Technical analysis

As was the case with candlestick patterns the broader technical analysis field is a highly controversial area in academia, with academics largely doubting the usefulness of these tools and the statistical significance of claimed results.

Technical indicators (e.g. oscillators and averages typically used to infer trend strength, future market direction and oversold/overbought market state) in many cases rely heavily on lagged OHLC price structures (where a *lagged price* is one measured at a previous (relative to the current) time step), making this form of analysis to a substantial degree an extension of the ancient Japanese candlestick approach.

On the positive side, Brock et al. (1992) measured the performance of technical analysis by applying a bootstrap methodology to the DJIA index, finding that trading range breaks and moving averages generate statistically significant returns when compared to four different benchmark models. Brown and Jennings (1989) confirm the usefulness of technical analysis, but only where prices are not fully informative and when market participants do not make rational decisions about the relationship between signals and price evolution. Kavajecz and Odders-White (2004) also confirm the usefulness of technical analysis by looking at support and resistance levels coinciding with order book depth concluding that moving averages have predictive power in relation to orderbook depth. Interestingly, Taylor and Allen (1992) surveyed senior London-based foreign exchange dealers finding a bias towards the use of technical analysis for shorter term price analysis compared to the use of fundamental data. Lui and Mole (1998) carry out a similar study in Hong Kong, with the same outcome as the work of Taylor and Allen. This raises the question of whether technical analysis is in a sense self-fulfilling, so that market movements are simply an artefact of popular technical indicators. According to Schulmeister (2007) the presence of chartists in the market would indeed result in technical analysis being a self-fulfilling prophecy: markets with a high use of technical analysis would see significant results due to the number of market participants relying on the same toolset.

Turning now to research that shows mixed results as to the value of technical analysis, an interesting study was published by Neftci (1991) which shows some instances of technical analysis can produce successful forecasting rules if the underlying timeseries process does not conform to a Gaussian distribution, although the general theme here was that technical analysis did not produce statistically significant predictions. Hao et al. (2013) investigate the profitability of technical trading rules such as trading range breakout (based off OHLC differences) and find these rules are more effective in emerging markets, and that shorter-term (minutes to hours) variants provide significantly more predictive power than longer term (over days) versions of the same indicator. However they also suggest that the chosen technical indicators would provide only weak predictive power in more efficient markets.

On the negative side, LeBaron (1999) investigated technical analysis over periods during which the Federal Reserve was inactive, showing that the ability to predict exchange rates using TA is significantly reduced. Park and Irwin (2007) conduct a study of the literature on the profitability of technical analysis, finding further evidence of poor predictive power when predictive systems are built from technical indicators alone. Menkhoff et al. (2007) acknowledge technical analysis is widespread but conclude the evidence that this form of predictive analytics holds any significant power is unconvincing.

Overall, then, there are mixed academic opinions as to the effectiveness of technical analysis, although use of technical indicators derived from an OHLC data structure is popular with practitioners. As was observed also in the section of the literature survey dealing with candlestick patterns, TA methods have been improved by applying more advanced methods such as genetic programming and neural networks to enhance the traditional or basic rules (Neely et al. (1997); Neely and Weller, 2001; Fernandez-Rodryguez et al. 2000). Many of the studies reviewed above suggest the use of technical analysis can be effective on certain markets but its usefulness is highly correlated with the level of market efficiency. More advanced methods are required in efficient markets to extract meaningful predictive power given the decay of observable inefficiencies over time. This would be consistent with Olson (2004), in which it is stated that the profitability of technical trading rules has indeed reduced over time.

2.2.3 Machine learning for prediction systems and trading

As has been discussed above there is some degree of evidence that OHLC data structures can be a basis for predictive systems, historically, first through candlestick patterns, and later as a component of many popular technical indicators. More recently predictive systems have made heavy use of machine learning techniques, as markets have become more efficient and extracting significant predictive power has become harder. Interestingly, although the complexity of predictive systems has increased, the use of OHLC data structures remains popular.

Hu et al. (2019) propose an interesting methodology by visualising the FTSE100 market as OHLC bars and then applying a deep convolutional autoencoder (CAE) to extract feature information from this dataset. Using the feature representations from the CAE they cluster the training data and select stocks with the highest Sharpe ratio (Sharpe, 1966) to construct a portfolio. Their portfolio was able to outperform the FTSE100 and competing investment vehicles by more than double. Other examples of using candlesticks in combination with deep learning is the work of Kim, T. and Kim H.Y. (2019) who propose a feature fusion LSTM-CNN combining features learned from different stock timeseries data and candlestick images to predict the SPDR S&P500 ETF price. They achieve a RMSE of 0.098 for predicting the five-minute forward return and successfully devise several trading strategies with the best generating profits in excess of 17% over a three-month period. They also discover that a candlestick chart is the most appropriate type of visualisation for the prediction of financial markets which is in line with the results of this thesis. The methodology of (Kim and Kim, 2019) shows an interesting approach to the use of candlestick patterns in the 21st century for extraction of predictive power.

Moving away now from the above approach of transforming OHLC data into images and applying CNNs to these image corpuses, Jasemi at al. (2011) combine Japanese candlestick analysis with a feed-forward neural network (ANN) using two different approaches. The first approach focuses on raw data, defined as ratios of OHL price levels to the bars close price, as network inputs. The second approach uses the candlestick representation as inputs to the network. Both the first and second approach had the same objective of predicting the 6-day trend, classified into three categories (ascent, neutral or decent). The authors find the signal approach (i.e. using OHLC representations of a candlestick) to be the most powerful, with accuracy scores into the 70% region (although the work does not account for dataset imbalance and so one should be cautious about the significance of this reported result). Nevertheless, this research demonstrates a successful hybrid combination of machine learning and candlestick patterns.

The use of OHLC differences, ratios and/or returns as features for machine learning algorithms has been a focus for many academics and practitioners, using a wide variety of learning algorithms from support vector machines (SVMs) (Prasaddas and Padhy, 2012; Huang, 2002; Li and Tam, 2017), LSTM RNNs (Rather et al., 2015; Bao et al., 2017; Chen et al., 2015; Gao et al., 2017; Gers et al., 2002; Khare et al., 2017; Li and Tam, 2017; Nelson et al., 2017, Zeng and Liu, 2018, Troiano, 2018), agent based approaches (Raudys and Zliobaite, 2006), deep belief networks (DBNs) (Shen et al., 2015), previously mentioned CNNs (Ding et al. 2015; Kim, T. and Kim H.Y., 2019) and ANNs (Refenes et al., 1994; Garliauskas, 1999; Fernandez-Rodrıguez, 2000; Yu, 2015). It is clear the use of advanced machine learning techniques is popular in designing predictive systems, which could be partly due to the need for more advanced techniques in extraction of predictive power as markets become more efficient.

One might give a particular emphasis here to the machine learning work of Sermpinis et al. (2015; 2013) and Duns et al. (2011), as the authors hold academic tenure and also actively run a fund where they presumably apply the techniques, or derivative versions of the techniques, proposed in their published works. An interesting approach to forecast foreign exchange rates is presented by Duns et al. (2011) and separately by Sermpinis et al. (2015; 2013). They focus on applying ANNs, SVMs and DBNs to a wide range of instruments achieving a maximum directional accuracy of 63.62%. Though one should again be cautious of directional prediction results reported solely in terms of accuracy, it is notable that when combined with simple trading strategies these predictions generate a maximum annualized return of up to 30% on the EURUSD currency market (a market this thesis also focuses on).

So far, the literature reviewed has focused on the use of single learning algorithms to predict, usually, the close price. Ensemble learning is the practice of combining many models' predictions into one with the aim of achieving a more stable model with enhanced predictive power. In Chapter 5 an ensemble classifier is proposed using multiple LSTM RNNs to predict a directional trend. Therefore, a review of interesting ensemble methods for financial prediction is appropriate. Cheng et al. (2012) conducts a comparison of ensemble methods in financial market prediction, finding bagging to be the most effective method when compared to random subspace and stacking, although he does note that the selection of ensemble algorithm is highly dependent on the desired usage. Picasso et al. (2018) combines technical analysis and machine learning together as an ensemble to predict market trend for the selection of a portfolio of NASDAQ stocks, finding the method to achieve up to 61.69% accuracy. This is impressive (given the usual caveats about reporting in terms of accuracy) given the efficiency of the market they focused on. A number of other works have focused on combining neural networks and other learning algorithms to achieve a higher, more robust predictive power (Schwaerzel and Rosen, 1997; Al-Hnaity et al, 2015; Gyamerah et al., 2019; Hegde et al, 2018; Yang el al., 2016; de Mello Assis et al., 2018; Xu et al., 2014; Yang el al., 2017; Yu et al., 2008). All have observed an increase in stability and enhanced robustness in out-of-sample datasets when compared to single predictors. Finally, in relation to the use of ensembles, it should be noted that the use in Chapter 5 of an ensemble of LSTM RNNs trained over different intra-day OHLC aggregations, each predicting a different horizon, is novel in previous work, to the author's knowledge, has not investigated the combination of different models which predict over differing time horizons and have differing frequency aggregations as features.

This thesis focuses on a range of asset classes with one being that of cryptocurrencies. The literature review will therefore turn now to the use of machine learning techniques for forecasting cryptocurrency price movements. Attanasio et al. (2019) explores the use of machine leaning techniques for quantitative cryptocurrency trading, finding that prediction models based on time series forecasting techniques perform better than classification models, which agrees with the work to be presented in this thesis. McNally et al. (2018) applies a Bayesian optimised LSTM RNN to predict the Bitcoin price direction achieving a directional accuracy of 52%. This appears somewhat disappointing when compared to other (Chowdhury et al., 2019; Mallqui and Fernandes, 2019; Sin and Wang, 2017) published work but demonstrates the difficulty of classifying cryptocurrency directional price movements.

Ensemble learning, used in this thesis, has also seen past success in the directional prediction of cryptocurrency prices. Chowdhury et al. (2019) investigate the use of machine learning (gradient boosted trees, ANNs, K-NN, ensembles) to predict future prices of the CCi30 cryptocurrency index and its constitutes. They find a maximum performance using ensemble learning (though the work appears flawed in that only in-sample results are reported, albeit with accuracies of 95%). An interesting approach is taken by Mallqui and Fernandes (2019) who predict the direction, maximum, minimum and close prices of the daily Bitcoin exchange rate using an ensemble of recurrent neural networks and tree-based algorithms. They find significant predictive power with this approach achieving a directional accuracy of 62.91%. Sin and Wang (2017) design an ANN ensemble called a "genetic algorithm based selective neural network ensemble" to predict the price of Bitcoin, showing an impressive 64% accuracy. A trading strategy is then constructed achieving 88% returns over a period of two months (although they failed to outperform a Buy & Hold strategy). Other examples of machine learning trading strategies focusing on cryptocurrency markets (Shah and Kang, 2014; Fischer et al., 2019) show similar results where a better risk adjusted return is achieved at the cost of lower returns when compared to a Buy & Hold strategy. (In this thesis it will be shown that a methodology using ensembles of LSTM RNNs is shown to outperform Buy & Hold strategies while maintaining a high Sharpe ratio.) Other notable works on the prediction of cryptocurrency prices—using methods outside the scope of this thesis—include Georgoula et al. (2015) and Steinert and Herff (2018), who use social media features, and Kristoufek (2015) and Phillips and Gorse (2018), who use wavelet coherence applied to social media datasets.

As has been mentioned previously, in this thesis (as will be shown in Chapter 4) a new mid-price derived from the structure of an OHLC bar is shown to be highly predictive. Notably there exists no prior literature by other academics or practitioners relating to a mid-price as proposed here.

Overall, on analysing the work discussed in this section, it appears higher frequency price and trend movements (e.g. intra-day movements) are more difficult to predict than those at a lower frequency (e.g. daily data). This may simply be an artefact of a large difference in the number of data points, where papers presenting work at a lower frequency may have discovered false artefacts, although one can only speculate. The level of market efficiency is also an important consideration as it appears that more impressive results have been achieved for emerging markets or markets with a high level of retail traders (e.g. Indonesia, Malaysia, China, Cryptocurrencies) compared to mature markets. Finally, the speed of market movements and liquidity fragmentation has an impact on how effectively predictive power can be extracted from a market due to patterns diminishing or completely breaking over a short period of time, usually because of a permanent change in market structure.

2.2.4 Conclusions of the literature survey

The evolution of systems for prediction of financial markets has continually increased in complexity, with a high correlation to increasing market efficiency, from the use of candlestick patterns in the 18th century to the widespread adoption of technical analysis, the use of advanced machine learning techniques, and hybrids between these most ancient and most modern approaches in the 21st century.

Evidence has been presented for (Xie et al., 2012; Chen et al., 2016; Lu, 2014; Prado, 2013) and against (Marshall et al., (2006; 2008); Horton, 2009; Fock et al., 2005; Duvinage et al., 2013; Detollenaere and Mazza, 2014) the use of candlestick patterns and technical indicators in the construction of financial predictive systems. A review by Park (2004) shows currency markets are best suited to technical analysis with futures being classed as an intermediate fit and stock markets being the poorest fit for this type of predictive system. Over time markets have become more efficient from decreasing transaction costs, “big data” analysis, and more advanced market participants, which Park suggests has negatively affected the predictive capabilities of technical analysis. The decline in the effectiveness of technical analysis is further emphasised by Neely (2009), evidencing the need for more advanced predictive systems which can generalise across asset classes. This is one of the main focuses of this thesis. The work presented in this thesis focuses on currencies and futures markets, as would be recommended by Park, albeit two of the three currency markets investigated are that of cryptocurrencies which were not invented when Park published his work. Insight into the effectiveness of candlesticks and OHLC data in cryptocurrency markets will therefore provide further insight into this area.

The increased quantity of published work using machine learning systems for financial market prediction (Prasaddas and Padhy, 2012; Li and Tam, 2017; Rather et al., 2015; Bao et al., 2017; Chen et al., 2015; Gao et al., 2017; Khare et al., 2017; Li and Tam, 2017; Nelson et al., 2017, Zeng and Liu, 2018, Troiano, 2018; Hu et al. 2019) is highly correlated with fewer works being published relating to candlestick patterns and technical analysis, likely due to market efficiencies, as discussed previously. However, the literature suggests that the use of OHLC data structures has continued to be of significant interest evidencing it to be a particularly powerful representation of time series data. This thesis will focus completely on the use of OHLC data, and the results to be presented will be in agreement with the observations of previous researchers that this data representation is indeed powerful.

One of the dangers of a machine learning predictive system is that of *overfitting* (when an algorithm is fit to specific artefacts of its training dataset, resulting in a model which cannot generalise well to out-of-sample data) and complexity explosions. This appears to be widespread in the literature given the number of papers stacking machine leaning algorithms together and procuring feature sets with complex transforms in their methodologies to achieve what is reported as reasonable predictive power. López de Prado (2018) emphasises the danger of overfitting and widespread false discoveries published in financial literature, albeit in the context of trading strategy parameters. It is evidenced that financial datasets are a difficult area in which to develop robust and generalisable systems as their governing dynamics continually change, hence the need for powerful systems which in addition maintain transparency and interpretability.

* * *

This review section will end by considering the relations between the systems developed in this thesis, heavily based on machine learning and OHLC data structures, and those systems evidenced as promising in the literature. The novel predictive methodologies proposed here differ from the past literature in a number of important ways:

1) The systems developed here focus on transparency and simplicity in the feature space while still achieving, and in many cases significantly outperforming, published works on the prediction of financial markets.

2) Consideration is given here to the sensitivities of each algorithm while still ensuring maximal predictive power can be extracted from a feature set.
3) The possibly misleading use of accuracy alone as a performance metric is here avoided, and statistical testing of predictive power is carried out, showing the results presented in this thesis to be significant.
4) No previous work has constructed ensembles by combining sub-models trained to predict different horizons using differing feature aggregation periods which is somewhat novel and achieves remarkable predictive power.
5) No previous work has developed predictive systems effective on both traditional and crypto asset classes, as will be done in this thesis.

With a focus on transparency, simplicity and interpretability this thesis provides insight in how to build prediction systems which can generalise, control for complexity, and operate at different data granularities while providing robust predictive power.

Chapter 3

Deep candlestick mining

The goal of this chapter is to predict the direction of the close price (up or down over the next period, here one hour) using as input novel candlestick patterns derived a process of feature mining within a large universe of potentially useful OHLC constructions (using ratios and difference of price levels). As discussed in the previous chapter, there are conflicts in the academic literature about the usefulness of candlestick patterns to predict directional price movements. However, these discussions in the literature are about traditionally defined candlestick patterns only. In this chapter, a data mining process named *Deep Candlestick Mining* (DCM) is developed using Extreme Gradient Boosted Trees, Long Short Term Memory Recurrent Neural Networks and k-means++, and is shown to be able to discover candlestick patterns significantly outperforming traditional ones. A test for the predictive ability of novel versus traditional candlestick patterns is devised using all significant candlestick patterns within the traditional or deep mined categories. The deep mined candlestick system demonstrates a remarkable ability to outperform the traditional system by up to 117% on hourly data.

3.1 Introduction

Japanese candlesticks are one of the oldest forms of pattern recognition techniques used to attempt to predict financial markets. The methodology was first proposed by Munehisa Homma around 1750 for charting the price behaviour of rice markets. Candlestick charts visualise an asset's price by aggregating period specific bars (e.g., 1-hour bars) consisting of open, high, low and close (OHLC) price levels. Formations of frequently identified sequential patterns are then used as a tool to predict future market direction. Many industry practitioners believe candlestick patterns are an effective predictive tool, though there has been much debate in the academic world as to their effectiveness, as discussed in section 2.2.1.

In this chapter a new process for discovering candlestick pattern formations is proposed, referred to as *Deep Candlestick Mining* (DCM), as a means to discover asset-specific predictive candlestick patterns using a number of ML techniques: Extreme Gradient Boosted Trees (XGBoost), Long Short Term Memory Recurrent Neural Networks (LSTM RNNs) and k-means++ (for technical model details see section 2.1.4). A feature mining process is first proposed using XGBoost and correlation filtering showing that a high predictive power can be extracted from financial markets using relatively simple and intuitive feature representations. This feature mining process is also used in chapters 4 and 5 due to its effectiveness in discovering powerful features while maintaining transparency in the feature space.

In the work of this chapter DCM-based prediction is shown to substantially outperform the use of traditional candlestick patterns on hourly data, over a range of different instruments, showing the methodology is robust to asset class. The technique additionally allows a visualisation of the candlestick patterns; in this way, DCM-based patterns allow a human level understanding, that might be especially appealing to industry practitioners used to the deployment of traditional candlestick patterns, of what black-box predictors have deemed to be important relationships discovered in financial market datasets using OHLC derived features.

3.2 Methodology: DCM Workflow

This section tabulates the workflow for the DCM process. Each stage after the first, in which the data are generated, is carried on a different dataset, in order that there is no unwanted data leakage and the resulting model can generalise well.

Table 3.1: Deep Candlestick Mining's process flow.

Process & Dataset	**Details**
Data Generation (all data)	1. Gather raw transaction data (section 3.3). 2. Aggregate this into hourly OHLC data.
Feature Mining (Dataset 1)	1. Identify a powerful set of features using XGBoost and a correlation filtering process (section 3.4.1). 2. For each feature set train LSTM RNN models (section 3.4.2). 3. Select the feature set which maximises the network's performance in validation (section 3.4.2).
Candlestick Generation (Dataset 2)	1. Cluster the feature set to identify global and local structure (section 3.5.1). 2. Validate each cluster's structure is real (section 3.5.1). 3. Drop clusters that do not satisfy cluster desiderata (section 3.5.2). 4. Index each candidate to the original OHLC timeseries and visualise its candlestick pattern (section 3.5.2). 5. Create a *Deep Candlestick Mining* pattern to represent the average shape of a cluster's candlestick patterns (i.e., the centroid) (section 3.5.2).
Test (Dataset 3)	1. Test traditional candlestick patterns individually (section 3.6.1) 2. Test DCM-based patterns individually (section 3.6.2).
Holdout (Dataset 4)	Compare a predictive system based on multiple DCM-based candlestick patterns to one based on multiple traditional candlestick patterns (section 3.7).

3.3 Data selection and usage

The work of this chapter concentrates on four different instruments, the German bund 10-year futures contract (FGBL), EURUSD, Bitcoin (BTCUSD) and

Ethereum (ETHUSD). As the cryptocurrency asset class is much newer than FGBL or EURUSD the datasets acquired from the Bitfinex exchange for BTCUSD and ETHUSD do not go back as far historically. An effort is made, where possible, to align the historical duration. As the validation, test and holdout datasets are important for candlestick discovery and out-of-sample comparisons these are each of the same duration, and hence each market would be exposed to roughly the same global economic environment. The train dataset's historical duration is aligned per asset class categorisation (traditional and cryptoasset). The traditional asset classes use more train data as there was enough data acquired to do so. Results were observed to be robust within reason to different dataset durations of train, test, validation and holdout. The table below shows each instrument’s dataset durations per category.

Table 3.1: Deep Candlestick Mining dataset durations.

ID	Train	Validate	Test	Holdout
FGBL	5 Years	6 Months	3 Months	3 Months
EURUSD	5 Years	6 Months	3 Months	3 Months
BTCUSD	3 Years	6 Months	3 Months	3 Months
ETHUSD	3 Years	6 Months	3 Months	3 Months

The train datasets are used for feature mining and training of the LSTM RNN. A validation dataset is then used to cluster the feature set. It is the clustering process which identifies the DCM patterns by differentiating global and local structure in the feature space. A test dataset is used for an out-of-sample predictive assessment at the individual pattern level, with the holdout dataset being reserved for a full portfolio of patterns predictive test.

3.4 OHLC feature mining

3.4.1 Identification of candidate features

All possible combinations of ratios and differences of one-hour OHLC data are calculated given L lags. The number of such combinations, and therefore the number of potential features, increased very rapidly with L, as can be seen in Figure 3.1, with examples for the case of L=2 in Figure 3.2.

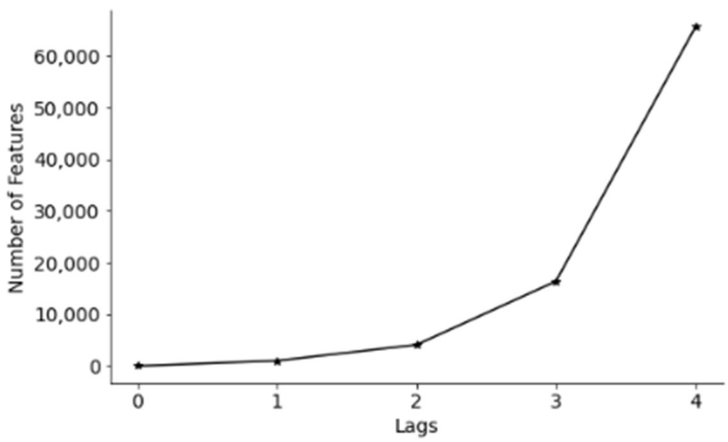


Figure 3.1: Number of potential features as a function of lag, L.

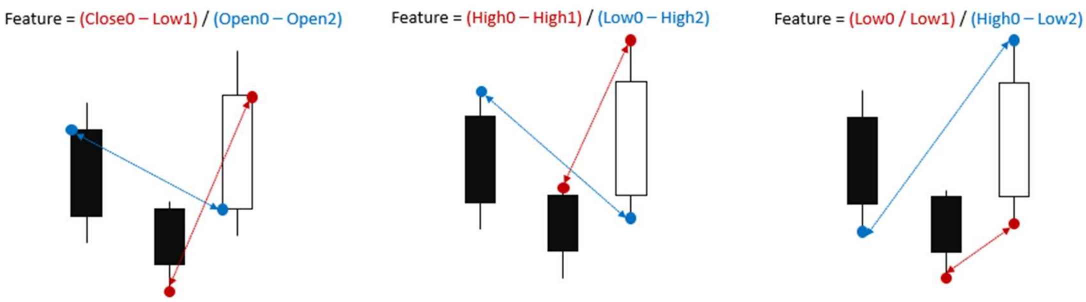


Figure 3.2: Examples of potential features for L=2.

XGBoost is then used to rank the importance of each feature to a target (in this case the future close price directional change), deriving the importance value from the total gain metric. The top N features—where N=100 is chosen here and throughout this work—are selected by inspection of the importance curve. As can be seen in Figure 3.3 the total gain metric curve noticeably flattens for all instruments beyond N=100. N should ideally be optimised for each asset when selecting a feature universe. However, to keep a consistent approach in demonstrating the Deep Candlestick Mining process we chose to use a constant N=100 here with no further optimization; the results presented below are therefore a general indication of the process's utility.

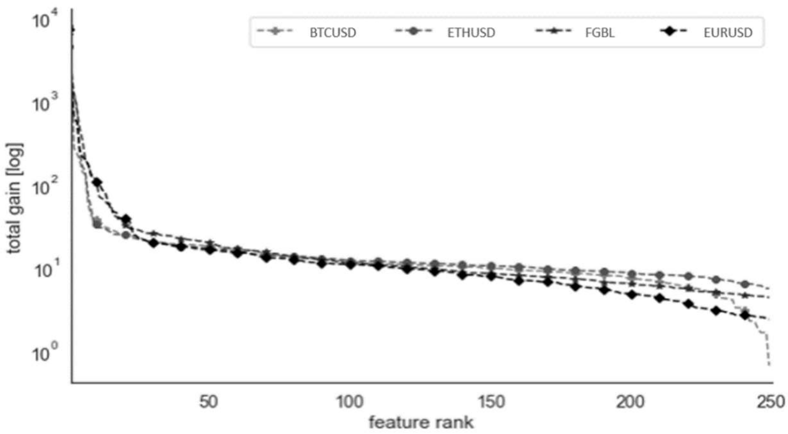


Figure 3.3: Three-lag importance mining curve for BTCUSD, ETHUSD, FGBL and EURUSD for the total gain importance metric.

Using the top N feature universe, a further filtering is then applied, focusing on the correlation of feature-to-target (ft) and feature-to-feature (ff) relationships. Explainability was a strong requirement in this work; thus, for example, the use of the full feature set with dimensionality reduction via PCA, would not be suitable due to the mixing of information from different feature dimensions. Features with a high feature-to-feature or low feature-to-target correlation may be redundant dimensions in the input feature set as they add minimal value to the predictor's discriminatory power. Figure 3.5 (on p. 50) shows these correlation relationships for Bitcoin; it is plain some features demonstrate very high inter-feature correlation and low correlation to target. Therefore, features are selectively chosen based on correlation relationships, so that a maximal predictive power may be extracted with less features, reducing the number of parameters (and model complexity) in the LSTM RNN. Features that pass the tests $|corr_{ft}| \geq c_1$ and $|corr_{ff}| \leq c_2$ (see Figure 3.5, on the following page) make the optimal feature universe, with c_1 and c_2 being optimised on a validation set, with resultant values $c_1 = 0.2$ and $c_2 = 1.0$.

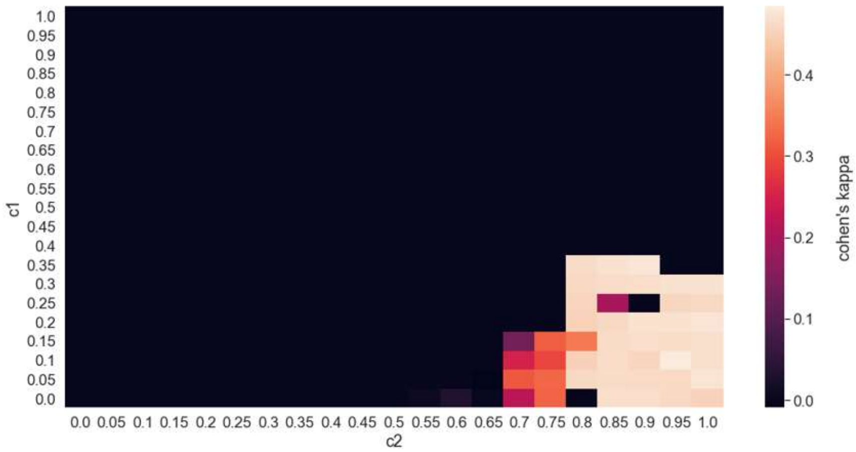


Figure 3.4: Filtered Bitcoin importance mining feature sets performance.

The dark colour in Figure 3.4 is where the correlation filtration process results in zero features. The bottom right of the figure shows that with no filtering (i.e., all N features from the importance mining) the LSTM RNN can even so work quite effectively, suggesting it can work out what features to up-weight and down-weight in the training process on its own. Over all datasets, this filtering achieved ~10% improvement in predictive power.

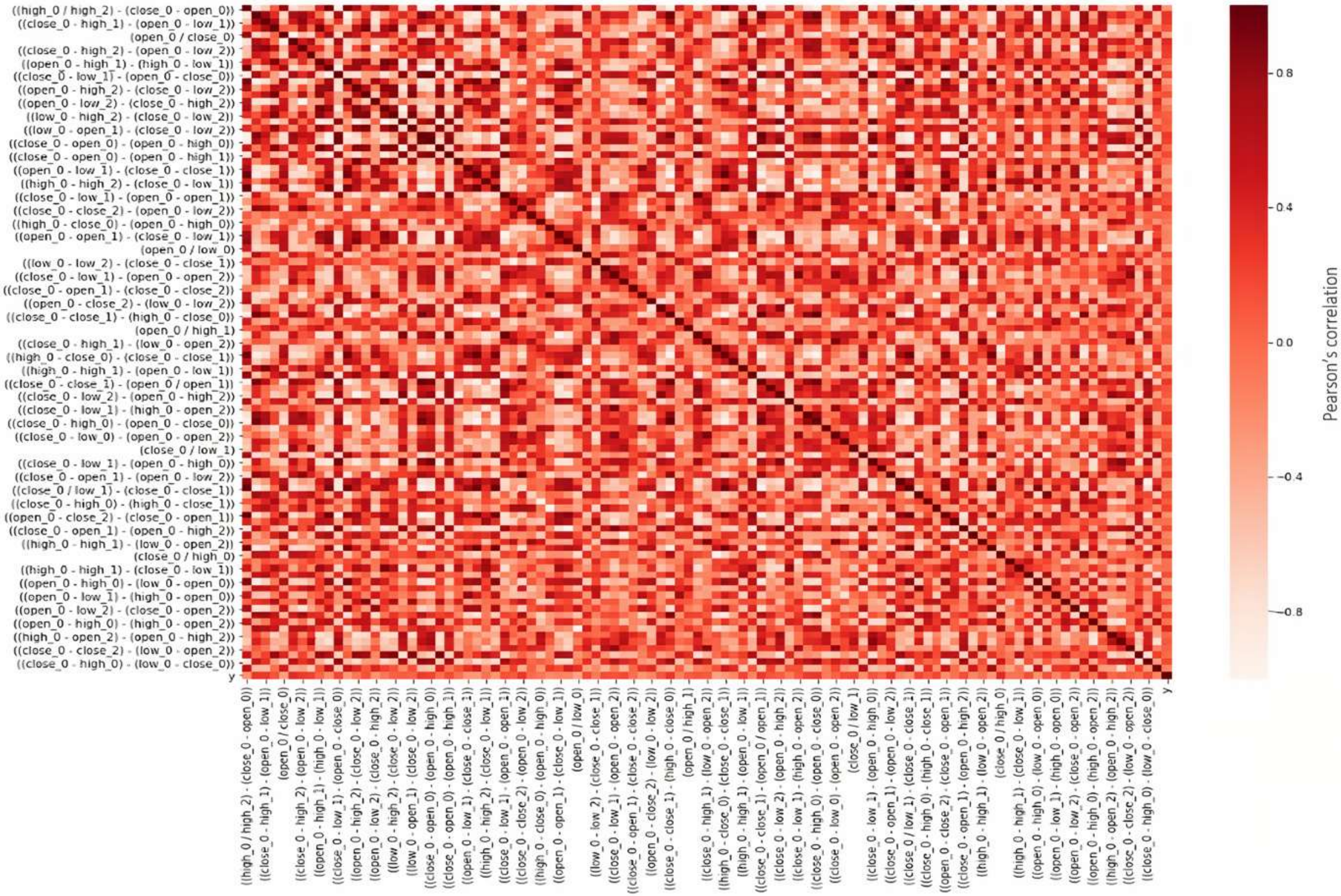


Figure 3.5: Top N Bitcoin correlation feature-to-feature and feature-to-target relationships.

3.4.2 Assessment of predictive value of features using LSTM

A LSTM RNN is used as the directional prediction model taking features generated by the feature importance mining step, 3.2.1, as input. The LSTM RNN is constructed using LSTM cells, Batch Normalisation layers and Dropout to improve regularisation and training speed. The resulting LSTM RNN is trained using ADAM, a first-order adaptive optimisation algorithm described in section 2.1.4.

Inputs to the LSTM RNN are standardised by removing the median and scaling the data to the interquartile range (i.e., between the 25^{th} and 75^{th} quantile). The processes of centring and scaling occur independently over each feature on the train set. The scaler statistics (median and interquartile range) are then saved for use in out-of-sample datasets. This type of scaling is more robust to outliers compared to common standardisation schemas which typically focus on removing the mean and scaling according to unit variance, which can falsely skew the sample mean and variance. It is important to consider the effect of outliers on the feature set in the current work as two of the instruments, Bitcoin and Ethereum, are of the cryptoasset class which is a dataset well-known for containing extreme outliers.

The targets for the LSTM RNN are -1 (for a down movement) and +1 (for an up movement). The network architecture is optimised using a distributed Bayesian Optimisation and Hyperband methodology (BOHB) (see section 2.1.4 for details). BOHB is used to optimise several hyperparameters of the LSTM RNN as specified in Table 3.2.

Table 3.2: Hyperparameters to be optimised by BOHB.

Hyperparameter	**Search Range**
Number of hidden layers	{1,2,3}
Number of units per layer	[2^3, 2^8]
Learning rate	[10^{-6}, 10^{-2}]
Dropout rate	[0,0.5]

Once the optimisation process completes, the sensitivities of the optimal parameter set are assessed. To do this, each parameter is frozen apart from the one which is being tested. Trials are run N times where the parameter is perturbed by a value drawn from a distribution with the mean and variance set as the optimal parameter value (i.e., $\mu(param\_value)$ and $\sigma(param\_value)$). This process is carried out to assess the systematic sensitivity of each parameter, ensuring the optimal solution is robust to a reasonable range of fluctuations around the perceived optimal. The test is critical to a successful financial machine learning model optimisation as financial datasets are very noisy and thus require robust parameter selections to provide valuable out-of-sample predictions. Other statistical sensitivity methodologies exist; however, it was determined the methodology proposed here maintained transparency, clarity and was a simple way to test the robustness of the LSTM RNN's predictive power.

3.5 Generation of novel candlestick patterns

3.5.1 Clustering

The LSTM RNN features which were used to perform the directional prediction on the validation set are now clustered using k-means++, where k is selected by maximising the Silhouette coefficient (SC) (see section 2.1.5 for details on this metric). An initial (parent) clustering is performed on the validation feature set with the intention of revealing global structure. A local (child) clustering is then performed which clusters each of the parent clusters. The optimal parent and child configurations are the ones which maximise the SC.

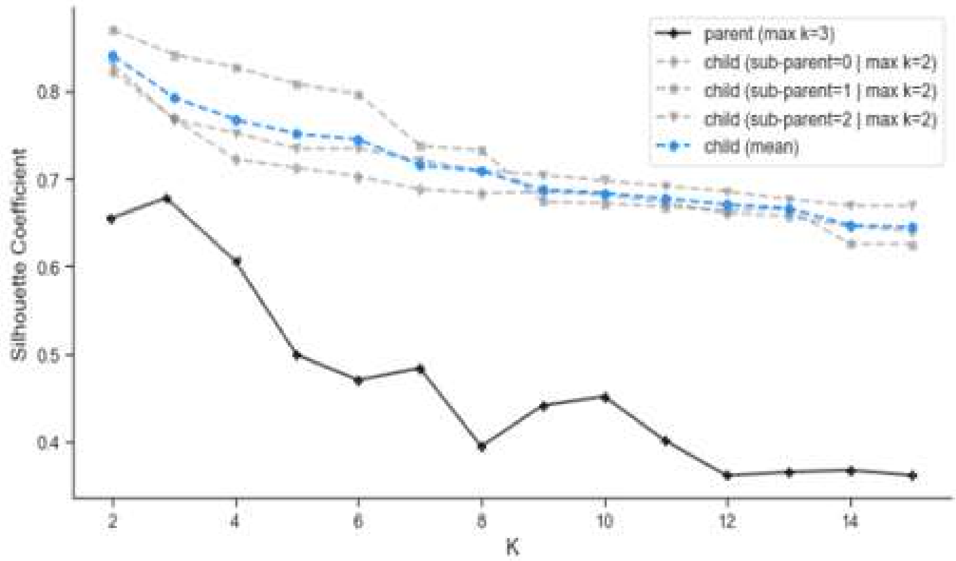


Figure 3.6: An example three-lag Silhouette coefficient optimisation for ETHUSD to demonstrate the process of parent and child clustering.

It is thought that by clustering for global and local structure interesting feature space relationships will be revealed that could allow us to classify specific types of directional movement more accurately. A second line of reasoning as to why a double clustering could be preferable is that “super clusters” might thereby be discovered. It is hypothesised that if these “super cluster” structures exist then they would provide robust and superior statistically significant predictive power, which would provide an edge over other market participants' models given the difficulty of identifying such structures in noisy datasets. Thus, these “super clusters” are likely to represent the most significant and powerful candlestick patterns. The clustering process is carried out for each instrument and lag lookback separately revealing structure in each.

The clustering of ETHUSD data (see Figure 3.6 above) shows an interesting result where the parent clustering sets $k = 3$ and the child clustering sets all $k = 2$. When a clustering algorithm sets $k = 2$ it can often mean there is no structure present, and the algorithm was forced to arbitrarily select a partitioning of the dataset. To validate there is indeed cluster structure in such cases each dimension of the feature set is plotted against its magnitude shown in Figure 3.7 below. Each column represents the parent cluster ID, and each row represents the child cluster ID. It can be observed that all structures are unique thus confirming that there is structure in this dataset. This process is carried out for each asset and lag resulting in all clusters being confirmed to be valid – i.e., in the optimal feature sets, unique structure was discovered and validated. Other clustering techniques and k-selection criteria could have been used; however, the optimal selection of a clustering algorithm and associated selection criteria is outside the scope of this work.

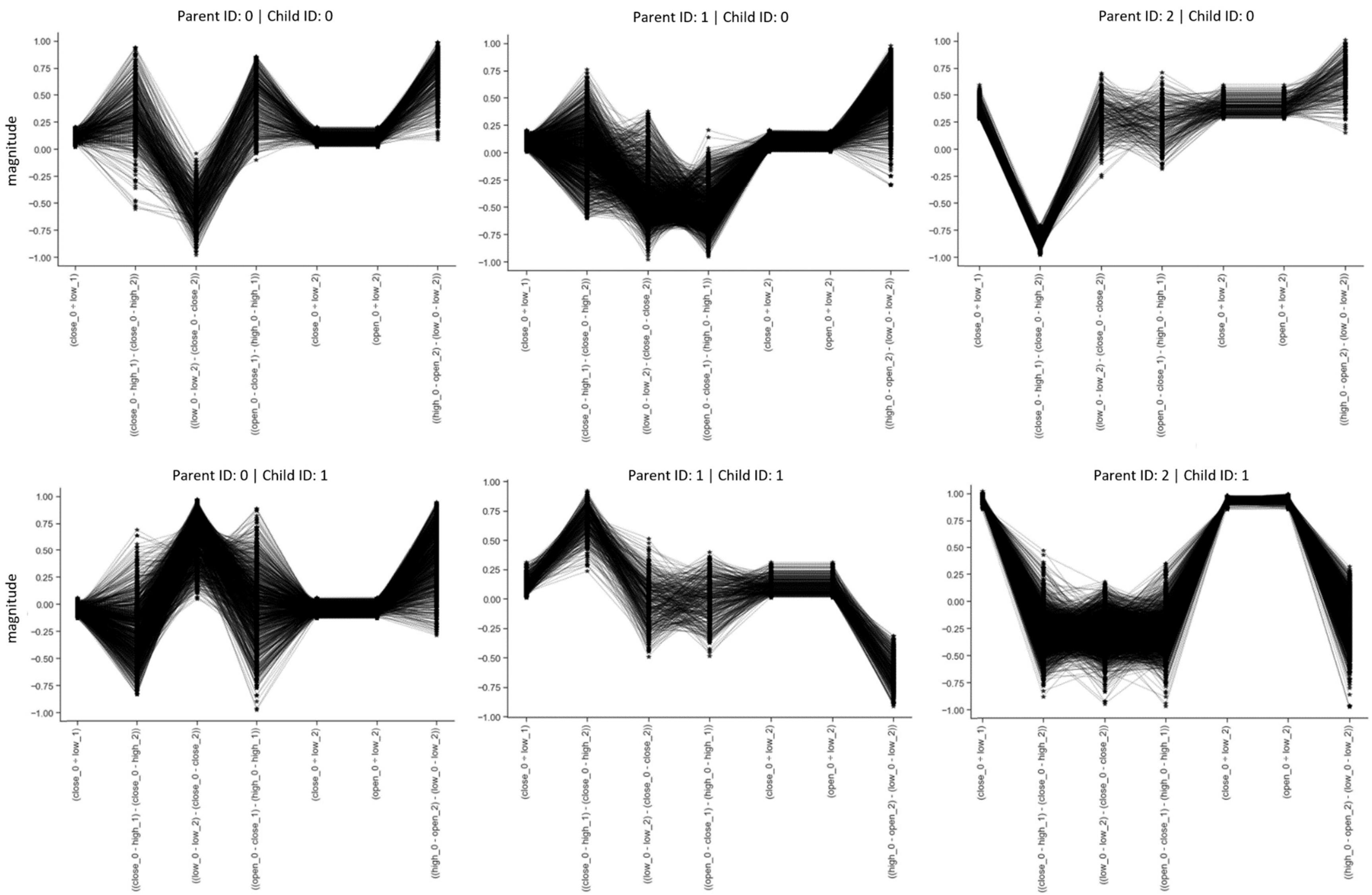


Figure 3.7: Selected clusters from ETHUSD confirming different structures in the magnitude space of each feature.

3.5.2 Cluster filtering and candlestick selection

After confirming that structure exists, using the methods of the previous section, each cluster is analysed further to determine if it encapsulates a valid candlestick pattern.

Initially, a visual similarity check is done. Each cluster member is indexed to the OHLC data it corresponds to, allowing a visual inspection of all candidate OHLC arrangements to confirm that the clustering process did indeed group together patterns of similar shape. A Silhouette coefficient (SC) threshold is then optimised per cluster to increase the consistency of shape. The optimal SC threshold is the one that corresponds to the lowest cut-off (maintaining sample size) where each cluster member's shape is deemed to be consistently representative of the average shape per cluster. After imposing this cut per cluster all instances were found to consistently represent similar shapes in the clusters they had been assigned to. A further validation is then carried out, requiring a valid cluster to satisfy the following criteria:

1. The LSTM RNN's Cohen's kappa > 0%.
2. The percentage of up movements deviates from 50%.

   This is calculated by indexing each candidate back to the original OHLC dataset, counting the number of candidates that represent an up movement and dividing by the total number of candidates belonging to that cluster.

3. The LSTM RNN's majority prediction agrees with the cluster's direction.

Criterion 1 has the purpose of verifying that the LSTM RNN was able to gain discriminatory power from the feature structure. If the network's Cohen's kappa score is <= 0%, then it was unable to use the feature structure effectively.

If this is true, then the candlestick pattern associated with the cluster would not represent a meaningful feature representation as the LSTM RNN deemed it insignificant and could not gain any informative information from it.

Criterion 2 verifies that the cluster candidates are biased in one direction. If the candidate's directional movements at $t + n$ are not skewed in one direction, on average, then there is no directional bias, resulting in uncertainty as to the future direction of price movement.

Criterion 3 verifies that the cluster and LSTM RNN agree on the directional movement at $t + n$. If this condition is not satisfied, then the derived candlestick pattern would be ineffective as the network derived the opposite signal from which the ground truth represents. In strongly trending markets this criterion could be satisfied when criterion 1 is not, hence the need for both.

The DCM-based candlestick patterns represent the centroids of each cluster which passes the tests. These patterns are essentially what the LSTM RNN would have seen if it had been looking at OHLC data as a human might look at a candlestick chart when a prediction was made.

3.6 Testing candlestick patterns individually

In this section results are presented first for traditional candlestick patterns and then for DCM-based patterns, in each case assessing the statistical significance of their predictive power in terms of accuracy (where each matched pattern's accuracy is calculated as the ratio of correct predictions to total predictions, for a pattern that has been verified as significant in predicting movement in a given direction).

3.6.1 Traditional candlestick patterns

To assess the power of the deep mined candlestick patterns against an appropriate baseline an assessment of 100 bull (predicting up) and bear (predicting down) traditional candlestick patterns (50 candlestick types overall)[1] were tested on FGBL, EURUSD, ETHUSD, BTCUSD hourly data. Significance levels were calculated using a binomial distribution (as in Hércules et al. 2013; see section 2.1.4 for details), where the null hypothesis was that candlesticks are no better than guessing, which translates to 50% directional accuracy. Significant candlesticks (see Table 3.4) are defined to be patterns with a directional predictive power significant at 10% or better.

It can be seen from Table 3.4 (below) that only four patterns were significant at the 5% and 10% levels for FGBL, EURUSD and ETHUSD. Analysis on BTCUSD revealed five significant patterns. Interestingly no pattern was significant at the 1% level. Hence, while there is some predictive ability in traditional candlestick patterns, it appears not to be widespread. This finding is in line with the negative bias most academic studies showed toward the usefulness of candlestick charting, as reviewed in section 2.2.1.

[1] 2 Crows; 3 Black Crows; 3 Inside; 3 Line Strike; 3 Outside; 3 Stars in South; 3 White Soldiers; Abandoned Baby; Advance Block; Belt Hold; Break Away; Closing Marubozu; Conceal Baby Swell; Counter Attack; Dark Cloud Cover; Down Side Gap 3 Methods; Downside Gap 2 Crows; Engulfing; Evening Star; Gap Side White; Hammer; Hanging Man; Harami; High Wave; Hikkake; Hikkake Mod; Homing Pigeon; Identical 3 Crows; In Neck; Inverted Hammer; Ladder Bottom; Long Line; Marubozu; Mat Hold; Matching Low; Morning Star; Piercing; Rise Fall 3 Methods; Separating Lines; Shooting Star; Short Line; Spinning Top; Stalled Pattern; Stick Sandwich; Takuri; Tasuki Gap; Thrusting; Tri Star; Unique 3 River

Table 3.3: Significant traditional candlestick patterns performance on the test dataset.

Pattern	**Number of Candles**	**Accuracy (%)**	**Type**	**Significance Level**
		FGBL		
Advanced Block	3	51.22	Bear	**
Inverted Hammer	3	52.23	Bull	*
3 Inside	3	52.85	Bear	**
Engulfing	2	53.53	Bull	**
		EURUSD		
Harami	2	54.94	Bull	*
Inverted Hammer	1	56.89	Bull	*
Matching Low	2	57.84	Bull	**
Advanced Black	3	53.53	Bear	*
		BTCUSD		
3 Line Strike	4	52.44	Bull	**
3 White Soldiers	3	55.40	Bear	*
Engulfing	2	50.73	Bull	*
Inverted Hammer	1	50.28	Bull	**
3 Outside	3	56.38	Bull	*
		ETHUSD		
Dark Cloud Cover	2	55.44	Bear	**
Doji Star	1	53.11	Bear	*
Engulfing	2	51.23	Bull	**
Engulfing	2	55.36	Bear	*

(*: significance at 10%; **: significance at 5%; ***: significance at 1%)

3.6.2 Deep candlestick mined patterns

Deep mined candlestick patterns are dataset-specific, being mined from the features that the LSTM RNN used as inputs. A list of significant patterns for the datasets considered is presented in Tables 3.5 and 3.6 below.

For the traditional instruments (FGBL and EURUSD) the DCM process was able to discover twelve significant patterns for FGBL and eight for EURUSD. The process additionally found five patterns to be significant at the 1% significance level on FGBL, and three achieving this level of significance on EURUSD. This contrasts with no significant patterns being found at the 1% level when using traditional candlestick patterns (see Table 3.4, above). Moreover, the DCM-based significant patterns achieve average directional accuracies of 62.17% and 59.39% on FGBL and EURUSD respectively. This is an improvement of 18.5% for FGBL and 6.43% for EURUSD when compared against the significant traditional candlesticks.

For the cryptocurrency instruments (BTCUSD and ETHUSD) the DCM process was able to identify ten significant candlestick patterns for BTCUSD and thirteen for ETHUSD, achieving average directional accuracies of 61.18% and 61.14% respectively. This represents an outperformance over the traditional candlesticks of 8.13% and 7.36% for BTCUSD and ETHUSD.

Table 3.4: Significant deep mined candlestick patterns for traditional assets.

Cluster	Number of Candles	Accuracy (%)	Type	Significance Level
		FGBL		
0,3	2	61.11	Bull	**
0,4	2	58.92	Bull	**
1,3	2	58.51	Bear	**
0,1	3	62.34	Bear	***
0,7	3	67.12	Bull	**
1,2	3	57.67	Bear	**
1,4	3	55.34	Bull	***
2,5	3	69.12	Bear	**
0,0	4	59.37	Bear	***
0,2	4	60.90	Bear	***
1,5	4	63.88	Bear	**
3,0	4	71.75	Bear	***
		EURUSD		
0,5	2	58.34	Bear	*
0,1	3	57.12	Bull	**
0,4	3	60.13	Bear	***
1,2	3	59.99	Bear	*
3,3	3	56.45	Bull	**
0,4	4	63.22	Bear	***
1,2	4	61.45	Bear	*
1,4	4	58.44	Bull	***

(*: significance at 10%; **: significance at 5%; ***: significance at 1%)

Table 3.5: Significant deep minded candlestick patterns for Bitcoin and Ethereum.

Cluster	Number of Candles	Accuracy (%)	Type	Significance Level
BTCUSD				
0,1	2	57.70	Bull	**
0,4	2	61.19	Bull	*
2,1	2	65.53	Bear	**
0,3	3	62.68	Bull	**
1,1	3	61.23	Bear	*
1,3	3	55.65	Bull	*
2,0	3	66.60	Bear	***
0,4	4	57.26	Bear	**
3,3	4	59.85	Bull	*
4,1	4	64.12	Bull	**
ETHUSD				
0,2	2	59.98	Bear	**
3,4	2	66.93	Bull	**
4,1	2	61.34	Bull	*
5,2	2	56.23	Bear	***
1,0	3	57.46	Bull	*
2,1	3	59.09	Bear	**
2,5	3	62.72	Bull	*
0,0	4	74.75	Bear	**
0,2	4	67.13	Bull	**
1,0	4	62.51	Bull	**
2,2	4	55.23	Bull	*
3,0	4	58.03	Bull	***
6,1	4	53.53	Bull	*

(*: significance at 10%; **: significance at 5%; ***: significance at 1%)

These results, comparing traditional to DCM-based candlestick patterns, therefore show substantial promise for the DCM process in discovering predictive candlestick patterns. DCM-based patterns were found to outperform on all instruments, with more patterns being significant at the 1% level on traditional markets. This too is an interesting discovery as it implies the process was able to find more robust structure in the feature space of traditional markets compared to the cryptocurrency markets. This again is in line with an earlier hypothesis (see Section 2.1.3) that, although cryptoasset markets appear to be more inefficient than mature instruments such as FGBL and EURUSD, the patterns on cryptoassets tend to decay faster, showing less robustness through time (i.e., are less robust to different market conditions). This could be related to the frequent structural breaks in the cryptoasset markets compared to the more consistent trading environment seen for traditional instruments.

Figures 3.8 and 3.9 present examples of t-SNE visualisations for three lagged (4 candle) candlestick patterns. ETHUSD and FGBL have been selected here as they showed the most interesting structure for each of the categories (traditional and cryptoasset class). The visualisations demonstrate that using a non-linear dimensionality reduction technique reveals more interesting structure analysis than in Figure 3.7. This can help validate the quality of each structure and confirm that the thought process behind the valid cluster criteria makes sense. The idea of “super clusters” mentioned previously is difficult to confirm in a high dimensional space without the use of a non-linear dimensionality reduction algorithm such as t-SNE. In both cases presented below a clear structure can be observed.

Figure 3.8 (p.66) shows a very pronounced structural shape is evident for cluster (0,2), which has a high accuracy of 67.13% at the 5% significance level.

This was investigated further revealing the cluster candidates to be of very similar shapes and directional representation. Further well-defined structure is observable across the different clusters for ETHUSD indicating there is robust structure in this dataset which could be exploited.

Figure 3.9 (p. 67) presents an equivalent structural analysis for FGBL, which also confirms the presence of well-defined groupings associated with select clusters. Interestingly, it can also be observed that structure associated with parent cluster 1 and 3 can be rather sparse on occasion. This can indicate that within clusters there is a dense section which very similar features have clustered around and that as the distance extends from the centroid the decay in pattern reliability (feature similarity) is accelerated.

Given the observations from these t-SNE charts that some clusters suffer from a certain sparsity away from a cluster's centroid, the decision to optimise a Silhouette coefficient (SC) threshold per cluster is a reasonable suggestion. The patterns which were clustered together and did not satisfy the SC threshold were investigated and found to represent those examples which were sparsely distributed by the t-SNE (see Figure 3.9). These examples did not conform to a common shape, which the other cluster candidates did, which would impact the shape and interpretation of a DCM-based pattern.

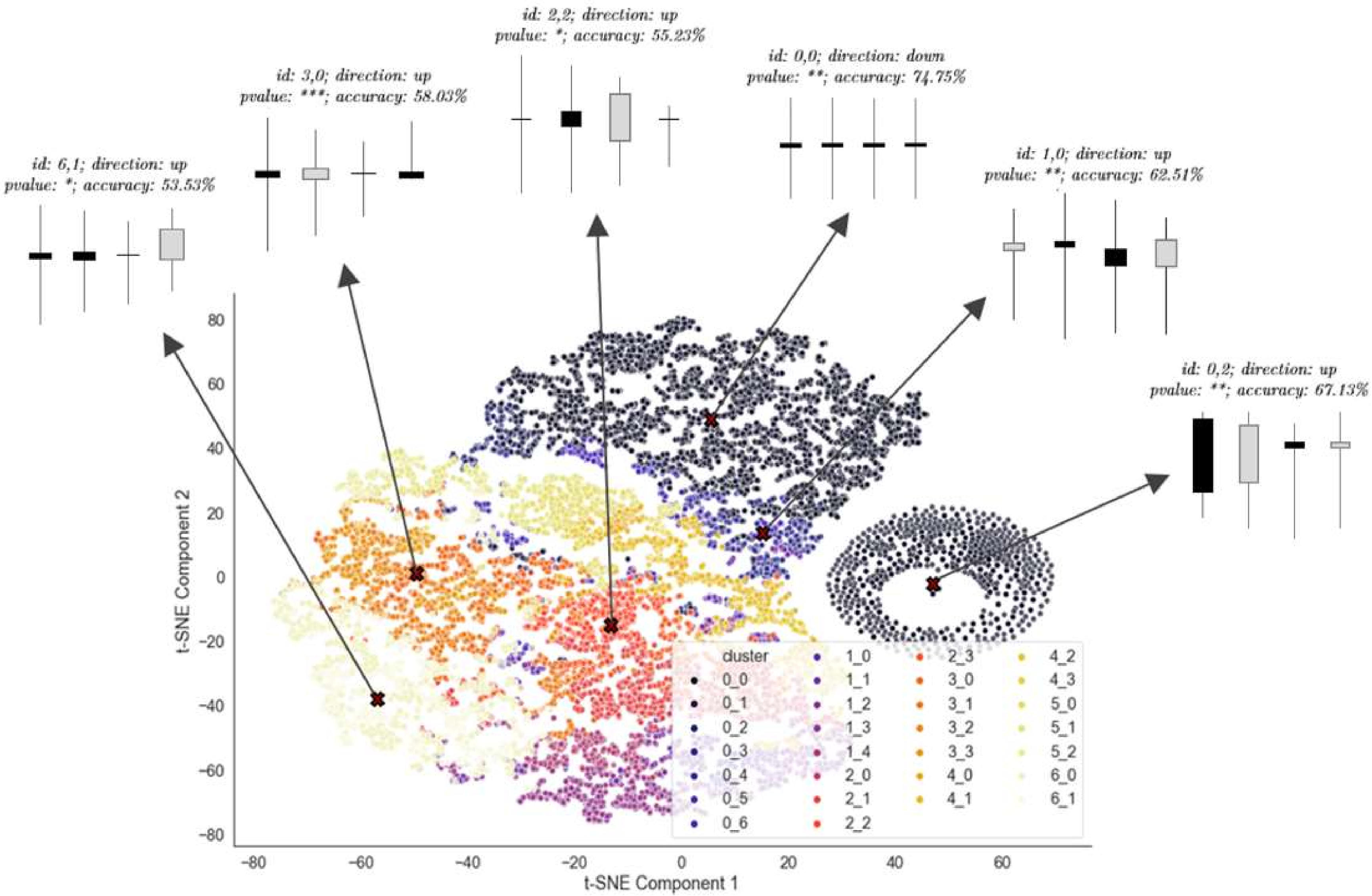


Figure 3.8: t-SNE **ETH** Deep Candlestick Mining.

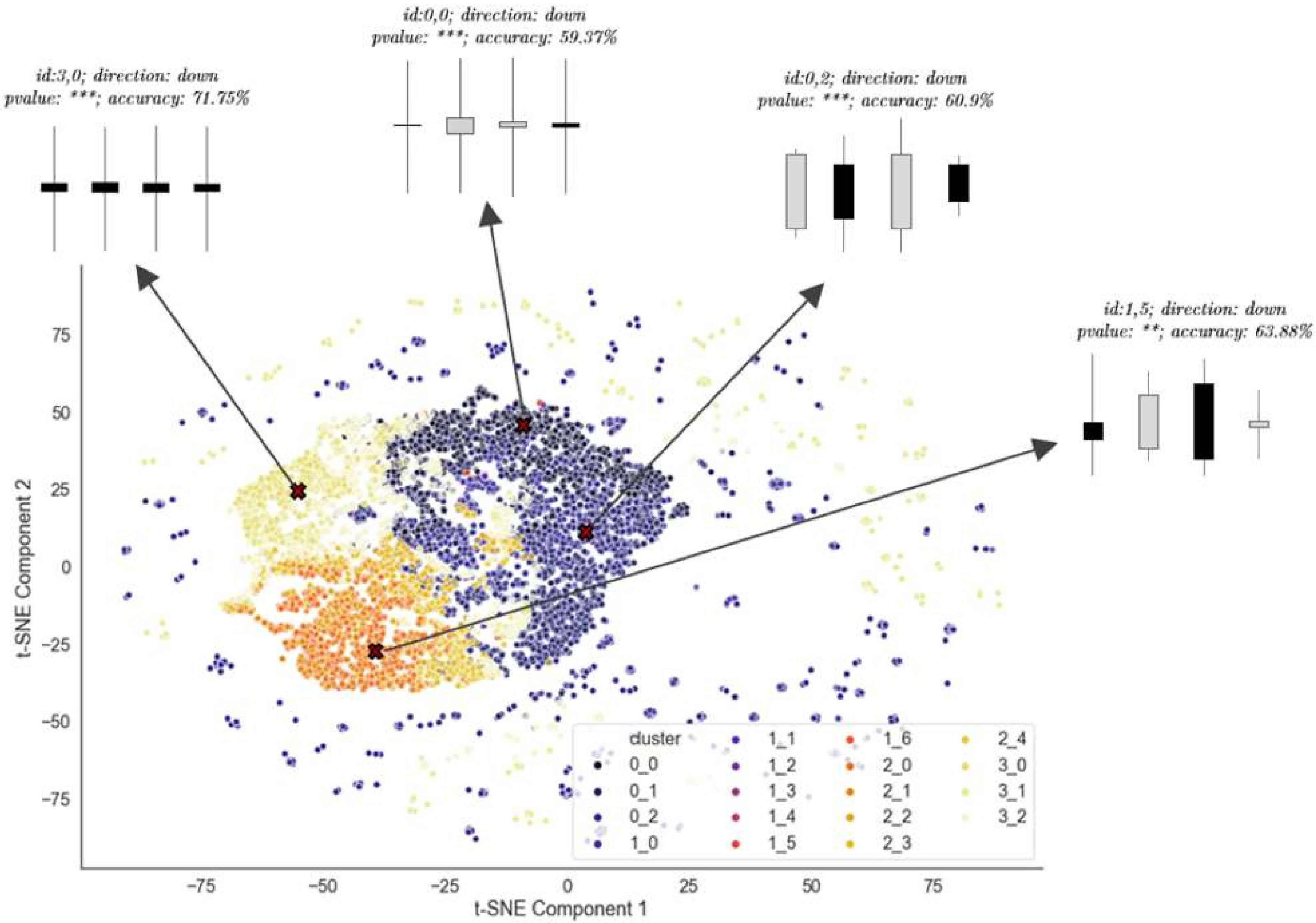


Figure 3.9: t-SNE **FGBL** Deep Candlestick Mining.

Table 3.6: FGBL Deep mined candlestick patters for all lags.

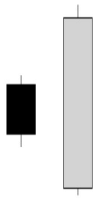
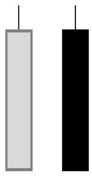
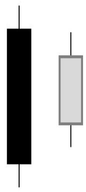
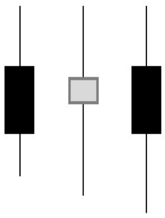
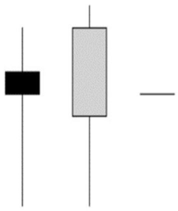
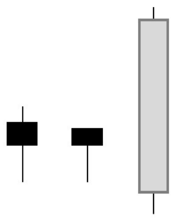
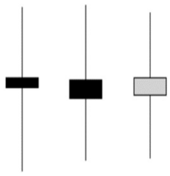
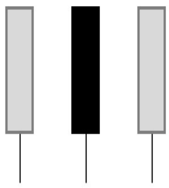
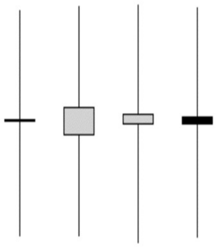
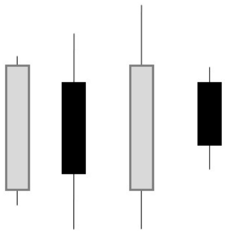
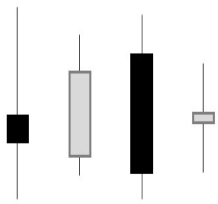
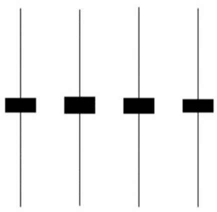
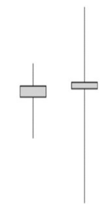
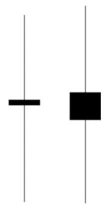
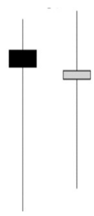
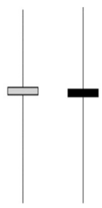
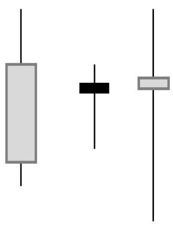
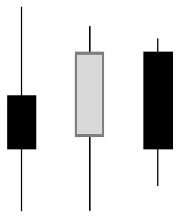
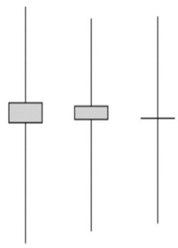
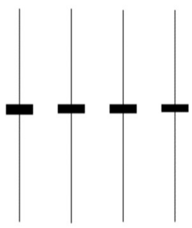
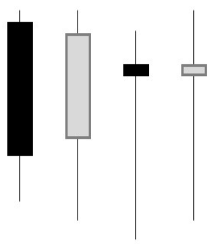
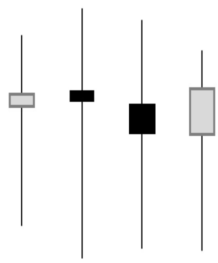
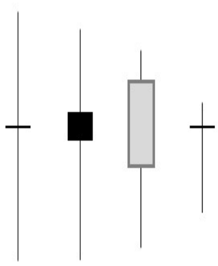
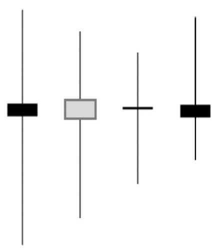
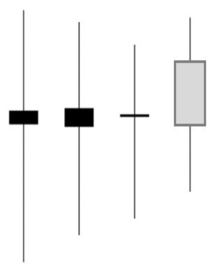
2 Candle Patterns		
		
Cluster: 0,3	Cluster: 0,4	Cluster: 1,3
3 Candle Patterns		
		
Cluster: 0,1	Cluster: 0,7	Cluster: 1,2
		
Cluster: 1,4	Cluster: 2,5	
4 Candle Patterns		
		
Cluster: 0,0	Cluster: 0,2	Cluster: 1,5
		
Cluster: 3,0		

Table 3.7: ETHUSD Deep mined candlestick patters for all lags.

2 Candle Patterns		
		
Cluster: 0,2	Cluster: 3,4	Cluster: 4,1
		
Cluster: 5,2		

3 Candle Patterns		
		
Cluster: 1,0	Cluster: 2,1	Cluster: 2,5

4 Candle Patterns		
		
Cluster: 0,0	Cluster: 0,2	Cluster: 1,0
		
Cluster: 2,2	Cluster: 3,0	Cluster: 6,1

Tables 3.7 and 3.8 show examples of novel candlestick patterns discovered by the deep candlestick mining (DCM) process for FGBL and ETHUSD (as these asset classes were presented in the t-SNE analyses). Similar shapes were discovered for BTCUSD and EURUSD but are omitted for brevity (though performance results can be observed in Tables 3.5 and 3.6, and candlestick visualisations can be found in Appendix A).

Table 3.6 displays the FGBL DCM-based patterns. FGBL was shown to have four significant traditional candlestick patterns in Table 3.3. It is interesting to note that the DCM-based patterns, while mainly discovering novel shaped representations, also rediscovered some traditional candlestick patterns in clusters 0,3 and 0,4 which are very similar to a bullish Engulfing and a bullish + bearish Inverted Hammer candlestick pattern, respectively.

Table 3.7 displays the ETHUSD DCM-based patterns. When analysing the traditional candlestick patterns for this instrument in Table 3.3 it was shown to have four significant patterns. In contrast, the DCM process discovered thirteen patterns of interest. Here, the DCM process again rediscovered traditional candlestick patterns which were not listed as significant in the previous analysis and combined them as part of a DCM-based pattern. For example, the traditional pattern “Doji Star” features heavily in many of the DCM-based pattern formation for ETHUSD.

The observation that (1) the DCM process can rediscover significant traditional patterns and (2) can find traditional candlestick patterns which were not significant on their own but significant when grouped as part of a larger formation is an important point, as it shows the DCM's ability to rediscover and reinvent pattern formations while also finding novel formations of candlestick patterns. Point number 2 may indicate significant (or even insignificant) candlestick patterns could be strong conformation signals, rather than simply representing a directional movement themselves, which the analysis presented earlier in this section suggests would yield poor results most of the time. Interestingly, in all cases the DCM approach appears to prefer more candlesticks in a pattern implying a greater information content is required to be predictive. Many popular traditional candlestick patterns only contain one or two candles which could be a reason why the vast majority have no power but when combined exhibit significant predictive power.

3.7 Traditional vs. deep mined candlestick system

Often a practitioner will use multiple candlestick patterns for making decisions. A comparison in this spirit between traditional and deep mined candlesticks was carried out by using all the patterns available in either category. The holdout dataset was used to assess the predictive power of both systems, in terms of Cohen's kappa. As can be seen in Table 3.8 the DCM system outperformed the traditional system by 83.6%, 98.7%, 110.06% and 117.89% on FGBL, EURUSD, BTCUSD and ETHUSD respectively.

Table 3.8: Performance comparison between a basket of traditional candlestick patterns vs. DCM candlestick patterns on the holdout dataset.

Asset	**Number of Traditional Patterns**	**Traditional Cohen's kappa**	**Number of Deep Mined Patterns**	**Deep Mined Cohen's kappa**
FGBL	4	4.45%	12	8.17%
EURUSD	4	7.67%	8	15.24%
BTCUSD	5	6.26%	10	13.15%
ETHUSD	4	6.55%	13	14.27%

3.8 Discussion

The deep candlestick mining (DCM) process introduced here has been shown to be remarkably effective at discovering statistically significant OHLC patterns. This is not in conflict with the many academic studies which claim candlestick patterns have no, or limited, predictive power (for example, Fock et al., 2005; Marshall et al., 2006; Horton, 2009) because the patterns that the DCM process discovers are largely novel (though some interesting correspondences with traditional candlestick patterns were discovered). DCM derived patterns outperformed the best (according to the earlier analysis) of the traditional patterns by 83.6%, 98.7%, 110.06% and 117.89% on FGBL, EURUSD, BTCUSD and ETHUSD respectively in relation to their ability to forecast directional movement better than random. The DCM process has many parts that could be further optimised to produce potentially better results. It would be expected these optimisations would be both instrument and time aggregation (daily, hourly, minute, etc.) dependent. The results here are therefore only an early indication of the promise of deep candlestick mining.

Chapter 4

A new methodology to exploit predictive power in (open, high, low, close) data

In the previous chapter it was shown that OHLC bars could be used to extract powerful candlestick patterns able to predict close price directional movements with a maximum (for EURUSD) Cohen's kappa of 15.24%. This same OHLC-based prediction framework, incorporating a feature discovery and mining process with LSTM RNNs, is used here to demonstrate the remarkable predictive power which can be extracted from a new target, termed *mid-price*, with directional forecasts of greater than 55% Cohen's kappa on hourly data aggregations. The predictive power of the new target is shown to hold across FGBL, EURUSD, BTCUSD and ETHUSD, emphasising that this power is not simply an artefact of one instrument but can easily generalise over several vastly different asset classes.

4.1 Introduction

The work of this chapter utilises ML technology in the form of XGBoost, LSTM RNNs and Bayesian optimization with Hyperband (BOHB) as key components in a process for trend detection which takes advantage of the relative ease of prediction of the mid-price (defined in terms of OHLC candlestick levels in Section 2.1.2) when compared to the traditional close price prediction target. It can thus be demonstrated that if the target is carefully chosen OHLC levels can have a remarkably high predictive potential, beyond that evidenced in the previous chapter.

4.2 Background

There exists no prior literature relating to a mid-price, as it will here be termed, based on a candlestick structure, as proposed here. The common definition of a mid-price is the price halfway between the bid and ask; this has no relevance to the current work. Here, a mid-price pertains to the mid-point between either the high and low or the open and close of an OHLC data aggregation. There have, however, been many studies focusing on the predictive power of candlestick patterns, as reviewed in Section 2.2.1, and studied in Chapter 3. As discussed in these earlier thesis sections, these studies have reported varying results, with most evidencing little or no value in these patterns as predictors of close price movements, although it was shown in Chapter 3 that if advanced machine learning methodologies are used then dataset-specific candlesticks appear to provide some substantial predictive power.

Mid-price motivation

As defined in Section 2.1.2, there are two version of the proposed mid-price to be used as a target to predict trend direction. The spur behind the use of this new financial timeseries price level as a training target was the observation that it contained less noise than traditional targets such as the close price.

To demonstrate this, the timeseries of the close price, mid-price-1, and mid-price-2 (see Section 2.1.2) for hourly data for ETHUSD are examined, to investigate the differences in standard deviation (i.e., noise content) of their price movements. Figure 4.1 shows these differences, where for each month in the ETHUSD dataset the standard deviation of the returns for hourly data was computed over the close-price, mid-price-1 and mid-price-2. It can be observed that there is a significant drop in noise content for both versions of the mid-price returns when compared to the close price returns.

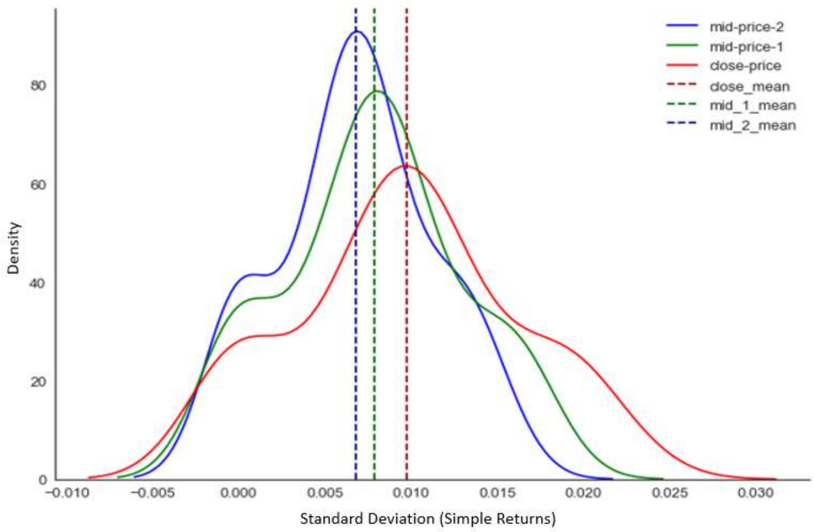


Figure 4.1: Differences in standard deviation per month on hourly data for ETHUSD.

Mid-price-2 is observed to contain the least noise content, a difference of 42.39% compared to the close-price returns. This is a substantial reduction in noise and is part of the reason that this new financial timeseries metric is much easier to predict. Similar results were obtained for many other examples of financial time series data, confirming mid-price-2, in particular, as a less noisy target than either mid-price-1 or the close-price.

4.3 Methodology

The methodology implemented here uses the tools (and data ranges) developed in Section 3.2 and 3.3 of the previous chapter.

4.3.1 Mid-price directional prediction

The model training is along the same lines as Section 3.3.2 of the previous chapter, as the feature mining discovery process and LSTM RNN hyperparameter optimisation are executed in the same way. The difference here is the target, which is one of the mid-price definitions rather than the close price. As in the previous chapter, the LSTM RNN output value is in the range $[-1, +1]$ with targets of -1 (down) and +1 (up). It should be noted that as we are predicting 1-hour intervals (a substantial duration for financial markets to evolve) there are no periods where the target variable remains the same in the datasets which these models are trained on. This is due to the underlying assets having a relatively high noise content when compared to other popular financial instruments, such as US Treasuries, where one might expect price evolution to be more muted.

4.4 Results

4.4.1 Baseline performance: use of close and OHLC lags as inputs

A baseline performance was established by investigating the prediction of both close- and mid-price targets (see Table 4.1 and 4.2) from close price lags and OHLC lags (defined as a full set of OHLC lags, for two preceding time steps, a total of eight features in all). Lagged inputs are defined by the equation below,

$$\delta_i = \frac{(p_i - p_{i-1})}{p_{i-1}}, \tag{4.1}$$

where p_i is the current price and p_{i-1} is the previous price.

The use of simple returns, as defined above, was selected for feature construction purposes as this definition is most used in the candlestick literature, which allows us to compare results more easily to those of others, and anchor the analysis. Log returns are another common feature construction, which are popular in many financial feature engineering studies as they are very effective in reducing seasonality in returns. However, both of these returns-derived series suffer from seasonality in their volatility patterns (solutions exist to eradicate this, in the form of fractional differencing methods, but this is outside the scope of this work) and exhibit little to no difference in their statistical properties when constructing intra-day features. Hence, it was decided that features would be constructed using the simple return, as this is more in line with the related literature, while maintaining similar statistical properties to log returns when applied to the financial timeseries under consideration in this study.

In this baseline experiment a lag setting of two was selected for initial experimentation. However, when constructing features over close and OHLC price levels, an imbalance in feature dimensionality occurs as the close-from-close feature set would have only two dimensions compared to the eight dimensions generated from the OHLC feature constructions (as lags are taken across all of the open, high, low and close price levels), causing a potential information shortfall. To remedy this, experiments (results not shown) were carried out where close lags were set at eight to have equivalence in feature dimensionality and informational content, but it was discovered results were robust to the number of close lags used, which made little to no difference to the result. Hence, all experiments used a lag setting of two. Table 4.1 and 4.2 present the baseline results across both traditional and cryptoasset markets where the metrics for F1, Recall and Precision are stacked, with the top values representing the performance for an upward prediction and the lower values representing the performance for a downward prediction.

Table 4.1: Baseline performance (i.e., close or OHLC lags used as inputs) for a one-hour prediction horizon on out-of-sample data for traditional markets; it is clear that the combination of input OHLC and target mid-price-2 gives the best performance.

I/O Configuration	F1	Recall	Precision	MCC	Cohen's kappa
FGBL					
Close from Close	0.5718 0.4296	0.6429 0.3744	0.5149 0.5038	1.80%	1.74%
Mid-1 from Close	0.6881 0.6963	0.6908 0.6937	0.6854 0.6990	38.44%	38.44%
Mid-2 from Close	0.7315 0.7342	0.7535 0.7137	0.7109 0.7560	46.70%	46.62%
Close from OHLC	0.6598 0.1051	0.9409 0.0588	0.5080 0.4906	-0.07%	-0.03%
Mid-1 from OHLC	0.7074 0.7276	0.6942 0.7407	0.7212 0.7149	43.55%	43.52%
Mid-2 from OHLC	**0.7467** **0.7556**	**0.7592** **0.7438**	**0.7347** **0.7678**	**50.27%**	**50.25%**
EURUSD					
Close from Close	0.4376 0.5497	0.3901 0.6091	0.4983 0.5008	-0.08%	-0.08%
Mid-1 from Close	0.7065 0.6965	0.7138 0.6892	0.6993 0.7040	40.32%	40.31%
Mid-2 from Close	0.7533 0.7531	0.7683 0.7386	0.7388 0.7682	50.69%	50.66%
Close from OHLC	0.4371 0.5722	0.3783 0.6489	0.5175 0.5118	2.82%	2.71%
Mid-1 from OHLC	0.7323 0.7209	0.7430 0.7102	0.7219 0.7319	45.36%	45.34%
Mid-2 from OHLC	**0.7569** **0.7527**	**0.7783** **0.7322**	**0.7366** **0.7744**	**51.08%**	**51.00%**

Table 4.2: Baseline performance (i.e., close or OHLC lags used as inputs) for a one-hour prediction horizon on out-of-sample data for cryptoasset markets; it is clear that the combination of input OHLC and target mid-price-2 gives the best performance.

I/O Configuration	**F1**	**Recall**	**Precision**	**MCC**	**Cohen's kappa**
			BTCUSD		
Close from Close	0.5487 0.4696	0.6114 0.4191	0.4977 0.5339	3.10%	3.03%
Mid-1 from Close	0.6464 0.6642	0.6560 0.6551	0.6371 0.6735	31.08%	31.07%
Mid-2 from Close	0.6951 0.7535	0.6538 0.7940	0.7420 0.7168	45.33%	45.03%
Close from OHLC	0.5087 0.5417	0.5062 0.5442	0.5111 0.5393	5.04%	5.04%
Mid-1 from OHLC	0.6797 0.7099	0.6731 0.7162	0.6865 0.7036	38.97%	38.97%
Mid-2 from OHLC	**0.7405** **0.7284**	**0.6928** **0.7952**	**0.7356** **0.7334**	**46.10%**	**46.97%**
			ETHUSD		
Close from Close	0.5453 0.4358	0.6066 0.3872	0.4953 0.4982	-0.63%	-0.62%
Mid-1 from Close	0.6515 0.6741	0.6360 0.6898	0.6678 0.6591	32.64%	32.60%
Mid-2 from Close	0.7459 0.7232	0.7804 0.6900	0.7144 0.7598	46.23%	46.03%
Close from OHLC	0.5509 0.4940	0.5862 0.4626	0.5195 0.5300	4.92%	4.88%
Mid-1 from OHLC	0.7064 0.7252	0.6898 0.7420	0.7238 0.7092	43.24%	43.19%
Mid-2 from OHLC	**0.7392** **0.7324**	**0.7512** **0.7206**	**0.7276** **0.7446**	**46.80%**	**47.18%**

Table 4.1 compares the prediction results for traditional markets FGBL and EURUSD, using close-from-close lags to predict the direction of the t+1 close price, to the prediction of mid-price-1 and mid-price-2. Both close-from-close and OHLC feature sets are used for all targets, to compare the relative predictive value each of these feature sets provides.

The close-from-close predictions were poor for both FGBL and EURUSD markets, with simple directional accuracy (defined as the mean precision score reported across both up and down classes) of 50.93% (MCC: 1.80%) and 49.95% (MCC: -0.08%) respectively. However, it should be noted that the poor performance derives primarily from the use of close price as a target rather than as a lagged input. Replacing the target at t+1 by either of the mid-prices, but retaining the simple close lag as input, results in an immediate and large improvement in directional accuracy, with an accuracy of 69.21% (MCC: 38.44%) and 73.34% (MCC: 46.70%) for mid-price-1 and mid-price-2, respectively, on FGBL. EURUSD markets exhibited similar behaviour with simple directional accuracy results of 70.16% (MCC: 40.32%) and 75.35% (MCC: 50.69%) for mid-price-1 and mid-price-2 respectively. It is thus possible to predict a mid-price to a high accuracy while continuing to use traditional close or OHLC price lags as features for FGBL and EURUSD, which is a remarkable baseline result given the noise content in close price returns and the fact that these predictions are made on some of the most liquid and popular financial markets globally.

Further to this result, it can be observed that OHLC lagged inputs increased predictive power (beyond those from close lagged inputs) by a further 13% for mid-price-1 and 7.64% for mid-price-2 on FGBL but actually reduced the model's predictive power for FGBL close prices. Improvements can also be observed for the mid-prices in the case of the EURUSD market, with mid-price-1 predictive power increasing by 12.5% and mid-price-2 increasing by around 1% when OHLC inputs were used. For prediction of close price for EURUSD, however, in contrasts to the result for FGBL, the use of OHLC inputs did give an improvement over close inputs.

Table 4.2 compares the same sets of input features and targets as Table 4.1 but on cryptoasset markets BTCUSD and ETHUSD. A similar pattern is again observable for these assets as for traditional markets, where a traditional feature formulation using close price lags as inputs to predict future close price directional movement performs poorly, but the use of the same inputs to predict the mid-prices performs very much better. Simple directional accuracy scores of 51.58% (MCC: 3.10%) and 49.67% (MCC: -0.63%) were achieved on BTCUSD and ETHUSD respectively when close price lags were used as inputs. The BTCUSD performance results are enhanced by 27.04% for mid-price-1 and 41.41% for mid-price-2, resulting in a simple directional accuracy of 65.53% (MCC: 31.08%) and 72.94% (MCC: 45.33%) for mid-price-1 and mid-price-2 respectively. The same performance improvement pattern is observable on ETHUSD with enhancements of 33.55% for mid-price-1 and 48.38% for mid-price-2, with simple directional accuracy of 66.34% (MCC: 32.64%) and 73.71% (MCC: 46.23%) being achieved for mid-price-1 and mid-price-2.

The results from Table 4.2 follow the same enhancement pattern as observed in Table 4.1 for EURUSD, with improvements being observed on all target variables when OHLC lags are used instead of close lags. The most significant BTCUSD enhancement was observed for mid-price-1 for which OHLC increased predictive power by 6.06%, to achieve a simple directional accuracy of 69.50% (MCC: 38.97%). The same pattern is observable for ETHUSD, which also saw the largest simple directional accuracy improvement for the mid-price-1 target, which increased by 7.99% to 71.65% (MCC: 43.24%). However, for both BTCUSD and ETHUSD, the highest predictive power is achieved through prediction of the mid-price-2 target. An interesting point to note when looking at the results for both the traditional and cryptoasset markets is that the classification metrics are well balanced and remain relatively stable with minimal imbalances, on average, towards either the up or down class. This result confirms that the models have been robustly trained and indicates performance results should be relatively agnostic to market state.

It can be seen from the tables above that using open, high, and low (OHL) lagged inputs in addition to the close input has only a very small effect on the LSTM network's ability to predict close direction; this may well explain why many traditional candlestick patterns appear not to be predictive (Marshall et al., (2006; 2008); Horton (2009); Fock et al., (2005)). There is however a more noticeable improvement in mid-price-1 and mid-price-2 predictions when additional OHL lagged inputs are used; this suggests that mid-price-2 predictions (the most effective) might be improved further by a more intelligent selection of OHLC based features.

4.4.2 Baseline feature lag optimisation

At this point only two lags have been considered. However, the number of lags of OHLC feature data could have an impact on predictive power and certainly has an impact on the complexity of the model (with fewer parameters being preferred). As mid-price-2 is the highest performing target from the previous section it was decided that the lag optimisation would take place with this target for all the markets considered in this study.

To investigate the optimal number of lags for the prediction of the mid-price-2 target, the training and evaluation process is run ten times for each instrument. Each time the process is run, an extra lag is included in the feature space. The results of this optimisation are for an out-of-sample test set (see Section 3.2.3 for data ranges) are shown below in Figure 4.1. Interestingly, the traditional asset classes are easier to predict, with higher predictive performance observed over most lag settings. This is likely due to there being less noise in the price series of traditional instruments compared to the highly volatile cryptoassets.

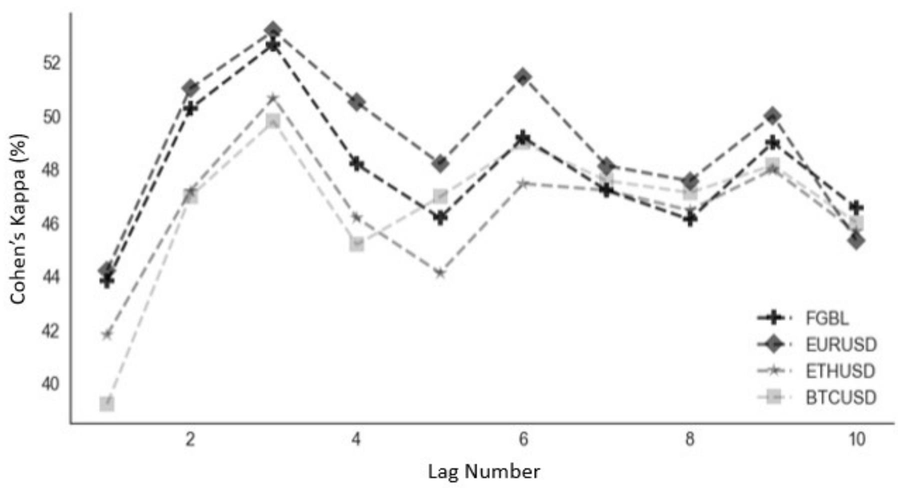


Figure 4.2: Out-of-sample feature lag optimisation results for the mid-price-2 target.

As can be observed in Figure 4.2, a lag of three provides the optimal predictive power across all instruments, achieving marginally enhanced improvements of around 5% for each instrument when compared to the originally chosen lag of two. As the optimisation results exhibit the same improvement patterns across all instruments and retain an exceptionally high level of predictability for a financial timeseries quantity, one can deduce that the predictive power of the mid-price target is not simply an artefact of one dataset, but generalisable across a universe of vastly different instruments. This is particularly interesting from a trading perspective as one can rely on the robustness of such a prediction methodology when devising trading strategies which utilise mid-price predictions. This level of robustness is rarely seen with traditional targets such as the close price due to a high noise content being encapsulated within the target thus demonstrating the unique power of the novel mid-price construction presented here.

There is an interesting connection between the results of this section and observations that can be made about candlestick patterns. In Chapter 3 we trained and optimised each neural network to predict the close price target based on a specified number of feature lags. This was done to align with common definitions of candlestick patterns, of which the most popular ones only contain three or four candlesticks. Figure 4.2 shows that Cohen's kappa has peaks at three, six and nine OHLC lags for all assets. However, the overall maximum is reached at three, implying three lags of OHLC data is optimal in this context, and many candlestick patterns are created from three lags of OHLC data, such as the Three Line Strike, explored in the previous chapter.

4.4.3 Use of mined OHLC features as inputs

Table 4.1 and 4.2 suggested the possibility that suitably configured OHLC data might further enhance the mid-price-2 prediction. In the experiments below mined data, as described in Section 3.3, were used, where the term Importance Mining in Table 4.3 refers to results using the top 100 importance-ranked factors, and correlation subset (ranked and filtered on a validation set) to a reduced input set with those same features now filtered.

Table 4.3: Feature mining performance results on an out-of-sample test dataset.

Asset	F1	Recall	Precision	MCC	Cohen's kappa
Importance Mining					
FGBL	0.7699 0.7698	0.7702 0.7696	0.7697 0.7700	53.98%	53.96%
EURUSD	0.7723 0.7736	0.7703 0.7757	0.7744 0.7715	54.60%	54.63%
BTCUSD	0.7614 0.7605	0.7627 0.7593	0.7601 0.7618	52.20%	52.26%
ETHUSD	0.7705 0.7690	0.7731 0.7665	0.7680 0.7715	53.96%	53.93%
Importance Mining + Correlation Filtering					
FGBL	0.7768 0.7767	0.7772 0.7764	0.7765 0.7770	55.36%	55.36%
EURUSD	0.7820 0.7815	0.7830 0.7806	0.7811 0.7824	56.36%	56.31%
BTCUSD	0.7683 0.7690	0.7674 0.7700	0.7694 0.7680	53.75%	53.77%
ETHUSD	0.7776 0.7779	0.7773 0.7783	0.7780 0.7775	55.55%	55.49%

It can be seen from Table 4.3 that the feature importance mining methodology improves the LSTM RNN's performance, resulting in an increase of MCC scores across all instruments considered. Improvements from the lag optimisation were most substantial for the cryptoasset category with MCC scores increasing by 6.30% and 7.89% for BTCUSD and ETHUSD respectively. This resulted in the models achieving final MCC scores of 52.20% for BTCUSD and 53.96% for ETHUSD. The predictive power of the traditional category was also increased with EURUSD seeing the largest improvement of 4.67% bringing the model's MCC score to 54.60% from the 52.16% achieved in the previous lag optimisation experiment.

In this section it has been shown that mined features can add significant improvements to out-of-sample model performance. However, the improvements due to the addition of correlation-based filtering are somewhat muted relative to other improvements demonstrated in this study. For the traditional asset category, the maximum improvement from correlation-based filtering was observed on FGBL with a final model MCC of 55.36%, up from 53.98%. The cryptoasset category also saw improvements of 2.90% and 2.94% on BTCUSD and ETHUSD, respectively, bringing the final model MCCs to 53.75% and 55.55% for BTCUSD and ETHUSD, respectively. In addition, the optimal values of the correlation thresholds c1 and c2 (see Section 3.3.1) were found to be 0.2 and 1.0, respectively, across all instruments considered, suggesting that these parameters are robust to varying noise levels in different markets. These various observations indicate both that it is the use of the mined features per se that is predominantly leading to the improvement in performance over the optimised baseline model, and also that the LSTM RNN can operate effectively without correlation-based input screening.

Although the relative performance improvements compared to the initial results presented in Table 4.1 and 4.2 were not large, it should be noted that the progressively-improved models were being required to correctly predict in harder and harder market situations in order to achieve incremental performance improvements[2], a task widely accepted to be of extreme difficulty for financial timeseries data. However, despite these observations, a maximum simple directional accuracy of 78.18% (MCC: 56.36%) was achieved on EURUSD using the enhanced Importance Mining + Correlation Filtering methodology. This is a

---

[2] For example, a model may perform well with simple features in a slow rolling market, but poorly in a market where volatility increases as the environment becomes harder to predict. Advanced BOHB optimisation methods also help to discover highly robust parameters with simple features allowing powerful models to be constructed which are robust many scenarios.

remarkable result for the novel mid-price targets, which have proved to be uniquely robust to vastly different markets while maintaining consistently high predictive power, a property rarely observed when predicting any financial quantity.

4.5 Discussion

It has been shown that use of the mid-price (as defined in in Chapter 2, Equation 2.1 and 2.2) as a target can result in up to a 78.18% simple directional prediction accuracy (MCC: 56.36%), observed in this case for EURUSD, using appropriate machine learning techniques. This result is robust to asset class and instrument with all instruments under examination. In addition, all instruments also benefitted from an enhanced feature selection process based on the work of the previous chapter. OHLC data was used to generate candlestick features via an XGBoost importance mining process, which increased the predictive power of an LSTM RNN from an initial median MCC of 46.46% (Mid-2 from Close lags) across all instruments to a median of 55.45% MCC, showing that appropriately formatted OHLC data has a high predictive value in relation to the mid-price targets. The median performance increase observed across all instruments from using mined OHLC features, rather than traditional close price lags, is 19.34%, a substantial improvement, implying that the use of contextual information embedded in the OHLC structure extracted using novel feature engineering techniques should be of extreme interest to practitioners and academics alike.

However, it was demonstrated also that mined OHLC data did not significantly increase predictive power when forecasting the traditional close price target, which is in line with Marshall et al., (2006; 2008), Horton (2009) and Fock et al., (2005). Hence the results here, while they may be surprising, are not at odds with the conclusions about candlestick charting drawn in other work. The usefulness of features derived from OHLC data is not in predicting the close price, but predicting the mid-price, which has been neglected in past research.

The discovery of the high predictive power of the mid-price is in itself a significant result given the prevailing sentiment that no aspect of an asset's price behaviour can be predicted substantially above random. It is not immediately obvious how to harness this high predictive power within a trading strategy, as a mid-price prediction is not located at a specific moment in time but only within an interval. However, a trading strategy built around the mid-price is by no means impossible, though it would necessarily require more for its execution than the simple prediction of this value, and the challenge posed by this will be the topic of the last chapter of this work.

Chapter 5

Trend prediction with Bayesian optimised multi-period conviction ensembles

In the previous chapter it was shown that a *mid-price-2* directional change could be predicted with a predictive power of up to 56.36% MCC (56.31% Cohen's kappa) when using a Bayesian optimisation and Hyperband (BOHB) optimised Long Short Term Memory Recurrent Neural Network (LSTM RNN). However, it was noted that it would be difficult to use this predictive power in the context of a trading strategy as a *mid-price-2* prediction is not located at a specific moment in time.

Here, we introduce a novel methodology combining multiple models of the type described in paragraph above, each predicting the directional *mid-price-2* change over a different time horizon to construct an ensemble. It is shown that by applying a Bayesian optimisation weighting scheme to the ensemble's sub-model predictions a predictive power of 73.26% Cohen's kappa can be achieved, an increase of nearly 30% from the previous chapter. To demonstrate the ensemble's utility a market neutral trading strategy is constructed by exploiting an artefact in the training of each LSTM RNN model allowing directional exposure to be sized in line with the ensemble's expected prediction accuracy. This is referred to as the model's *conviction*. The resulting trading strategy outperforms a Buy & Hold strategy while maintaining the consistency of a Cash

& Carry strategy, a popular institutional trading strategy aiming to capture the difference between a futures contract and its spot price.

5.1 Introduction

Machine learning technologies for predicting trends in financial markets have primarily focused on traditional targets such as the close price and the mid-price (defined as the point between the bid price and the ask price). The reported success in predicting these quantities has been limited, as reviewed in Chapter 2, and evidenced in Chapter 3 and 4 of this work. However, practitioners and academics alike have continued to investigate increasingly advanced pattern recognition techniques in the pursuit of gaining an edge against other market participants, through more advanced modelling methods intended to provide a higher predictive power, which ultimately translates into larger financial rewards. Many of the more advanced pattern recognition methods use complex feature and model constructions to extract maximal predictive power from noisy datasets. These approaches have a high level of complexity but even so often result in only marginal outperformance of a random predictor when forecasting traditional price targets.

In this chapter an ensemble of BOHB-optimised LSTM RNN models is constructed to predict directional trend movement defined as the binary (up / down) *mid-price-2* directional change, as formulated in the previous chapter, rather than focusing on the traditional targets. Each BOHB-optimised LSTM RNN predicts a different horizon of a *mid-price-2* directional change, trained using the pipeline presented in Chapter 4. The ensemble's weighted output is mapped to specific horizons between a pre-specified range of possible horizons, with the optimal ensemble being the one that maximises predictive power at a specific target horizon (e.g., 60 minutes). To demonstrate the practical usages of this unique predictive power, a trading strategy is formulated that shows how one can exploit the power of a *mid-price-2* prediction in the context of a trading strategy subject to institutional risk management practices.

The proposed methods of this chapter show that simple, but well-constructed, predictive modelling pipelines can provide a remarkably high level of predictive power rarely observed on financial datasets, while also being able to add significant trading value when appropriate execution methods are considered.

5.2 Background

There have been many academic studies focusing on the power of OHLC data, primarily in the context of candlestick patterns as reviewed in Section 2.2.1. However, there exists no published literature on the predictability of a *mid-price-2* directional change in the context of OHLC data, outside of the work presented in Chapter 4 of this thesis.

The work presented in this chapter extends the findings presented in Chapter 4 to an ensemble setting where multiple models predicting different horizons of the *mid-price-2* are combined through a weighting scheme. As each model is trained using its own feature aggregations, with a granularity specific to that model, the resulting ensemble encapsulates information about many different model expectations over different horizons, using multiple data aggregations. This ensemble construction is itself novel; there is no prior literature relating to ensemble construction in this manner for financial market prediction, to the best of my knowledge.

In this chapter several ensemble weighting schemes are investigated, with Bayesian optimisation (Brochu et al., 2010) found to produce the best results. The results presented are further evidence of the remarkable predictive power which can be extracted from OHLC data structures if exposed to a well-designed machine learning pipeline focused on novel targets.

5.2.1 Ensemble construction

The feature mining, model construction and hyperparameter optimisation process proposed in the previous chapter is used here to train all the ensemble's sub-models independently.

To construct an ensemble of *mid-price-2* models three weighting schemas (equally weighted, Cohen’s kappa weighted, and Bayesian weighted) are investigated. Each method assigns a weight to each of the sub-models. These weights are then normalised to produce the ensemble's output. The equally weighted and kappa weighted schemes are straightforward to implement. However, the Bayesian weighting scheme is more involved. The objective of this scheme is to maximise the ensemble's Cohen’s kappa on a validation dataset. In this work a Gaussian Process (GP) (Rasmussen and Williams, 2006) Upper Confidence Bound (UCB) (Snoek et al., 2012) is used to construct acquisition functions that minimise regret while maximising the ensemble's predictive capabilities. The process constructs a posterior distribution (e.g., GPs) to describe the function that is to be optimised, which then improves iteratively with more observations, resulting in the process becoming more confident about certain regions of the parameter space, which are then explored further.

5.2.2 Data sources and usage

The source of the data for the traditional assets under consideration is the same as in previous chapters. However, for the cryptoassets a different exchange is now used, OKEx, rather than Bitfinex, which was previously used. The reason for this change is Bitfinex's falling behind in the development of their market infrastructure in recent years, and since the start of this thesis. As a result, there is no mature futures market listed on their exchange with acceptable liquidity to interact with, a critical component of the trading strategy presented here. In addition, the number of active Bitfinex users has severely declined, resulting in their spot (and futures) market liquidity and price integrity suffering. Hence, the exchange was changed to OKEx, which provides all the same products as Bitfinex, plus new ones needed for this chapter, with high quality liquidity to interact with. Data ranges for training, validation and testing remain in line with Table 3.1 of Chapter 3.

5.2.3 Trading simulator

A market neutral trading strategy was developed specifically for use in this chapter, to demonstrate how to exploit the power of a *mid-price-2* prediction. The strategy is reliant on placing limit orders at set price levels (the exact methodology is explained in section 5.3.5). To accurately simulate any trading strategy an order book replay mechanism should be implemented, to replay exact market transactions rather than relying on aggregated data, which introduces errors into the simulation process and can ultimately result in false discoveries. Hence it is of extreme importance to have an accurate simulation environment available for testing algorithmic trading concepts.

Here, an ultrafast memory optimised C++ orderbook replay package was developed for the purpose of replaying quote and trade feeds for accurate simulation in an efficient manner. The package can replay any granularity of feed from level 3 (which combines trades and quotes, allowing one to know what orders a trade executed with) to level 1 (which is just a quote feed with the bid and ask prices). The package has been written in optimised C++ to deal with the very large quantities of data available from financial markets (e.g., for one cryptoasset on OKEx there can be tens of millions of updates per day, hence efficiently written code is essential).

5.2.4 Cryptocurrency perpetual futures

The market neutral trading strategy utilised in this chapter uses two instruments to trade with per base currency. For example, the Bitcoin strategy trades are placed on BTCUSD, a spot market, and BTC-USDT-SWAP, a linear perpetual futures market (Tse, 2020).

The objective of a perpetual future is to provide a derivative market for illiquid assets which are anchored to their spot prices, or fair values. It was first introduced by economist Robert Shiller in 1992. The product has rarely been utilised in traditional finance but has received a lot of attention in cryptoasset markets from when it was first introduced in 2016 by BitMEX (Hayes, 2016), due to the high leverage available (up to 100x), low fees, tight spread (difference between the bid and ask price) and lack of expiry (the contract never expires) making it easy to manage positions.

As cryptoasset markets typically suffer from fragmented liquidity (the same asset can be bought and sold at slightly different prices on different exchanges) the perpetual future is not anchored to the spot price of a specific exchange but rather to a basket of prices from different exchanges, to help stabilise the pricing mechanism (usually an equally weighted average of the index basket) and protect against manipulation. To ensure the perpetual futures contract remains anchored to this index a funding rate is built into the contracts mechanics to incentivise market participants to take contrarian positions to the market trend and keep the perpetuals price close to the spot price. Every N hours the market participants holding positions on the perpetual contract will either pay or receive an amount equal to the difference between the average index price and the perpetual futures price. The cash flows (referred to as the funding rate) do not come from the exchange but rather the participants who hold positions in the contract. Perpetual futures contracts have several unique mechanics summarised in Table 5.1, on the following page.

As can be observed from the table, there are many variables to consider when trading perpetual futures contracts, from different payoff dynamics that will impact the cost of hedging to funding rate frequency and pay-out currency. Linear contracts are more typical in traditional financial markets, whereas inverse contracts are very popular in cryptoasset markets as there is no requirement to interact with a regulated currency (e.g., USD) as all interactions take place using the cryptocurrency (e.g., BTC). A full review is outside the scope of this work although a more detailed review is available from multiple authors (Shiller, 1993; Alexander et al. 2021).

Table 5.1: A review of perpetual futures contract mechanisms.

Mechanic	Description	
Funding Rate	As discussed above, this is equivalent to the average displacement between the index and contracts price.	
Funding Frequency	Most cryptocurrency exchanges choose a funding rate of 3 payments per day (every 8 hours). There are some exchanges which offer payments at different frequencies.	
Funding Sign	If the funding rate is positive (perpetual futures contract price is above the index price), the longs pay the funding fee to the shorts. If negative (perpetual futures contract price is below the index price), the shorts pay to longs.	
Contract Types	There are two popular types of perpetual contract available, referred to as *Linear* and *Inverse*, with subtle differences in how position p&l is calculated, and funding payments are delivered.	
Example	**Linear**	**Inverse**
Quoted	BTC/USD	BTC/USD
Contract Value	0.1 BTC	10 USD
Delta*	1	$1/Spot$
Gamma**	0	$-1/Spot^2$
P&L	$BTCPositionSize * (ExitPrice - EntryPrice)$	$USDPositionSize * \left(\frac{1}{EntryPrice} - \frac{1}{ExitPrice}\right)$
Payments	USD	BTC

*Delta: The rate of change of the contracts price to the index price.
**Gamma: The rate of change of the contracts delta to the index price.

5.2.5 Market neutral futures strategies

One of the most popular institutional market neutral trading strategies is called a *Cash & Carry.* This strategy aims to generate profits by exploiting the difference between a futures price and a spot price, the difference often being referred to as the *basis.* (Thus, the trade is also sometimes simply referred to as a Basis trade.) The strategy is profitable because a futures market must expire at the spot price, or some average of the spot price leading up to the contract expire, allowing traders to lock in the basis as profit. There are two main states of a futures market, as will be explained. The state of *contango* is when the futures price is above the spot price whereas the state of *backwardation* is when the futures price is below the spot price. Institutions favour the Cash & Carry strategy as profits can be generated from financial markets without taking any directional risk.

For a profit to be generated from the basis, trades must be executed in a way that depends on the market's state, in order to maintain market neutrality. These trades are summarised below in Table 5.2, where it should be noted that all positions are closed upon a futures contract's expiration.

Table 5.2: Cash & Carry trades per market state.

State	**Spot Market**	**Futures Market**[3]
Contango	Buy (Long)	Sell (Short)
Backwardation	Sell (Short)	Buy (Long)

[3] A subtle point to note is that the dollar quantities placed on each market differ. The spot market is simple, as a trader simply needs to buy one unit of the asset, denoted by some dollar amount, for example $100. To properly hedge this trade and capture the basis the futures market trade must also price in the basis. For example, if the basis is 10% and the market is in contango then a trader would buy the spot market with $100 and short the futures market with $110 (i.e., the equivalent of 1 unit long and 1 unit short).

The strategy above works well on calendar futures markets (in which contracts expire on a specific date) as there can in this case be a large difference between futures and spot prices ('basis'). However, the work of this chapter is focused on perpetual futures (this type of contract being explained in section 5.2.4), which have little to no difference in their price compared to the spot market, and never expire. As explained in section 5.2.4, the perpetual futures contract makes payments at regular periods (for example, every eight hours) to compensate traders for anchoring the futures price close to, or at, the spot index price. Therefore, by executing the same trades as explained in Table 5.2, in the states of contango or backwardation the perpetual basis trade will earn income from these regular payments rather than profit from the price differences. This type of Cash & Carry trade is commonly referred to as a *Funding Capture* and is one of the most popular institutional cryptocurrency trades to execute. The reason for its popularity is that there is no directional exposure taken but the trade can even so return consistent, and sometimes large, profits due to the highly volatile nature of cryptoassets resulting in large displacements of the perpetual futures market relative to the spot market. A Funding Capture strategy is characterised by frequent gains and high Sharpe ratios if the correct positioning has been executed.

5.2.6 Evaluation metrics

All metrics used here to evaluate the predictive performance of models are as in the previous chapter, with their formal definitions stated in Section 2.1.5. However, the evaluation of a trading strategy has not been addressed in the previous chapters. Therefore, important metrics to consider when assessing the performance of a trading strategy are detailed below in Table 5.3.

Table 5.3: A review of trading strategy evaluation metrics.

Metric	Description
Winning Days	The ratio of winning days to total days as a percentage. A large value is favourable.
Mean Daily Return	The average return that a strategy earnt per day. A large value is favourable.
Longest Drawdown	The maximum number of days a strategy failed to match its previous high return. A small value is favourable.
Max Drawdown	The maximum percentage loss from a strategy's peak return to its trough, before generating a new peak. A small value is favourable.
Sharpe Ratio	A strategy's average return as a proportion of the strategy's return standard deviation. The metric is often annualised by multiplying the value by the square root of the number of days the market is open for trading (in crypto markets this is 365 days a year; in traditional markets this is around 252 depending on whether it is a leap year and specific market nuances). A large value is favourable.
P&L	A strategy's total return over the full period it was deployed. A large value is favourable.

5.3 Methodology

The overall process used in this chapter, including the implemented trading strategy, is summarised (for a given asset) below:

Table 5.1: Process flow for trading from trend prediction.

construct ensemble model

select data (section 5.2.2)

for each of n models, each with a different prediction horizon:

- prepare the data: convert to OHLC format with time period equal to the horizon
- generate the feature universe (section 5.3.1)
- identify a powerful set of features using XGBoost and correlation filtering (section 5.3.1)
- use selected features to predict mid-price-2 (section 5.3.2)

calculate weightings of the models in the ensemble (section 5.3.3)

trade with the ensemble model (sections 5.2.5 and 5.3.5)

optimise trading interval (section 5.4.2)

at each such interval:

- make a prediction using the weighted ensemble model
- assess the confidence level of this prediction (see discussion and explanation in section 5.3.4)
- use this as part of the input into a trading decision, following the steps outlined Table 5.4

To construct the ensemble, n models are required to be trained, using the methodology presented in Chapter 4. Once the models have been trained individually a weighting scheme is applied to their outputs to create the ensemble. In this work sixteen sub-models are trained on different horizons from 15 minutes to 4 hours in steps of 15 minutes (i.e., 15, 30, 45, …, 225, 240). As in Chapter 3, example charts are included in this section to help demonstrate the process. For final results see the Results section.

5.3.1 Feature mining and selection

Figure 5.1 shows that each sub-model of the proposed ensemble has a similar feature ranking pattern with all of them flattening around N=100, after which a gradual decline in importance is observed. This is the same N setting as was proposed in previous chapters for different horizons, reinforcing that this setting is robust to model horizon (and asset, as similar patterns were observed for all assets under consideration in this chapter).

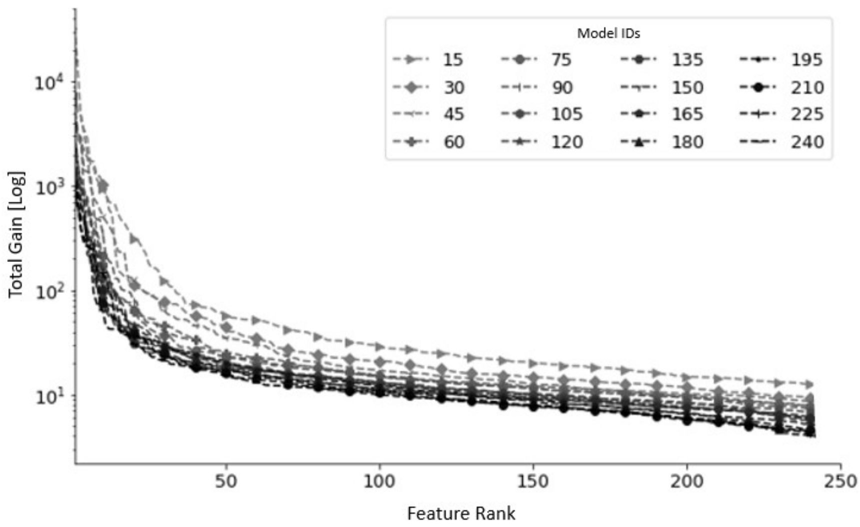


Figure 5.1: Feature importance rankings of all sub-models for the BTCUSD ensemble.

The selected N=100 features per model are then exposed to the correlation filtering framework, as detailed in the previous chapter.

5.3.2 Multiple mid-price-2 directional predictions

Again, the methodology implemented to train each individual model follows that of the previous chapter. The difference here is that instead of one model there are sixteen models, each trained on a different horizon.

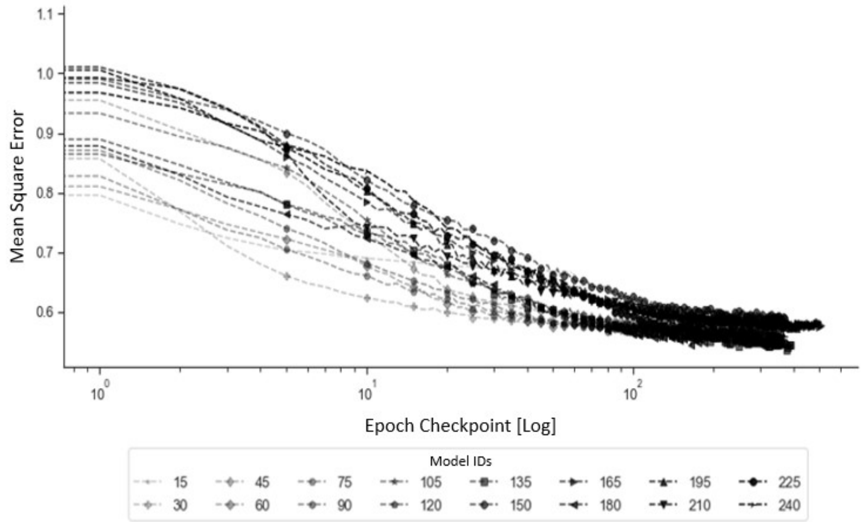


Figure 5.2: BOHB-optimised LSTM RNN validation loss per model horizon for BTCUSD.

Figure 5.2 shows that each sub-model was successfully trained to provide a stable validation loss across its relevant horizon. This is of particular interest to note, as often when the target horizon increases the variance in the predicted target label also increases. It is shown here that, in contrast, the proposed methodology can produce robust models at any of the horizons considered, a favourable characteristic to observe, especially if the predictions are to be used for trading as one can allocate capital confidently.

5.3.3 Ensemble weighting scheme

Three ensemble weighting schemas (equally weighted, Cohen's kappa weighted, and Bayesian weighted) will be considered for the construction of the ensembles. An *equally weighted ensemble* results in each model receiving a weight equivalent to $1/n$, where n is the number of models included in the ensemble. A *Cohen's* *kappa weighted ensemble* results in each model receiving a weight equivalent to its Cohen's kappa on a validation dataset, and a *Bayesian optimised weighted ensemble* allocates a value between [0,1] to each of the models in the ensemble, where the objective is to maximise the ensemble's Cohen's kappa on a validation dataset. The output of the chosen weighting scheme is then normalised using the following formula:

$$\frac{\sum_{i=0}^{n} w_i m_i}{\sum_{i=0}^{n} w_i}, \tag{5.1}$$

where i is a model index, n is the total number of models included in the ensemble, m_i is model i's output and w_i is the weight assigned to the sub-model. Bayesian optimisation was selected, as opposed to a simple grid-search, due to the high dimensionality of the optimisation problem.

5.3.4 Ensemble conviction mapping

Each of the model outputs (LSTMs, optimised using the BOHB process of section 2.1.4) is assessed by adding the output to the relevant bucket (buckets being of size 0.1, stretching from -1 to 1, the output range of the model predictions, and hence 20 buckets in total), taking all predictions that reside in each bucket and assessing the simple accuracy of that bucket. The hypothesis to be tested here is that output values near the limits (close to either -1 or 1, the ends of the bucket range) will be more accurate than output values near zero (in the middle of the bucket range). If this were true, the resulting shape of the accuracy as a function of the bucket label (referred to in Figure 5.3 as 'Network Output') should resemble a parabola, and this does, in fact, appear to be reliably the case, as evidenced in Figure 5.3.

As the example in Figure 5.3 below shows, the remarkable similarity of the bucketed outputs (right) to the hypothetical parabolic mappings (left) does, in fact, allow a higher confidence to be allocated to a prediction when a model outputs a value tending towards the limits of its output range. *We name this property the model's 'conviction mapping'.*

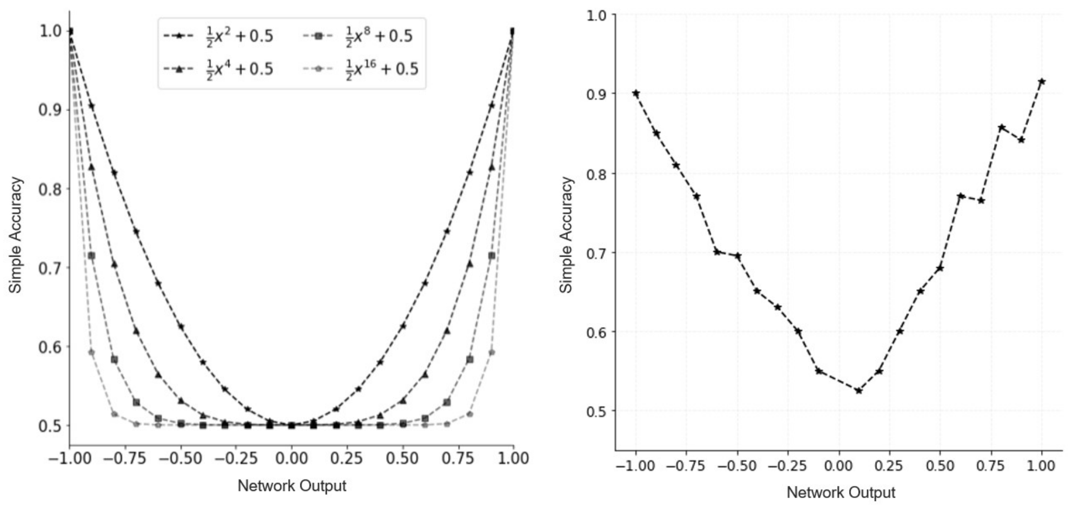


Figure 5.3: [Left] Different ideal hypothetical outputs drawn from parabola functions; [Right] Example 15-minute predictive model output on a validation dataset.

In this chapter the output of an ensemble of models will be used to inform a trading strategy. As discussed and evidenced above, when the output of a component model is near the limits of its range it can be assumed that its prediction is highly likely to be correct and thus capital should be aggressively assigned in line with the prediction. It is found that for each sub-model in each asset's ensemble a robust conviction mapping is embedded over all horizons, although some are more stable than others.

Examples from each sub-model of the BTCUSD ensemble (see Figures 5.4 to 5.7 below) are selected to demonstrate that conviction stability persists over different long- and short-term horizons and that predictive results are robust to the train, validation, and test datasets.

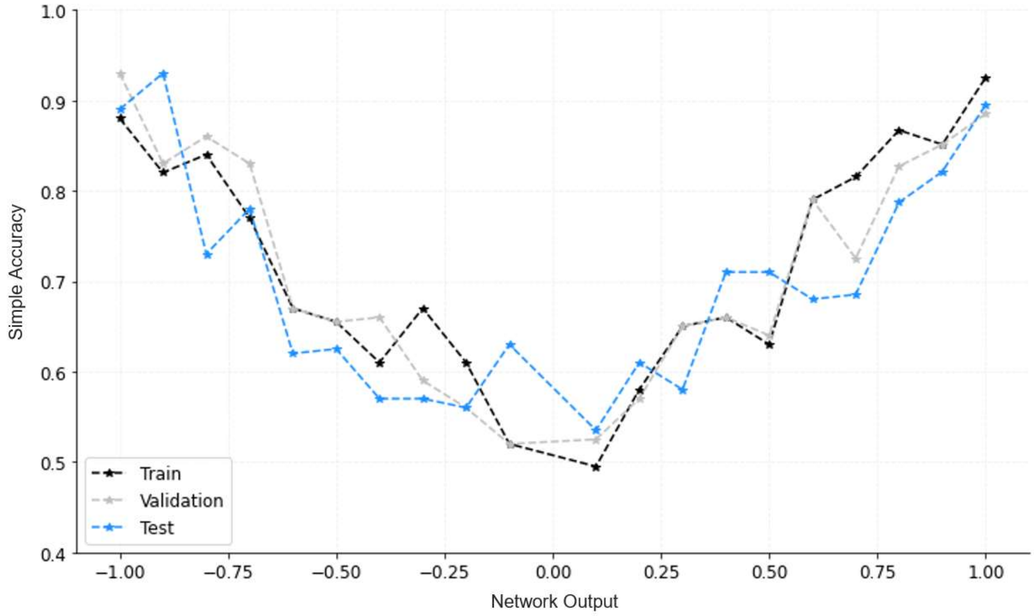


Figure 5.4: Conviction mapping for the 30-minute BTCUSD model.

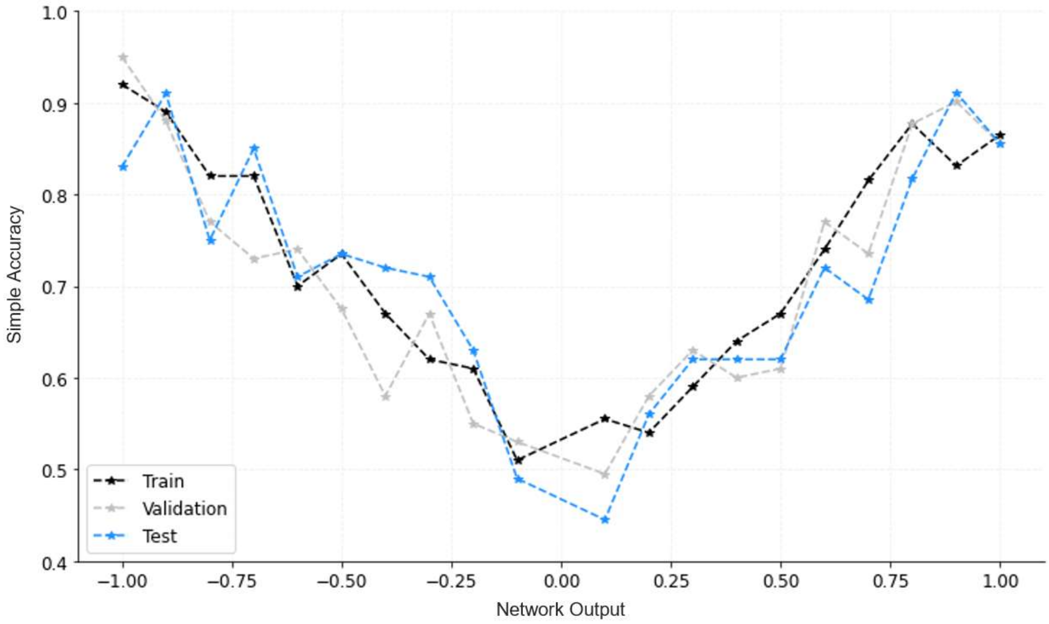


Figure 5.5: Conviction mapping for the 60-minute BTCUSD model.

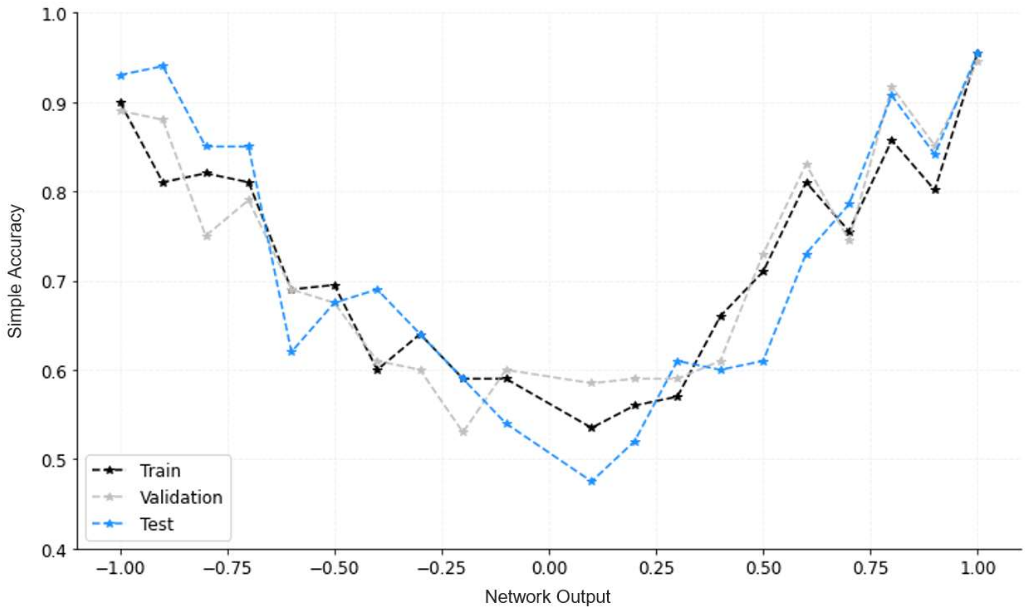


Figure 5.6: Conviction mapping for the 120-minute BTCUSD model.

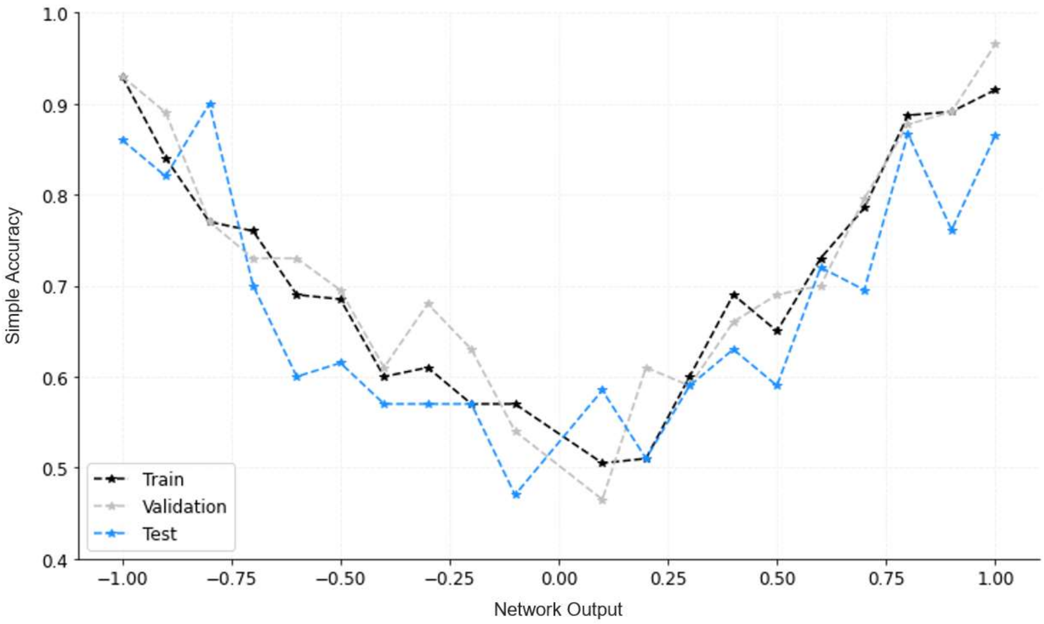


Figure 5.7: Conviction mapping for the 240-minute BTCUSD model.

5.3.5 Trading strategy

In this work a novel twist on the popular Funding Capture strategy is proposed by allowing restricted directional exposure to be opened. This is done by reducing position size on either the spot or perpetual futures market within a pre-specified risk budget (e.g., a risk budget equal to no more than a 1% directional exposure) conditional on a conviction ensemble's prediction and the market state. Reducing the position size on either of the above markets has the net effect of opening a directional exposure while still allowing the baseline strategy to earn an income from the funding rates. The strategy is designed to satisfy an institutional style of trading with flexible limits on the amount of directional exposure that can be taken, referred to above as the risk budget.

To help explain the *directional mechanic* further a hypothetical trade is detailed. Assume the conviction ensemble is predicting an upward movement, and a trader is $500 short on the perpetual contract and $500 long on the spot market positioned for the contango state. To benefit from this prediction the trader will reduce the perpetual contract position size from $500 to $250, which will open the equivalent of a 25% long directional exposure while leaving 25% short on the perpetual contract, which will continue to earn the funding rate. Assuming the prediction of 'up' is correct, and the next period happened to be a funding rate payment, then the trader would receive p&l from the long directional exposure and the funding rate paid to short positions. Hence, the proposed strategy exploits multiple sources of income in a controlled manner, allowing a trader to manage risk effectively, exploit the power of a *mid-price-2* prediction and continue to benefit from the baseline funding rate.

The above explanation of the strategy's directional mechanic does not consider the order *placement mechanic* which is now explained. As a *mid-price-2* prediction does not occur at a specific moment in time the execution of opening or closing a directional exposure is critical to the success of the strategy. The execution placement rules are detailed in Table 5.4, on the following page.

An example order placement scenario is shown in Figure 5.8. For this example, it is assumed the market is in contango, the prediction is for an upward move, and conviction is greater than 0.5. At $t0$ a limit order is placed on the bid side of the perpetual future at the previously realised $MidPrice2_{t-1}$. This has the effect of buying back a portion of the short position resulting in long directional exposure.

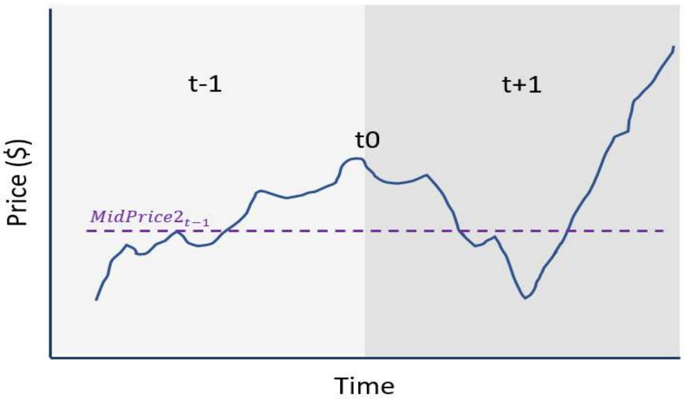


Figure 5.8: A hypothetical limit entry order placement.

Using the same example as above, if the prediction happened to be for a downward movement, then an order would be placed on the ask side of the spot market, selling a portion of the spot position, resulting in an opening of short directional exposure. As the market is already above this point (giving favourable execution) the order would simply be placed at the best ask price and held there until filled. The strategy re-balances directional exposures at each discrete prediction period.

Table 5.4: Trading and execution logic review.

Rule	Detail	Reason
Opening	Placed via limit orders at or around the previous *mid-price-2*.	Directional risk is only opened if orders can be executed around the previous *mid-price-2* as this aligns the powerful prediction target with prices that can be executed at.
Placement*	For every 1% reduction from perfect conviction the limit order price will deepen by the minimum tick size of the market (BTCUSD: $0.1; ETHUSD: $0.01).	The trading strategy becomes less aggressive as uncertainty increases resulting in a reduced probability of a bad order being filled.
Sizing**	Orders are sized as a function of ensemble conviction. Sizing is computed as: $conviction \times risk\ budget$.	More dollars of the assigned risk budget are utilised as directional confidence increases.
Thresholding	Order placement will only be activated when the conviction ensemble breaches 0.5 for long or -0.5 for short.	Adds further protection against bad predictions in line with expected model accuracy.
Closing	Closing orders are placed aggressively via market orders.	Protects strategy gains against uncertainty with immediate effect.

*All price levels are calculated from the spot market rather than the perpetual futures market for ease as the perpetual contract will often be closely anchored to the spot price and there is more data available for spot to train the models on.
**If a risk budget has been previously assigned and there is an increase of conviction then the sizing will be the difference between current and desired position size.

The strategy presented in the results section of this chapter specifies very conservative directional budgets of 1%, 5% and 10%. This is to keep the strategy closely in line with a market neutral mandate while demonstrating how much improvement can be made when allowing small directional exposures to be opened, if controlled through a *mid-price-2* conviction ensemble. Note that the funding rate and directional p&l is continually invested.

The strategy is deployed on the cryptocurrency markets due to high frequency data being available for simulated trading, allowing limit order placement logic to be modelled, due to the cryptocurrency asset class becoming very popular, and to demonstrate the power of a *mid-price-2* directional prediction in the context of a popular trading concept.

All transaction costs[4], spread costs and market impact costs are accounted for in the trading simulation. Adverse selection is not accounted for, although would be an interesting iteration to the strategy's order placement logic.

5.4 Results

In this section individual model performance results are presented first. These are then used in the construction of the conviction ensembles. The ensembles are then used within a trading strategy, to demonstrate their practical value.

[4] The highest level of transaction cost is used for the simulation. On OKEx this is set at 0.08% for spot limit orders, 0.1% for spot market orders, 0.02% for perpetual limit orders and 0.05% for perpetual market orders.

5.4.1 Individual model performance

To help assess the power of the proposed conviction ensemble methodology each sub-model's performance is first investigated, providing a baseline benchmark in relation to the ensemble performance.

The results for the individual models are shown in Table 5.5 on the following page for each asset on both validation and test data sets, where the ID column represents the model's horizon in minutes (e.g. an ID of 15 represents a 15 minute forecast horizon). It can be observed that very high Cohen's kappa scores (representing percentages better than random) are achieved at all horizons and for all assets under consideration. These performance results are rarely observed when predicting financial quantities due to the noise content in financial data. It is thus shown that the power of a *mid-price-2* prediction can scale and be robust to multiple short- and long-term horizons over multiple asset classes.

Table 5.5: A **Cohen's kappa** comparison of all ensembles' sub-models per horizon and asset.

ID	Validation	Test	ID	Validation	Test
			BTCUSD		
15	55.45%	54.99%	**135**	50.97%	50.09%
30	54.86%	51.36%	**150**	51.54%	52.32%
45	55.18%	51.37%	**165**	56.06%	50.63%
60	52.36%	54.40%	**180**	49.38%	50.81%
75	47.33%	51.19%	**195**	47.18%	54.09%
90	49.64%	52.08%	**210**	52.76%	53.80%
105	55.76%	54.84%	**225**	49.73%	49.22%
120	50.25%	48.75%	**240**	55.87%	53.16%
			ETHUSD		
15	55.25%	51.08%	**135**	50.72%	53.67%
30	54.44%	48.69%	**150**	54.73%	51.53%
45	51.64%	52.82%	**165**	56.12%	50.12%
60	54.66%	56.01%	**180**	53.39%	55.08%
75	55.17%	52.71%	**195**	54.09%	54.02%
90	52.30%	49.79%	**210**	51.75%	55.53%
105	52.52%	51.13%	**225**	48.60%	47.89%
120	52.92%	54.78%	**240**	56.05%	53.88%
			FGBL		
15	55.01%	50.86%	**135**	50.50%	53.44%
30	54.21%	48.49%	**150**	54.50%	51.32%
45	51.42%	52.59%	**165**	55.87%	49.91%
60	54.43%	55.78%	**180**	53.16%	54.85%
75	54.94%	52.49%	**195**	53.87%	53.79%
90	52.08%	49.58%	**210**	51.53%	55.29%
105	52.29%	50.92%	**225**	48.39%	47.69%
120	52.69%	54.56%	**240**	55.80%	54.64%
			EURUSD		
15	56.78%	52.50%	**135**	52.12%	55.17%
30	55.95%	50.04%	**150**	56.25%	52.96%
45	53.08%	54.29%	**165**	55.68%	53.52%
60	56.18%	57.57%	**180**	54.87%	56.61%
75	56.71%	54.17%	**195**	55.59%	55.52%
90	53.76%	51.18%	**210**	53.19%	57.08%
105	53.98%	52.56%	**225**	49.95%	49.23%
120	54.39%	56.31%	**240**	53.66%	55.46%

The highest validation score in Table 5.5 is on EURUSD at the 75-minute horizon, with a Cohen's kappa of 56.71%. The highest individual test score is also observed on EURUSD, with the 210-minute model achieving a 57.08% Cohen's kappa. Confusion matrices relating to these best-performing horizons are shown in Table 5.6 below.

Table 5.6: Confusion matrices for best-performing prediction horizons.

Model	Confusion Matrix		
75 Min Horizon (56.71% Kappa)		*Up*	*Down*
	Up	671	193
	Down	181	683
210 Min Horizon (57.08% Kappa)		*Up*	*Down*
	Up	240	70
	Down	63	247

These results are remarkable as they evidence that long-term predictions can be made regarding a market trend (through the *mid-price-2* metric) with confidence higher than that displayed for many short-term prediction models. In addition to this display of robustness, the average difference in performance between the validation and test datasets ranges from the 2.37% Cohen's kappa for EURUSD to a maximum of 2.58% Cohen's kappa for ETHUSD, giving confidence that results observed on validation datasets can be reproduced on test datasets and in a production setting.

5.4.2 Conviction ensemble performance

Conviction ensembles are assembled by selecting a weighting scheme to apply to a universe of individual models which produce the ensemble output. The result is a model which is informed by information extracted from multiple horizons and multiple feature granularities, a very powerful combination which provides important contextual information to the ensemble. It is thought that this contextual information extracted at multiple granularities is one of the main reasons the ensemble outperforms the already impressive individual modelling results presented above in Table 5.5.

As has been mentioned earlier, to construct the ensemble three weighting schemes (detailed in section 5.3.3) will be tested, applied to each sub-model universe per optimisation. Each model will receive a weight dictated by the selected scheme to its last available output, with the combined output then being normalised to ensure the resulting model operates within the same bounds as the sub-models. For each weighting scheme the predictive performance is measured at different horizons ranging from 15 minutes to 4 hours. The optimisation results, for the three weighting schemes, and for each asset under consideration, are shown below in Figures 5.9 to 5.12.

It can be observed from that the optimal horizon is 60 minutes for all asset classes and found to persist over all weighting schemes. The maximum Cohen's kappa appears to be robust to small variations in horizon for all the weighting schemes considered, with the 45-minute and 75-minute horizons, both achieving performance scores in the top 10^{th} percentile of model results.

The best performance is attained using the Bayesian optimisation weighting scheme. This is likely due to the scheme being able to dynamically test different weights for each model as part of the trail-and-error optimisation process, a flexibility the other schemes cannot provide as they are static. As the optimal 60-minute horizon exhibits a shallow peak for all weighting schemes it can thus be assumed the optimal results using any of the three weighting schemes considered are robust to reasonable deviations and not simply an artefact of a specific dataset. This is further enforced by the fact that the optimal horizon scales across multiple asset classes which encapsulate very different market dynamics.

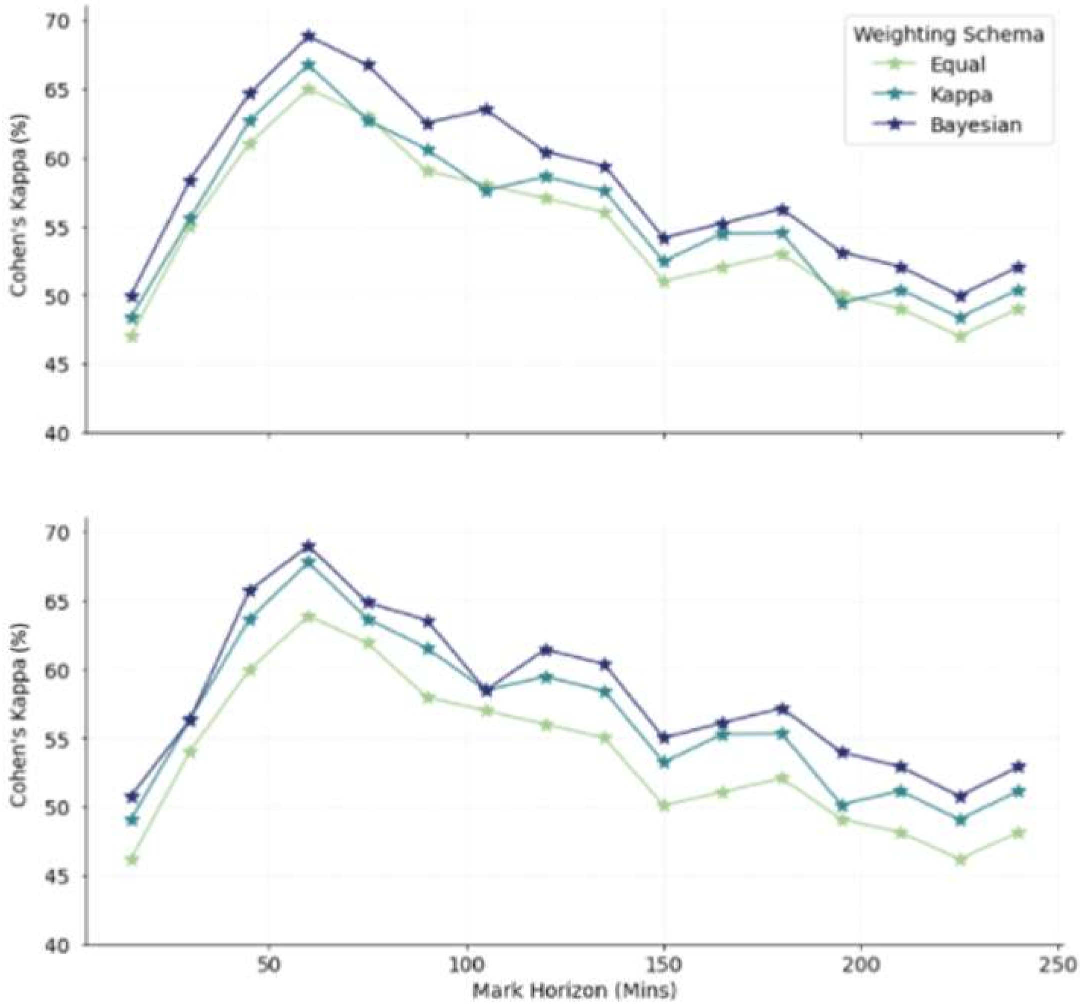


Figure 5.9: **BTCUSD** | [Upper] Validation performance; [Lower] Test performance.

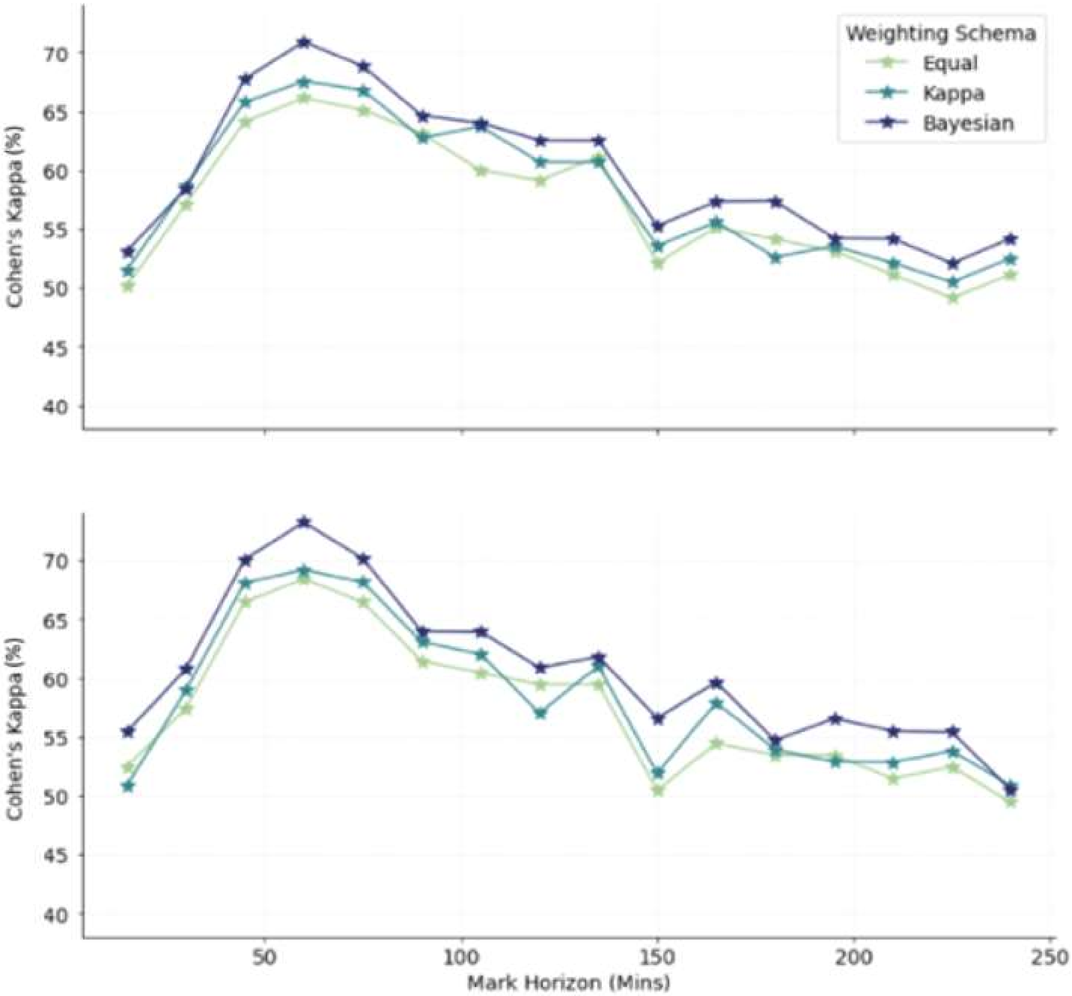


Figure 5.10: **ETHUSD** | [Upper] Validation performance; [Lower] Test performance.

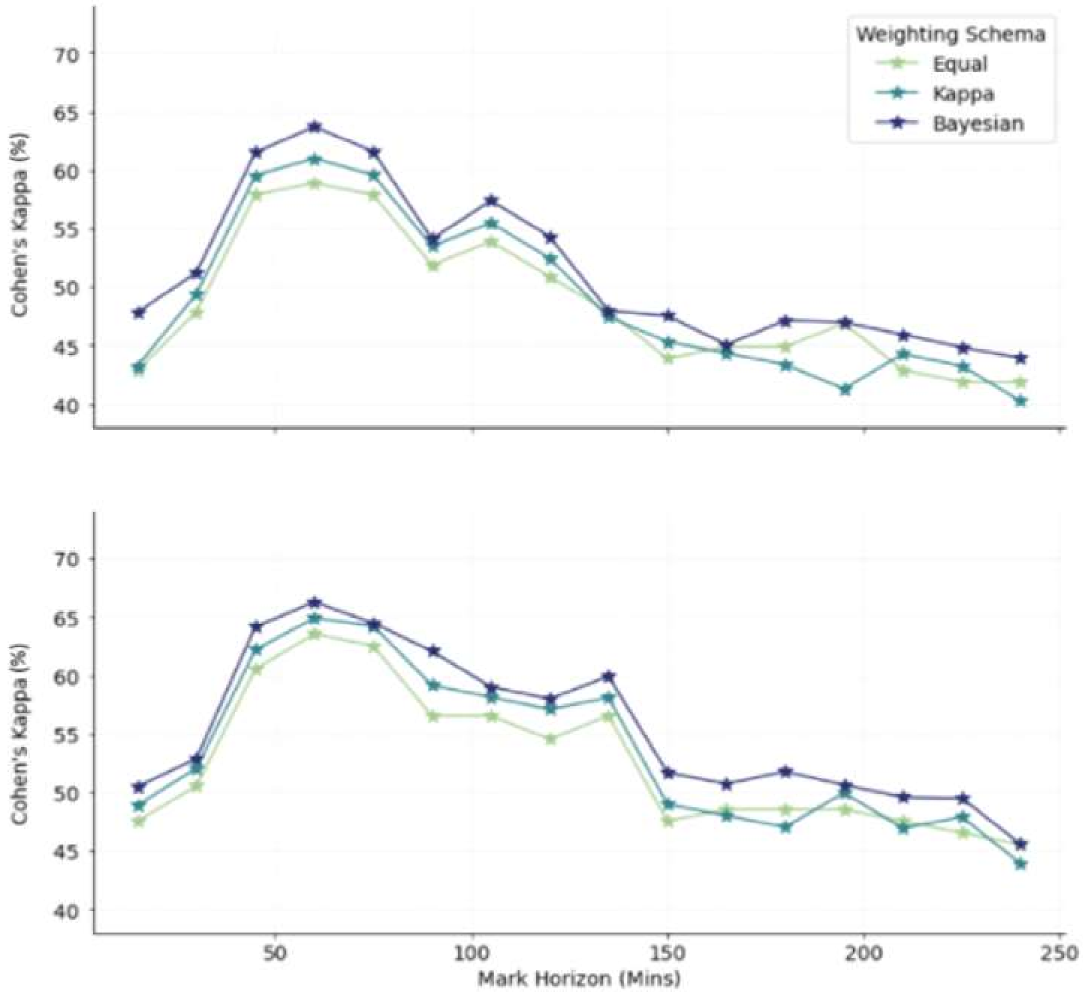


Figure 5.11: **FGBL** | [Upper] Validation performance; [Lower] Test performance.

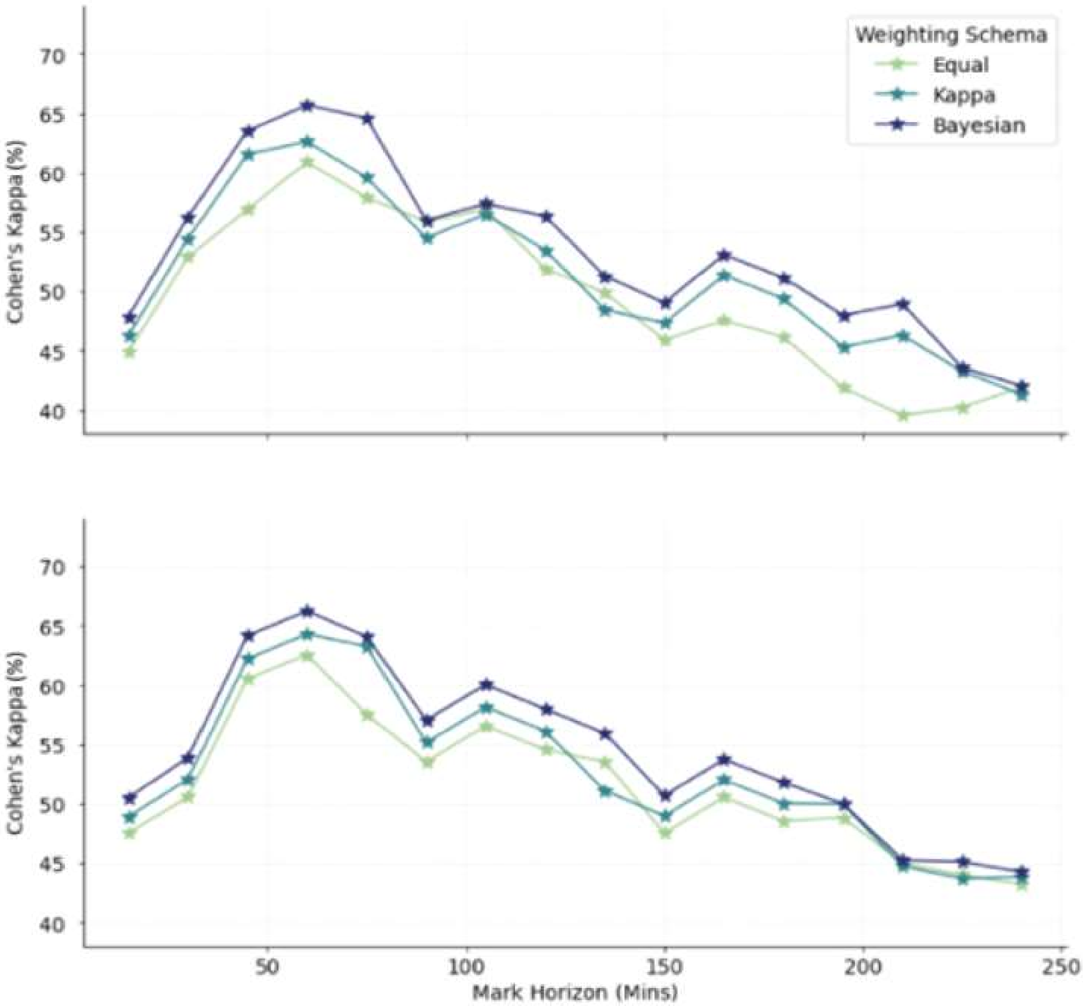


Figure 5.12: **EURUSD** | [Upper] Validation performance; [Lower] Test performance.

The charts above show that the Bayesian optimisation weighting scheme is superior across all horizons, asset classes and weighting schemes on both the validation and test datasets. However, the predictive power is surprisingly robust to variations of weighting scheme, indicating that the ensemble models provide a consistent yet high predictive power and are not reliant on overly complex optimisation schemes which can lead to instability, again demonstrating the unique properties encapsulated in the methodology and novel target construction. Table 5.7 below summarises the optimal performances per weighting scheme and asset, all of which achieved maximum performance at the 60-minute horizon mark.

Table 5.7: A comparison of different weighting scheme performances.

Asset	**Validation Cohen's Kappa**	**Test Cohen's Kappa**
Equally Weighted		
BTCUSD	65.00%	63.84%
ETHUSD	66.13%	68.45%
EURUSD	60.89%	62.53%
FGBL	58.89%	63.57%
Cohen's Kappa Weighted		
BTCUSD	66.74%	67.74%
ETHUSD	67.51%	69.19%
EURUSD	62.59%	64.26%
FGBL	61.00%	64.90%
Bayesian Weighted		
BTCUSD	68.81%	68.94%
ETHUSD	70.92%	73.26%
EURUSD	65.70%	67.01%
FGBL	63.70%	66.26%

The maximum performance observed on a validation dataset was on ETHUSD, using the Bayesian weighting scheme, where a 70.92% Cohen's kappa was achieved. This translated into a remarkable **73.26%** Cohen's kappa on the test dataset which was the highest observed performance over all asset classes and weighting schemes investigated. The mean absolute performance difference (in terms of Cohen's kappa) between the validation and test datasets across all assets was 2.45%, 2.06%, and 1.59% on the equally weighted, kappa weighted, and Bayesian weighted schemes, respectively, showing that there is a robustness to the validation results, as they can be closely reproduced on unseen data, a promising property to observe. Interestingly, the smallest difference was observed for the Bayesian weighting scheme and the largest difference observed for the equally weighted scheme, implying that the extra complexity of the Bayesian weighted scheme was worth paying, given that the results both provide higher predictive power and are more robust.

The predictive models built for the cryptoassets outperformed those for the traditional assets by providing an additional 3.09%, 3.89% and 4.47% Cohen's kappa on average across the equally weighted, kappa weighted and Bayesian weighted schemes respectively. This is contradictory to the individual model performances, where the maximum predictive power was observed for the traditional asset classes. Thus, although multi-period contextual information extracted from multiple data granularities is important when predicting all asset classes, the benefits are most strongly observed for cryptoassets.

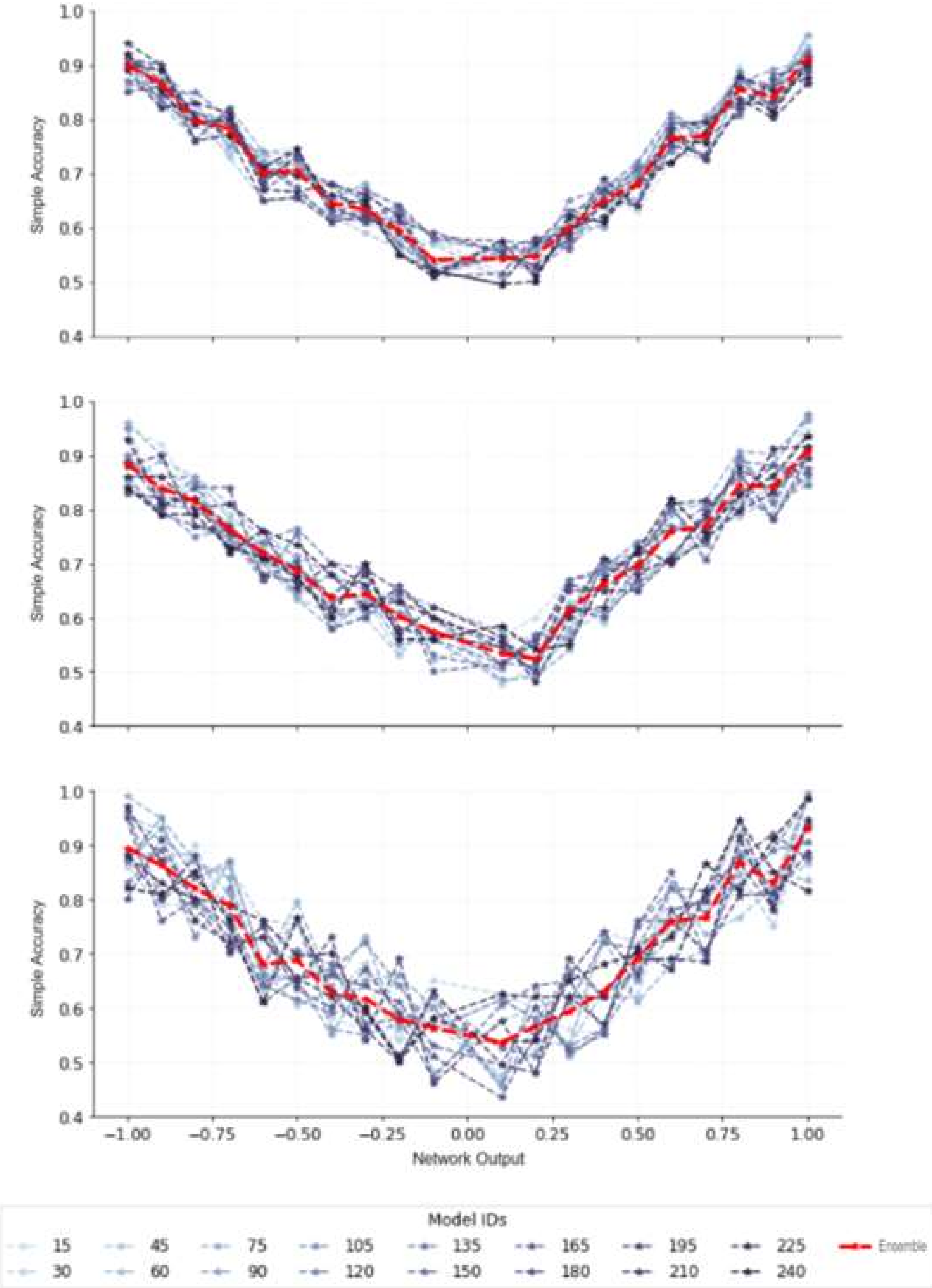


Figure 5.13: **BTCUSD** | [Upper] Train conviction mapping; [Middle] Validation conviction mapping; [Lower] Test conviction mapping.

5.4.3 Conviction ensemble stability

To verify the stability of the ensemble conviction mappings, the train, validation, and test conviction accuracies are plotted for each ensemble sub-model in Figure 5.13, with those for the Bayesian weighted ensembles added to the charts to demonstrate the enhanced stability. The results are presented for BTCUSD (ETHUSD was observed to follow the same pattern).

It can be seen from Figure 5.13 that the tightest distribution of model performances is on the train dataset. This is to be expected as it is an in-sample result. As the models are applied to unseen datasets ranging further from the training dataset, the individual model performance distributions widen. The Bayesian weighted ensemble addresses these concerns effectively by adjusting for individual model instabilities. This stabilising effect is critical to successful usage in real world scenarios such as trading where the output will be used to size and place orders that require stable and accurate predictions. If the output were unstable it would result in overtrading, higher trading costs and less accurate order sizing and placement logic, as these mechanics are a function of model certainty.

An interesting point to note, though outside the scope of this work, is that a model re-training mechanism could be designed as a function of an ensemble's sub-model’s performance variance, which would effectively control for instability propagating through the ensemble's predictions as time evolved.

5.4.4 Market neutral trading strategy

As described in section 5.3.5 a market neutral trading strategy was designed to enhance a classical Funding Capture trade while maintaining the benefits of the baseline strategy’s funding rate income. The results from this strategy presented here use the optimal Bayesian weighted ensemble to inform order placement and sizing of directional exposures every 60 minutes, as this is the horizon where the highest predictive power was discovered. Simulation results using a $100k portfolio are presented below for BTCUSD and ETHUSD.

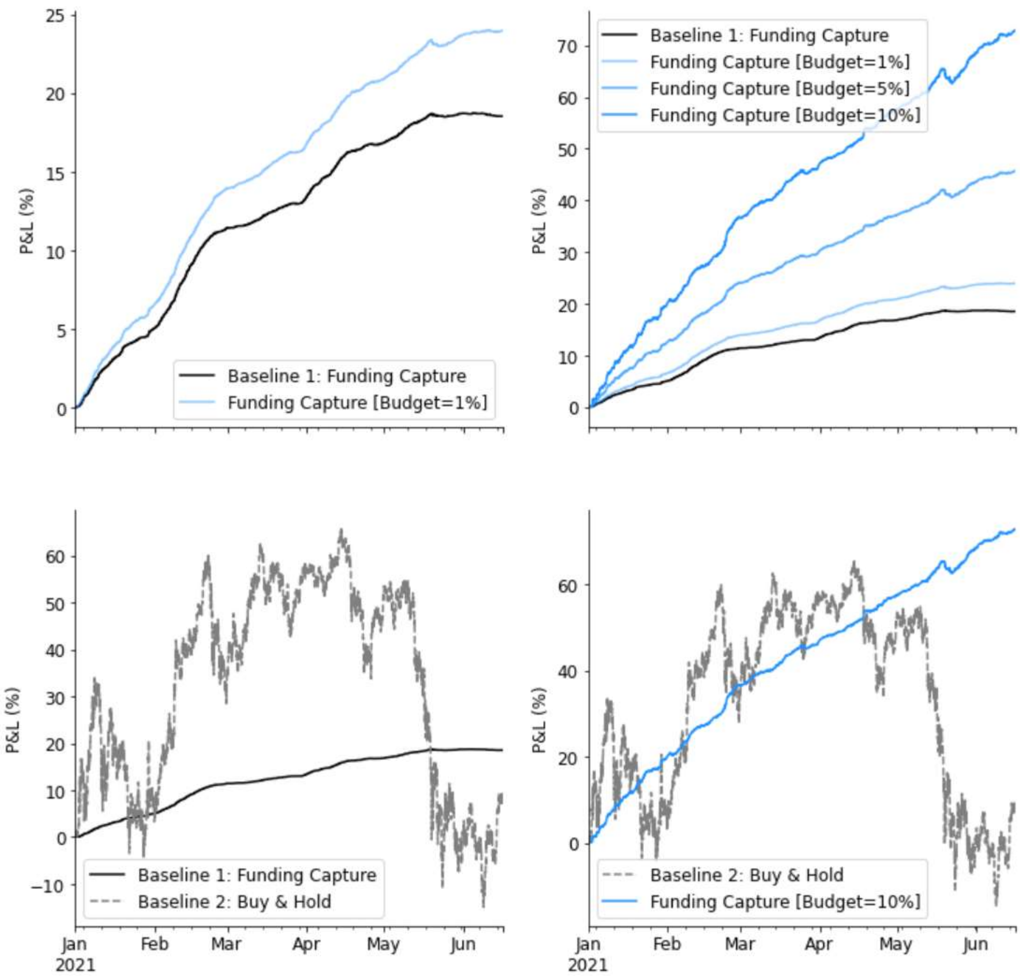


Figure 5.14: **BTCUSD** performance comparison for the different strategy variants, in each case using a Bayesian weighted conviction ensemble to make trading decisions. Upper charts use Funding Capture as the baseline strategy while lower charts use Buy & Hold.

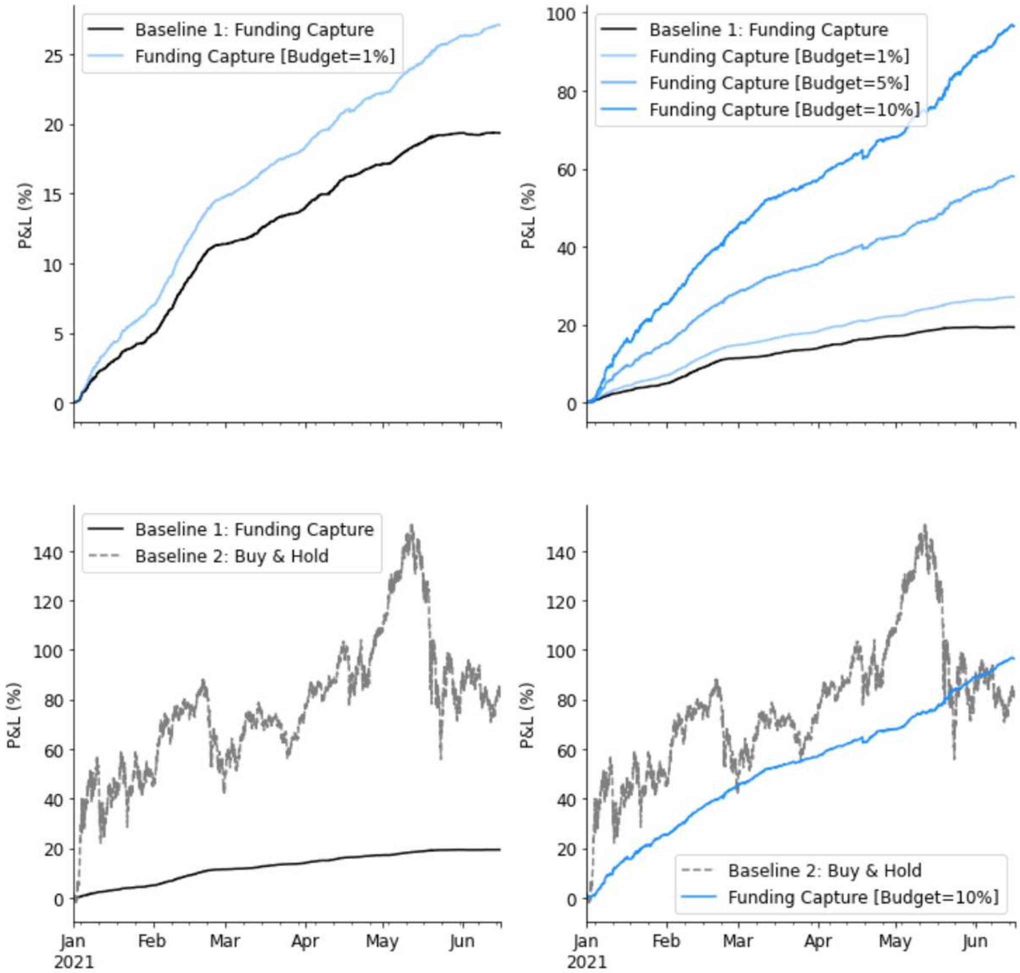


Figure 5.15: **ETHUSD** performance comparison for the different strategy variants, in each case using a Bayesian weighted conviction ensemble to make trading decisions. Upper charts use Funding Capture as the baseline strategy while lower charts use Buy & Hold.

As can be seen from Figures 5.14 and 5.15, the two baseline strategy profiles are very different. The Buy & Hold strategy exhibits large swings in its performance but can provide high returns, whereas the Funding Capture strategy typically generates very consistent returns with lower upside. However, as BTCUSD had a large selloff in May (due to Elon Musk's tweets, and rumours that Tesla had sold their previously acquired Bitcoins) the Funding Capture strategy outperformed the Buy & Hold strategy during this period while maintaining its more desirable risk profile. This is not the case for ETHUSD, for which the Buy & Hold strategy outperformed Funding Capture. There is thus a trade-off between the amount of risk taken and the consistency of return delivered by a strategy.

The conviction ensembles effectively combine the best features of the baseline strategies, allowing directional risk to be safely taken while still allowing for consistent returns to be generated from the perpetual futures funding rate. The proposed Funding Capture strategy with different directional budgets provides a strong alternative to simply buying the underlying or fully hedging directional exposure, as the classical Funding Capture strategy does. As already noted, the proposed strategy can provide returns similar to Buy & Hold, with Sharpe ratios comparable to the classical Funding Capture strategy, essentially providing the best of both worlds. This successful fusion is achieved through the power of a mid-price conviction ensemble allowing for controlled directional exposure to be opened within the bounds of a pre-specified risk tolerance, giving a flexibility that the baseline strategies do not provide. Table 5.8 below summarises each strategy's performance statistics.

Table 5.8: An out-of-sample comparison of trading performance across the different strategies under consideration.

Asset	**Winning Days**	**Mean Daily Return**	**Longest Drawdown**	**Max Drawdown**	**Sharpe Ratio**	**P&L**
Baseline 1: Funding Capture						
BTCUSD	89.09%	0.10%	10 Days	-0.18%	19.31	18.43%
ETHUSD	89.70%	0.11%	9 Days	-0.12%	20.79	19.25%
Baseline 2: Buy & Hold						
BTCUSD	47.27%	0.04%	62 Days	-45.79%	0.59	6.50%
ETHUSD	53.94%	0.37%	41 Days	-31.41%	1.95	84.57%
Funding Capture [1% Budget]						
BTCUSD	92.73%	0.13%	9 Days	-0.27%	22.56	23.85%
ETHUSD	94.55%	0.14%	3 Days	-0.09%	24.46	26.92%
Funding Capture [5% Budget]						
BTCUSD	92.73%	0.23%	8 Days	-0.88%	21.19	45.56%
ETHUSD	92.73%	0.27%	2 Days	-0.55%	20.33	57.53%
Funding Capture [10% Budget]						
BTCUSD	90.91%	0.33%	8 Days	-1.45%	18.50	72.67%
ETHUSD	91.52%	0.41%	3 Days	-0.97%	17.34	95.67%

It is clear from the descriptive statistics in Table 5.8 that at a baseline level the Funding Capture strategy offers a much more robust return stream compared to Buy & Hold. The basic Funding Capture strategy offers Sharpe ratios up to 20.79 on ETHUSD compared to a maximum Sharpe ratio of 1.95 on ETHUSD using the Buy & Hold strategy. The advantages of the Funding Capture strategy are further emphasised through other descriptive statistics such as the number of winning days (an average of 89.39% across both assets for Funding Capture compared to 50.61% for Buy & Hold), the maximum drawdown (an average of -0.15% across both assets for Funding Capture compared to -38.60% for Buy & Hold), and the longest drawdown (an average of 9.5 days for Funding Capture compared to 51.5 days for Buy & Hold). However, the ETHUSD Buy & Hold strategy outperformed in p&l and mean daily return, as would be expected in times of bull markets, which were experienced during the period that these tests were run. The BTCUSD Buy & Hold strategy did not outperform in p&l or mean daily return making it the worst performing strategy. As previously mentioned, this is due to strong negative sentiment surrounding Tesla's Bitcoin holdings as mentioned previously that resulted in the price of BTCUSD crashing in May.

The benefits of using conviction ensembles over the classical Funding Capture strategy are now analysed. Across the three directional budgets explored, the p&l of the baseline Funding Capture strategy is increased on average by 35%, 173%, and 346% for directional budgets of 1%, 5% and 10% respectively. These large outperformances come with very little damage to the strategy's core properties, with only a marginal increase in the max drawdown, and in fact a shortened drawdown duration in cases, most apparently on ETHUSD. Hence, although a deeper max drawdown is observed, this drawdown is generally recovered from more quickly than in the case of the basic Funding Capture baseline strategy. Across all the directional exposures considered the number of winning days increased by 4% on average, while mean daily returns also increased by 29%, 138%, and 251% for 1%, 5% and 10% directional budgets, respectively. These enhancements are directly attributed to directional exposures provided through the conviction ensembles. Remarkably, for taking only a 1% directional exposure the baseline Funding Capture strategy can be improved by 35% in p&l. It can also be seen that by opening just a 10% directional exposure the strategies p&l can marginally outperform an ETHUSD Buy & Hold return of 84.57% by 13.12%, taking the p&l to 95.67% while maintaining exceptionally high Sharpe ratios.

Figures 5.16 and 5.17, on the next page, address the consistency of each strategy type comparing the baseline strategies with the minimum (1%) and maximum (10%) directional exposure budget explored in this work.

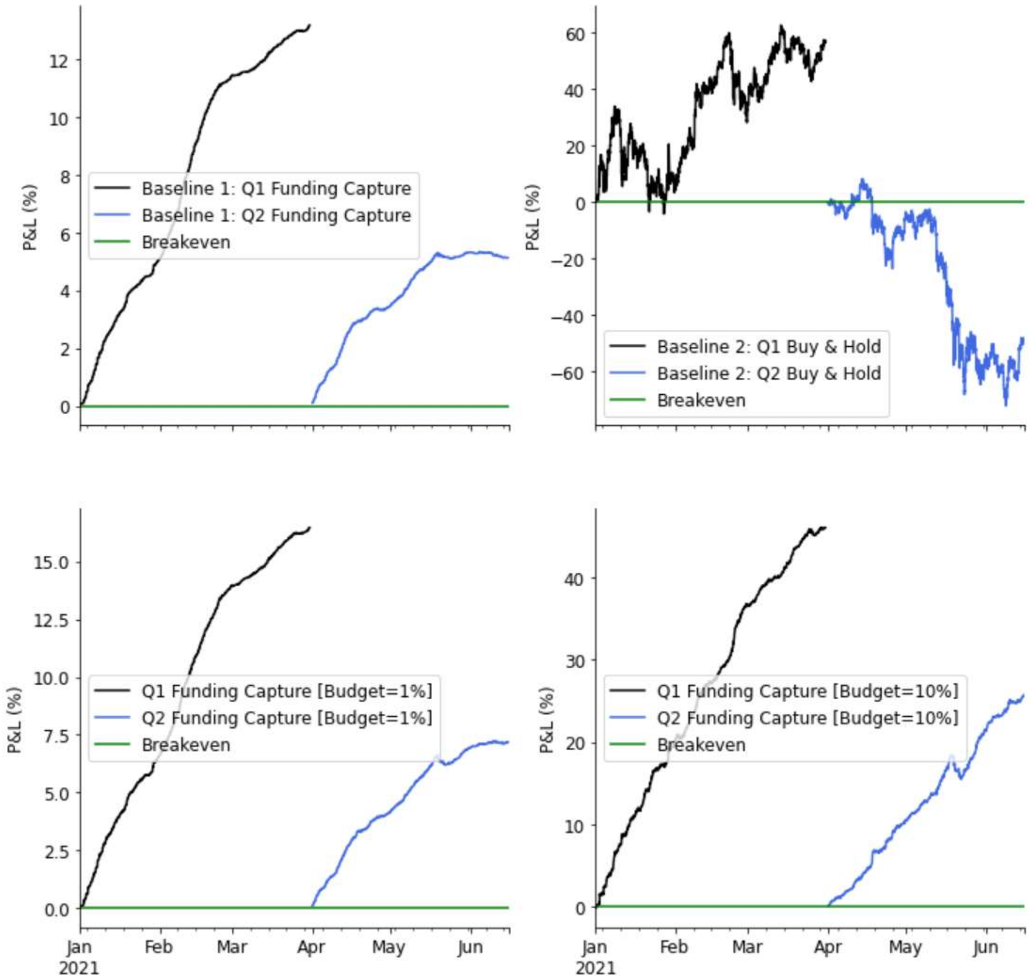


Figure 5.16: **BTCUSD** quarterly performance comparison over different strategy variants.

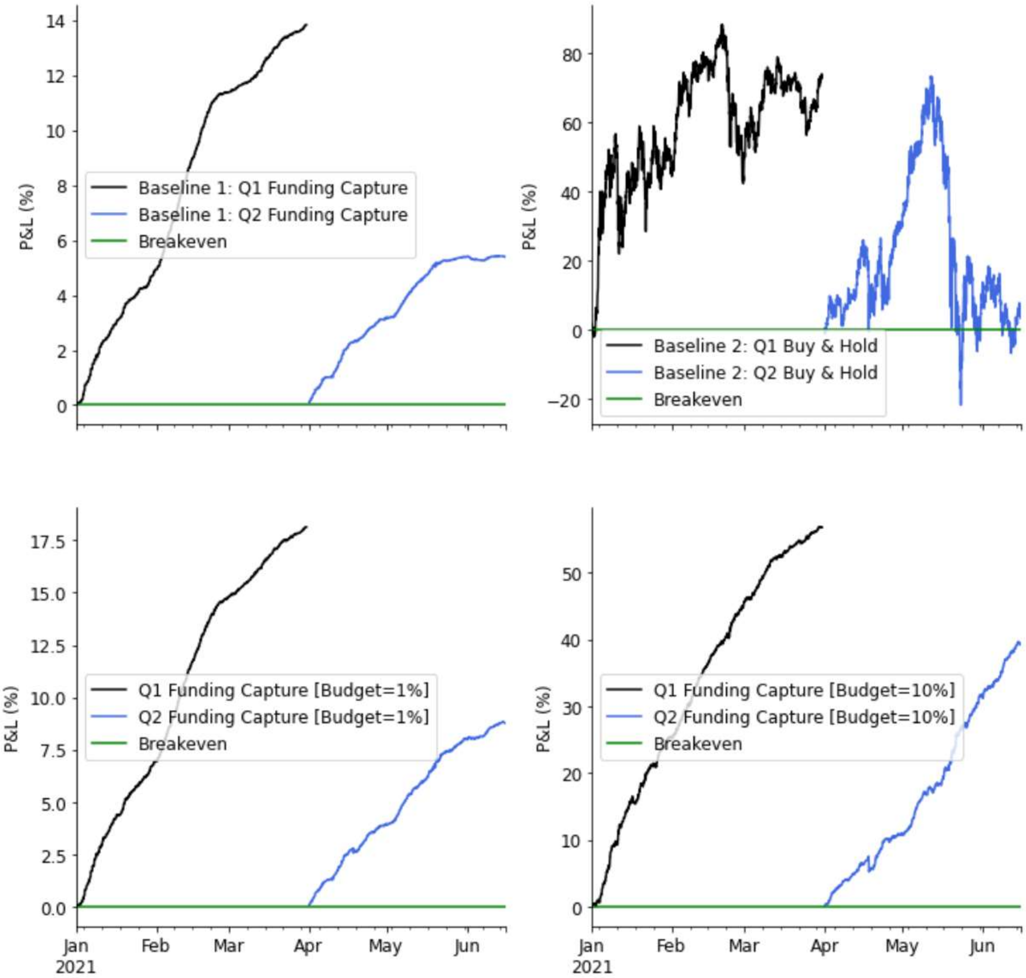


Figure 5.17: **ETHUSD** quarterly performance comparison over different strategy variants.

Figure 5.16 shows that BTCUSD performed very well in the first quarter of 2021, giving a baseline Buy & Hold strategy a p&l of ~60%, an outperformance over all other strategies considered over the same period (although if they are compared with respect to Sharpe ratio the proposed Funding Capture strategy series again outperforms). However, as mentioned earlier, the BTCUSD return for Q2 severely underperformed, losing ~50% and reducing the year-to-date return of a BTCUSD Buy & Hold strategy to only 6.5%. The same pattern is observable in the ETHUSD returns in Figure 5.17, where the Buy & Hold strategy performs well in the first Quarter of 2021 and then remains flat to the end of Q2. This observation emphasises how important timing is when considering a Buy & Hold strategy. ETHUSD demonstrates this sensitivity most clearly, where ~40% of the return is generated in the first week of 2021. Therefore, an investor who deployed a Buy & Hold strategy from January 1st, 2021, would be ~40% better off than an investor who deployed the same strategy a week later. This is demonstrated visually in Figure 5.18 below.

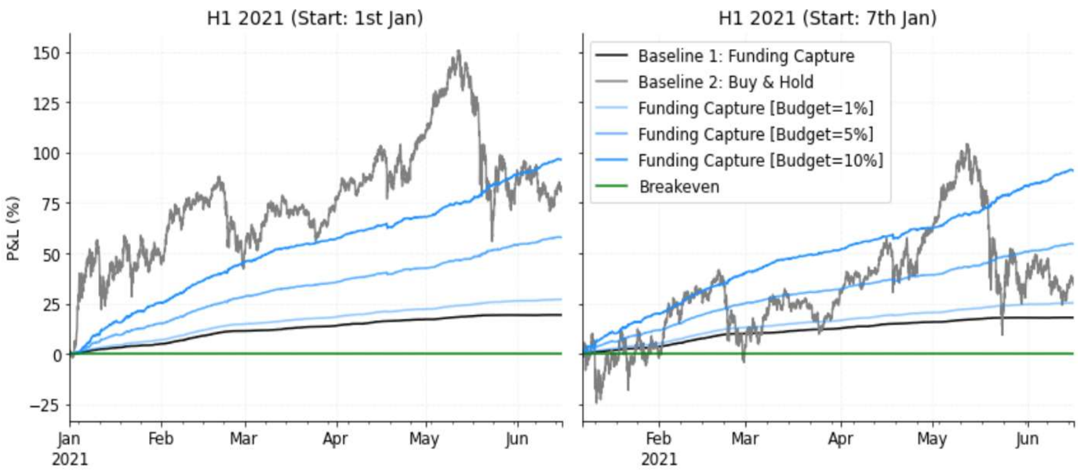


Figure 5.18: **ETHUSD** performance sensitivity demonstrates the weakness of a Buy & Hold strategy compared to a systematic strategy which can deliver consistent returns.

The timing sensitivity demonstrated in Figure 5.18 is a very undesirable property for a strategy as it highlights that an investor's returns are mainly a function of a lucky entry point, making this type of strategy unsustainable for many investors. In the ETHUSD example above it can be seen that an investor entering the market on the 7^{th} January 2021 would take on a very large amount of risk by holding such a volatile asset, and only marginally outperform an investor who deployed the Funding Capture strategy with a 1% directional exposure budget, a very undesirable result for the Buy & Hold investor.

Gaining exposure to extreme returns through a Buy & Hold strategy is therefore highly conditional on optimally timing the opportunity to enter the market, which is often attributed to luck. When comparing this to the Funding Capture strategy with directional budgets, the benefits of fusing the Funding Capture with the conviction ensemble's superior predictive power becomes apparent, with the 10% directional budget being able to outperform Buy & Hold while delivering high Sharpe ratios, low drawdowns and consistent returns which are robust to timing. The success of combining the Funding Capture strategy with directional budgets could only be achieved with an extremely high predictive power that is also consistent and robust, a characteristic not realistically achievable through traditional methodologies and target constructions.

5.5 Discussion

In this chapter it has been shown that the power of a *mid-price-2* prediction holds across multiple time horizons from 15 minutes to 4 hours, demonstrating that this novel target can provide robust, consistent, and scalable predictive value. The development of Bayesian optimised conviction ensembles extended the framework for *mid-price-2* predictions. Several different ensemble weighting schemes were considered, of which the Bayesian weighted scheme was the most effective, outperforming any individual model's predictive power by an average of 22.93%

across all assets under examination, with the ensemble achieving a maximum Cohen's kappa of 73.26% on out-of-sample ETHUSD data.

The work of Chapter 4 had demonstrated that a high predictive power could be achieved through a *mid-price-2* prediction, but had noted that making use of this predictive power would not be trivial since a *mid-price-2* is not located at a specific point in time. Here, a novel twist on two classical trading strategies was developed, powered by a Bayesian optimised conviction ensemble, which outperformed both baseline strategies considered, while maintaining each of their desirable properties, resulting in high Sharpe ratios and high returns with minimal directional risk taken. This outperformance is directly attributed to the power of a *mid-price-2* prediction and could not have been achieved with traditional targets, due to such targets being unable to provide a high and consistently robust predictive power. The construction of conviction ensembles and the resulting trading strategy has many components which can be further optimised, producing potentially better results. Therefore, the results here only indicate the potential power and trading value that conviction ensembles could provide.

Chapter 6

Conclusions and future work

In this final chapter, the main contributions of this thesis are summarised, future work is proposed, and concluding remarks are made.

6.1 Discussion and summary of contributions

This thesis has focused on the design and development of novel machine learning methodologies for extracting predictive power from OHLC data structures, a subject that has seen much controversy (Marshall et al., (2006; 2008); Horton, 2009; Fock et al., 2005; Duvinage et al., 2013; Detollenaere and Mazza, 2014). The aim of this work has been to provide novel methodologies for extracting value from these data structures, in several real-world contexts from pattern mining to prediction and algorithmic trading. Throughout this thesis empirical results have been presented which demonstrate that the novel techniques, and novel targets, proposed here have successfully been able to extract valuable information from OHLC data structures, a result in contrast to many other works in the literature (see Section 2.2.4).

In Chapter 3, *Deep Candlestick Mining* (DCM) was introduced, with the aim of discovering both new and traditional statistically significant predictive candlestick patterns. A novel double-clustering process was first developed to identify general and specific structures in data. The approach was applied to understand what OHLC feature input states a BOHB-optimised LSTM RNN found to be useful for predicting the close price change. Using the proposed methodology, a relationship between LSTM RNN predictions and arrangements of OHLC data structures, known as candlestick patterns, was discovered allowing for a deeper understanding of what specific OHLC arrangements provided predictive value to the LSTM RNN. The novel DCM-based candlestick patterns were represented by the centroids of each cluster, constructed by averaging each cluster's candidate OHLC arrangements, deemed valid if several validation filters were passed. A test was then formulated to compare the predictive power of traditional vs. DCM-based candlestick patterns, showing DCM-based patterns to significantly outperform. The result demonstrates that there is significant value in OHLC data structures if used in combination with novel methodologies not widely known to the community.

In Chapter 4, a new prediction target was proposed in the form of *mid-price-1*, the point between the high and low of an OHLC structure, and *mid-price-2*, the point between the open and close price levels. It was found that these targets had reduced noise content compared to traditional ones like the close price, making them easier to predict. The highest predictive power was achieved through the *mid-price-2* target, in combination with (importance mined + correlation filtered) OHLC features, suggesting again that there is value in OHLC data structures if used in a novel context, where here the novelty lies in both the features and the target *mid-price*. However, although the methodology provided a level of predictive power rarely observed when forecasting financial quantities, it was suggested it would be hard to use in practice as the *mid-price* does not reside at a specific point in time.

In Chapter 5, the concepts from the previous chapter were extended into an ensemble setting by training N BOHB-optimised LSTM RNN’s to predict *mid-price-2* directional changes at N different horizons, as previously presented in Chapter 4 for N=1, and then combined using a weighting scheme to construct the ensemble. It was shown that the results presented in Chapter 4 for a 60-minute horizon were robust to a very wide range of horizons from 15-minutes to 4-hours achieving remarkably high model performance scores for each horizon. This further emphasises the robust and consistent predictive power which can be achieved from a novel *mid-price-2* target using OHLC features if modelled through a context not widely considered before. Using these models, an ensemble was constructed by exploring three different weighing schemes (equally weighted, Cohen's kappa weighted, and Bayesian weighted) with the Bayesian weighting scheme shown to provide the best results over all considered horizons. A Bayesian-weighted ensemble achieved a predictive power exceeding 70% MCC at the 60-minute horizon on ETHUSD, an improvement of 30% over any result in the previous chapter. The remarkably high predictive power presented, while surprising, does not contradict the surrounding academic literature as the performance was achieved by using a novel target and ensemble construction process never used before in the context of OHLC data structures. To demonstrate the utility of *mid-price-2* Bayesian optimised conviction ensembles, the predictions were used to control directional exposure, execution, and position sizing in a novel twist of a popular institutional market neutral cryptoasset trading strategy, allowing the strategy to earn higher returns from controlled directional risk exposures. It was shown that by opening only small directional exposures, enhancements of up to 346% over the baseline strategy could be achieved in total return, while maintaining exceptionally high Sharpe ratios. The approach offers a way to achieve returns at the level of Buy & Hold without sacrificing consistency of return, a novel yet highly flexible strategy proposition. The success of this strategy addressed concerns in Chapter 4 that it might be hard to harness the power of *mid-price-2* predictions in a real-world scenario.

The primary result of this thesis, presented in Chapter 5, and based on work in the previous two chapters, shows that a remarkably high predictive power can indeed be extracted from OHLC data structures and used successfully in an algorithmic trading setting, if formulated in combination with novel targets and modelling techniques. Overall, the work presented here has shown substantial promise in the development of new methodologies for predictive systems using OHLC data structures, although there is still scope for improvement, as discussed below.

6.2 Future work

There are many opportunities for the work presented in this thesis to be further developed and explored, due to its novelty. As concepts are shared across chapters proposals are presented per component of the modelling pipeline used throughout this thesis, rather than related to a specific chapter.

All proposed enhancements are designed to develop the baseline framework presented in this thesis into a dynamically evolving pipeline for enhanced financial market prediction and trading.

6.2.1 Feature construction

Multi-exchange OHLC structures. Currently, a single exchange's data feed is used to construct the OHLC data structures which form the baseline feature inputs. While this approach is often acceptable for traditional assets, which are usually listed on only one exchange, it is commonplace for cryptoassets to be listed on multiple exchanges. Hence, the construction of OHLC data structures using the whole market (often 10+ exchanges for cryptoassets) should be considered to increase information content embedded in the baseline OHLC structure.

Price discovery weightings. As there are many exchanges around the world catering to different client types, trading styles, products (spot, futures, options, etc.), and market views, an OHLC structure constructed over a universe of exchanges, as suggested above, may start to show higher variance in its price levels, making prediction harder as more uncertainty is embedded in the feature space and targets. To adjust for this, it may be valuable to explore a dynamic weighting scheme to account for informational flows between exchanges and adjust for exchange nuances (such as the Kimchi premium (Mourdoukoutas, 2018)) resulting in a price discovery aware OHLC data structure.

Sampling method. This work has focused on *clock time sampling*, where OHLC structures are generated every clock period (e.g., every hour) as it is the default sampling convention and used widely in the surrounding literature. However, it would be useful to investigate the use of *event time sampling*, such as 'dollar volume time sampling', which generates OHLC structures every N dollars (e.g., every \$10 million traded). The main advantages of event time sampling are enhanced statistical properties for modelling (data is more i.i.d., which is often advantageous for machine learning algorithms (Lopez de Prado, 2018)), and informational stability as each OHLC period represents a constant quantity of market events (clock time does not, which could result in subtle modelling instabilities). It is expected enhanced control can be achieved through event time sampling, as models can make decisions in line with the natural rate at which markets process information rather than at an arbitrarily imposed rate, as is the case with clock time.

6.2.2 Feature selection

In this work a very large universe of features is exposed to an XGBoost model which subsequently ranks them for importance against a specific target (e.g., close price, *mid-price-2*, …). The ranked features are then processed further through a correlation filtering framework. However, given the dimensionality of the feature universe and the inherent noise content, it may be appropriate to investigate the use of different methods such as LSTM RNN Deep AutoEncoder (Srivastava et al, 2015) architectures to first process the feature universe to obtain a representation with reduced noise and dimensionality, through the AutoEncoder's latent space, then assess importance mappings to a desired target. This approach may be better able to exploit sequential relationships and subtle conditional information in a large feature universe, resulting in a richer feature space.

6.2.3 Structural identification

k-means++ has been used to design a novel double-clustering process which identifies general, and then specific, structure in data. This approach, while functional, does not scale well to larger datasets. Hence, it would be valuable to explore more scalable clustering algorithms for the purpose of structural identification in large feature sets, such as PQk-means (Matsui et al, 2017).

6.2.4 Target construction

Mid-price-1 and *mid-price-2* were proposed in this work as novel targets, showing that remarkable predictive power can be achieved by predicting their directional change. This is likely for several reasons, including a reduced noise content and the fact that they are unknown quantities to most market participants, thereby reducing competition in the prediction and subsequent trading based on their values. Given the success of these proposed quantities in this work it would be valuable to investigate a series of novel *mid-price* constructions focusing on mid-points between known, and lesser known, OHLC price levels such as mid-points between high to upper body, upper body to mid-price-2, mid-price-2 to lower body, lower body to low, etc. By sub-dividing each level of an OHLC structure, novel targets can be constructed, allowing a much larger proportion of the OHLC structure to be predicted via novel quantities, rather than well-known traditional ones, likely resulting in more accurate prediction of trends.

6.2.5 Ensemble construction

In Chapter 5 the best ensembles were constructed by applying a Bayesian weighting scheme to a universe of individually trained BOHB-optimised LSTM RNNs. While this was shown to be a highly effective methodology it could be extended to (1) consider regimes, (2) dynamically weight sub-models, and (3) dynamically train sub-models through asynchronous schedules. These propositions are discussed in more detail below.

Regime aware ensembles. It would be worthwhile investigating the combination of the DCM's double-clustering methodology applied to the task of market regime identification with ensemble weighting schemes. The combination of these two novel techniques would result in regime aware ensembles, potentially enhancing prediction accuracy as ensembles would be weighted per regime to maximise predictive power.

Dynamic ensemble weighting. An extension to the above-proposed regime aware ensembles could be to use a reinforcement learning agent to construct a dynamic weighing scheme. The agent could take the regime cluster IDs and sub-model predictions as input to derive a dynamic weighting for the ensemble. The reward function would be designed to maximise an objective of interest such as the MCC, or even a desired trading performance metric, such as the Sharpe ratio.

Asynchronously staggered ensembles. The synchronisation of a classical ensemble's sub-models is sensitive to the data each model is trained on. Typically, the dataset each sub-model is trained on is the same. It would be valuable to explore a methodology to de-synchronise the sub-models' training ranges, to break up any unwanted correlation in model predictions. Each model would be trained over different data ranges and only be available at certain times during the prediction period, as they would auto-re-train at different times, based on variables such as performance and consistency. The proposal is based upon the observation that the most recent training data is not necessary the most representative of current trading conditions and that a range of models exploiting different data properties across different data ranges could improve the ensemble's robustness and accuracy.

The above proposals build on the current work by extending the core ensemble design principals, making the framework more dynamic, allowing it to evolve with the market and adjust for different regimes. These proposals will very likely result in enhanced robustness of the predictive power, offering a greater level of control when used in a trading strategy.

6.2.6 Trading strategy design

In this thesis, a popular institutional market neutral trading strategy was enhanced by the addition of directional risk budgets controlled by conviction ensembles. This provided the strategy with a directional income, in addition to the baseline funding rate capture, by exploiting the remarkably accurate and robust predictive power of a *mid-price-2* conviction ensemble. Future work should focus on two main areas – namely, execution and risk management.

Execution. Enhancing the limit order placement model (LOPM) is of critical importance to increase the dollar capacity of this strategy, and to protect against information leakage as a result of large orders being placed at specific price levels. Initial investigation should focus on layering limit orders into the order book as a function of distance from previously realised *mid-price-2* levels and conviction. By encouraging more limit order placements, some more aggressively priced than the previously realised *mid-price-2*, a more continuous execution model could be achieved, enhancing capacity and adding a third source of income to the strategy through fee rebates, as some exchanges pay for limit orders to be placed into their order books to enhance liquidity – typically the job of a market maker. In addition to enhanced pricing logic, a multi-exchange approach should be integrated into the LOPM to exploit pricing inefficiencies as a result of fragmented liquidity observed in cryptoassets.

Risk Management. An investigation into dynamic risk budgets as a function of conviction and regime should be considered. This would allow an enhanced level of control, conditional on regime, allowing for larger directional income budgets to be allocated at times when market conditions are stable, and for directional risk budgets to be reduced or turned off at times when market direction is uncertain.

Overall, the aim of this trading strategy is to demonstrate a novel use case of a conviction ensemble in the context of a tightly risk-controlled trading strategy. The strategy's additional income streams (directional exposure and fee rebates) are only attractive if a remarkably high level of predictive power can be achieved, not possible with traditional predictive modelling methodologies or targets.

6.3 Concluding remarks

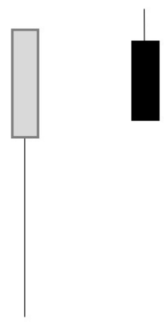
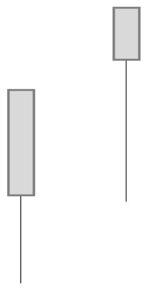
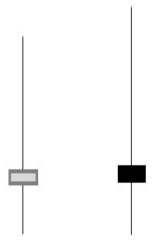
The prediction of financial markets is, and will remain, a difficult challenge due to market dynamics being influenced in part by random and subtle forces. However, the work presented here suggests that there are ways to exploit information embedded in OHLC data structures to accurately predict and trade financial markets if novel methodologies are used.

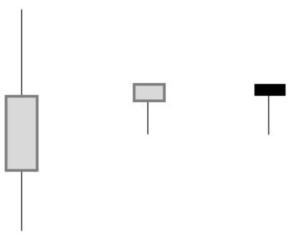
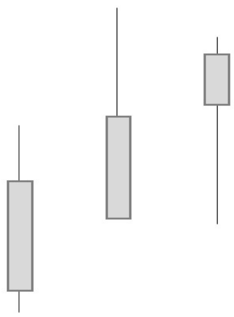
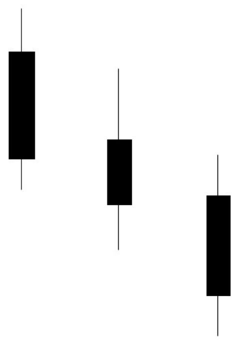
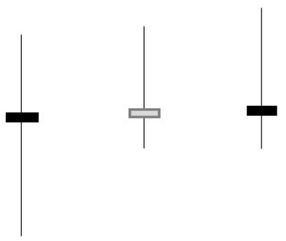
Appendix A

Additional Research Results

Significant candlestick pattern catalogues are presented on the following pages supplementing the catalogues presented for FGBL and ETHUSD in Chapter 3.

Table A.1: BTCUSD Deep mined candlestick patters for all lags.

2 Candle Patterns		
		
Cluster: 0,1	Cluster: 0,4	Cluster: 2,1

3 Candle Patterns		
		
Cluster: 0,3	Cluster: 1,1	Cluster: 1,3
		
Cluster: 2,0		

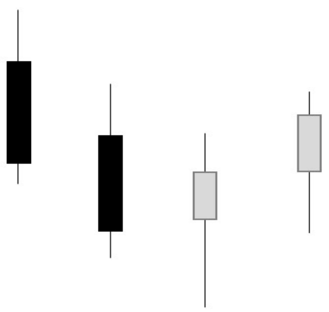
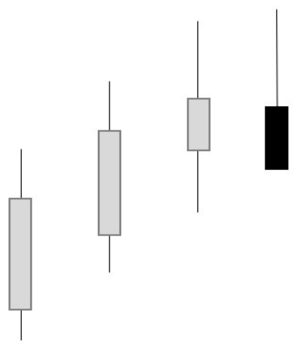
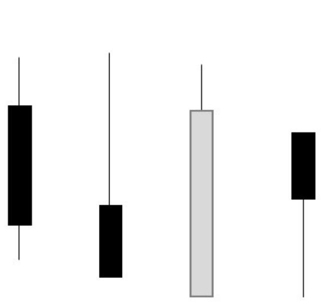
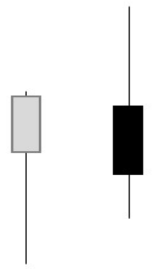
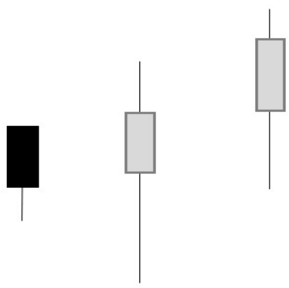
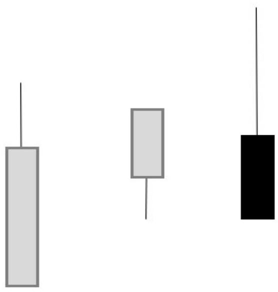
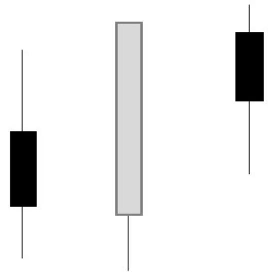
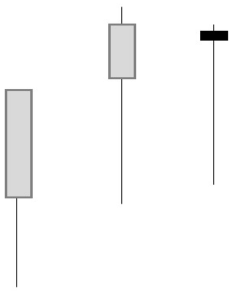
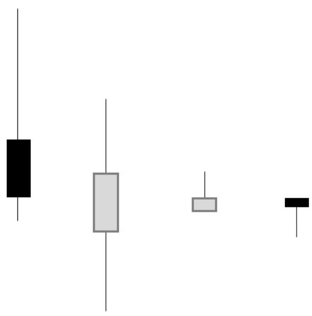
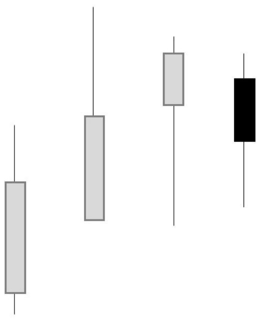
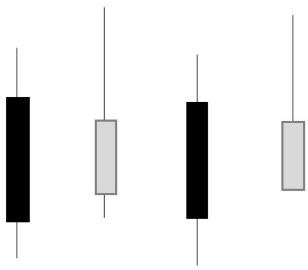
4 Candle Patterns		
		
Cluster: 0,4	Cluster: 3,3	Cluster: 4,1

Table A.2: EURUSD Deep mined candlestick patters for all lags.

2 Candle Patterns		
		
Cluster: 0,5		

3 Candle Patterns		
		
Cluster: 0,1	Cluster: 0,4	Cluster: 1,2
		
Cluster: 3,3		

4 Candle Patterns		
		
Cluster: 0,4	Cluster: 1,2	Cluster: 1,4

Bibliography

Alexander, C., Deng, J. & Zou, B. (2021) Hedging with Bitcoin Futures: The Effect of Liquidation Loss Aversion and Aggressive Trading. Arxiv. Available from: https://arxiv.org/abs/2101.01261. [Accessed: 11th August 2021].

Al-Hnaity, B. & Abbod, M. (2015) A novel hybrid ensemble model to predict FTSE100 index by combining neural network and EEMD. *In 2015 European Control Conference*. Linz, Austria. 15-17 July 2015. IEEE. 3021-3028.

Arthur, D. & Vassilvitskii, S. (2007) k-means++: the advantages of careful seeding. *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*. Philadelphia, PA, USA. 1027–1035.

Attanasio, G., Cagliero, L., Garza, P. & Baralis, E. (2019) Quantitative cryptocurrency trading: exploring the use of machine learning techniques. *Proceedings of the 5th Workshop on Data Science for Macro-modeling with Financial and Economic Datasets*. Amsterdam, Netherlands. 30 June 2019. ACM. 1:1-6.

Bank of England (2018) What are cryptoassets (cryptocurrencies)? Available at https://www.bankofengland.co.uk/knowledgebank/what-are-cryptocurrencies. [Accessed: 4th September 2019]

Bao, W., Yue, J. & Rao, Y. (2017) A deep learning framework for financial time series using stacked autoencoders and long-short term memory. *PLoS ONE*. 12 (7). Available from https://doi.org/10.1371/journal.pone.0180944 [Accessed 12th August 2019].

Bergstra, J., Bardenet, R., Bengio, Y., & Kégl, B. (2011) Algorithms for hyper-parameter optimization. *In Proceedings of the 25th International Conference on Advances in Neural Information Processing Systems.* 2546–2554.

Bigalow, S.W. (2011) *Profitable candlestick trading: pinpointing market opportunities to maximize profits.* New York, John Wiley and Sons Inc.

Brochu, E., Cora, V.M. & De Freitas, Nando (2010) A tutorial on bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. *arXiv* preprint arXiv:1012.2599.

Brock, W., Lakonishok, J. & LeBaron, B. (1992) Simple technical trading rules and the stochastic properties of stock returns. *Journal of Finance.* 47 (5), 1731–1764.

Brown, D. & Jennings, R., (1989) On technical analysis. *Review of Financial Studies.* 2 (4), 527–551.

Byrd, R.H., Chin, G.M., Nocedal, J. & Wu, J. (2012) Sample size selection in optimization methods for machine learning. *Mathematical programming.* 134 (1), 127–155.

Chen, G.H., Nikolov, S. & Shah, D. (2013) A latent source model for nonparametric time series classification. *Advances in Neural Information Processing Systems.* 1088-1096.

Chen, K., Zhou, Y. & Dai, F. (2015) A LSTM-based method for stock returns prediction: A case study of China stock market. *Proceedings of the 2015 IEEE International Conference on Big Data.* Santa Clara, CA, USA, 29 Oct. – 1 Nov 2015. IEEE. 2823-2824.

Chen, S., Bao, S. & Zhou Y. (2016) The predictive power of Japanese candlestick charting in Chinese stock market. *Physica A: Statistical Mechanics and its Applications.* 475, 148-165.

Chen, T. & Guestrin, C. (2016) XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco, California, USA, 13-17 August 2016. ACM. 785–794.

Cheng, C., Xu, W. & Wang, J. (2012) A Comparison of Ensemble Methods in Financial Market Prediction. *In Fifth International Joint Conference on Computational Sciences and Optimization*. Harbin, China. 23-26 June 2012. IEEE. 755-759.

Chowdhury, R., Rahman, M.A., Rahman, M.S., Mahdy, M.R.C. (2019) Predicting and Forecasting the Price of Constituents and Index of Cryptocurrency Using Machine Learning. *Arxiv*. Available from: https://arxiv.org/abs/1905.08444. [Accessed: 16th August 2019].

Clemen, R.T. (1989) Combining forecasts: A review and annotated bibliography. *International journal of forecasting*. 5 (2), 559-583.

de Mello Assis, J., Pereira, A.C.M. & Couto e Silva, R. (2018) Designing Financial Strategies based on Artificial Neural Networks Ensembles for Stock Markets. *In 2018 International Joint Conference on Neural Networks*. Rio de Janeiro, Brazil. 8-13 July 2018. IEEE. 1-8.

Detollenaere, B. & Mazza, P. (2014) Do Japanese candlesticks help solve the trader's dilemma? *Journal of Banking & Finance*. 48, 386-395.

Ding, X., Zhang, Y., Liu, T. & Duan, J. (2015) Deep learning for event-driven stock prediction. *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*. Palo Alto, California, USA. 25–31 July 2015. ACM. 2327-2333.

Dittmar, R., Neely, C. J. & Weller, P. (1996) Is technical analysis in the foreign exchange market profitable? a genetic programming approach. *C.E.P.R. Discussion Papers, CEPR Discussion Papers*. 1480.

Ester, M., Kriegel, H.-P., Sander, J. & Xu, Xiaowei. (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*. Portland, Oregon, USA. 2^{nd} – 4^{th} 1996. AAAI Press. 226–231.

Falkner, S., Klein, Aaron. & Hutter, Frank. (2018) Bohb: Robust and efficient hyperparameter optimization at scale. *In Proceedings of the International Conference on Machine Learning*.

Fernandez-Rodrıguez, F., Gonzalez-Martel, C. & Sosvilla-Rivero, S. (2000) On the profitability of technical trading rules based on artificial neural networks: Evidence from the Madrid stock market. *Economics letters*. 69 (1), 89–94.

Fock, J., Klein, C. & Zwergel, B. (2005) Performance of candlestick analysis on intraday futures data. *The Journal of Derivatives*. 13 (1), 28-40.

Friedlander, M.P. & Schmidt, M. (2012) Hybrid deterministic-stochastic methods for data fitting. *SIAM Journal on Scientific Computing*. 34 (3), 1380–1405.

Fu, X., Zhang, S., Chen, J., Ouyang, T. & Wu, J. (2019) A Sentiment-Aware Trading Volume Prediction Model for P2P Market Using LSTM. *IEEE Access*. 7, 81934-81944.

Gao, T., Chai, Y. & Liu, Y. (2017) Applying long short term memory neural networks for predicting stock closing price. *In 8th IEEE International Conference on Software Engineering and Service Science*. Beijing, China. 24-26 November 2017. IEEE. 575-578.

Garliauskas, A. (1999) Neural network chaos and computational algorithms of forecast in finance. *Proceedings of International Conference on Systems, Man, and Cybernetics.* Tokyo, Japan. 12-15 October 1999. IEEE. 638-643.

Georgoula, I., Pournarakis, D., Bilanakos, C., Sotiropoulos, D.N., Giaglis, G.M. (2015) Using time-series and sentiment analysis to detect the determinants of Bitcoin prices. *SSRN.* Available from http://dx.doi.org/10.2139/ssrn.2607167. [Accessed: 16th August 2019].

Gers F.A., Eck D., Schmidhuber J. (2002) Applying LSTM to Time Series Predictable Through Time-Window Approaches. *In Tagliaferri R., Marinaro M. (eds) Neural Nets WIRN Vietri-01. Perspectives in Neural Computing.* London: Springer. 193-200.

Geurts, P., Ernst, D. & Wehenkel, L. (2006) Extremely randomized trees. *Machine Learning.* 63, 3-42.

Greaves, A. & Au, B. (2015) Using the Bitcoin transaction graph to predict the price of Bitcoin. *Stanford Univ.*

Gu, Q., Zhu, L. & Cai, Z. (2009) Evaluation measures of the classification performance of imbalanced data sets. *Communications in Computer and Information Science.* 51, 461-471.

Guo, Z., Wang, H., Liu, Q. & Yang, J. (2014) A Feature Fusion Based Forecasting Model for Financial Time Series. *PloS ONE.* 9 (6). Available from https://doi.org/10.1371/journal.pone.0101113 [Accessed 16th August 2019].

Gyamerah, S.A., Ngare, P. & Lkpe, D. (2019) On Stock Market Movement Prediction Via Stacking Ensemble Learning Method. *In 2019 IEEE Conference on Computational Intelligence for Financial Engineering & Economics.* Shenzhen, China. 4-5 May 2019. IEEE. 1-8.

Hayes, Arthur. (2016) Announcing the Launch of the Perpetual XBTUSD Leveraged Swap. BitMEX. Available from. https://blog.bitmex.com/announcing-the-launch-of-the-perpetual-xbtusd-leveraged-swap [Accessed 10th April 2021]

Heaton, J. (2005) *Introduction to Neural Network for Java.* St. Louis, Heaton Research. Inc.

Heeswijk, M., Miche, Y., Knuutila, T.L., Hilbers, P.A.J., Honkela, T., Oja, E. & Lendasse, A. (2009) Adaptive Ensemble Models of Extreme Learning Machines for Time Series Prediction. *Proceedings of the 19th International Conference on Artificial Neural Networks*. Limassol, Cyprus. 14-17 September 2009. Heidelberg: Springer. 305-314.

Hegde, M.S., Krishna, G. & Srinath, R. (2018) An Ensemble Stock Predictor and Recommender System. *In 2018 International Conference on Advances in Computing, Communications and Informatics*. Bangalore, India. 19-22 September 2018. IEEE. 1981-1985.

Hinton, G., Bengio, Y and LeCun, Y (2015) Deep Learning. *Advanced Neural Information Processing Systems Tutorial.* Montréal, Montréal, Canada. 7-12 December, 2015.

Hochreiter, S. & Schmidhuber, J. (1997) Long short-term memory. *Neural Computation*. 9 (8), 1735–1780.

Horton, M. (2009) Stars, crows, and doji: The use of candlesticks in stock selection. *The Quarterly Review of Economics and Finance*. 49 (2), 283-294.

Hsieh, T.-J., Hsiao, H.-F. & Yeh, W.-C. (2011) Forecasting stock markets using wavelet transforms and recurrent neural networks: An integrated system based on artificial bee colony algorithm. *Applied Soft Computing*. 11 (2), 2510–2525.

Hu, G., Hu, Y., Yang, K., Yu, Z., Sung, F., Zhang, Z., Xie, F., Liu, J., Robertson, N., Hospedales, T., et al. (2017) Deep stock representation learning: From candlestick charts to investment decisions. Available from: https://arxiv.org/abs/1709.03803 [Accessed 12th August 2019].

Huang, W., Nakamori, Y. & Wang, S.Y. (2005) Forecasting stock market movement direction with support vector machine. *Computers & Operations Research*. 32 (10), 2513–2522.

Ioffe, S., Szegedy, C. (2015) Batch normalization: accelerating deep network training by reducing internal covariate shift. *Proceedings of the 32nd International Conference on International Conference on Machine Learning*. Lille, France, 6-11 July 2015. ACM. 448-456.

Jasemi, M., Kimiagari, A.M., & Memariani, A. (2011) A modern neural network model to do stock market timing on the basis of the ancient investment technique of Japanese Candlestick. *Expert Systems with Applications*. 38, 3884-3890.

Jiang, X., Pan, S., Jiang, J. & Long, G. (2018) Cross-Domain Deep Learning Approach For Multiple Financial Market Prediction. *In International Joint Conference on Neural Networks*. Rio de Janerio, Brazil. 8-13 July 2018. IEEE. 1-8.

Kavajecz, K. A. (2004) Technical analysis and liquidity provision. *Review of Financial Studies*. 17 (4), 1043–1071.

Keskar, N., Mudigere, D., Nocedal, J., Smelyanskiy, M & Tang, P.T.P. (2017) On Large-Batch Training For Deep Learning: Generalisation Gap And Sharp Minima. *International Conference on Learning Representations*. Toulon, France. 24-26 April 2017.

Khare, K., Darekar, O., Gupta, P & Attar, V.Z. (2017) Short term stock price prediction using deep learning. *In 2nd IEEE International Conference On Recent Trends in Electronics Information & Communication Technology.* Bangalore, India. 19-20 May 2017. IEEE. 482-486.

Kim, K.-J. & Lee, W.B. (2004) Stock market prediction using artificial neural networks with optimal feature transformation. *Neural Computing & Applications.* 13 (3), 255-260.

Kim, T. & Kim H.Y. (2019) Forecasting stock prices with a feature fusion LSTM-CNN model using different representations of the same data. *PLoS ONE.* 14 (2). Available from https://doi.org/10.1371/journal.pone.0212320 [Accessed 12th August 2019].

Kingma, D. P. & Ba, J. L. (2015). Adam: A Method for Stochastic Optimization. *International Conference on Learning Representations.* 1–13.

Kristoufek, L. (2015) What are the main drivers of the Bitcoin price? evidence from wavelet coherence analysis. *PloS ONE.* 10 (4). Available from https://doi.org/10.1371/journal.pone. 0123923 [Accessed 12th August 2019].

Kuncheva, L.I. (2004) Classifier Ensembles for Changing Environments. *In 5th International Workshop on Multiple Classifier Systems.* Cagliari, Italy. 9-11 June 2004. Berlin: Springer. 1-15.

Kusuma, R.M.I., Ho, T.T., Kao, W.C., Ou, Y.Y. & Hua, K.L. (2019) Using Deep Learning Neural Networks and Candlestick Chart Representation to Predict Stock Market. Available from: https://arxiv.org/abs/1903.12258. [Accessed 12th August 2019].

Lan, Q., Zhang, D. & Xiong, L. (2011) Reversal Pattern Discovery in Financial Time Series Based on Fuzzy Candlestick Lines. *Systems Engineering Procedia.* 2, 182-190.

LeBaron, B. (1996) Technical trading rule profitability and foreign exchange intervention. *National Bureau of Economic Research, Inc, NBER Working Papers.* 5505.

Li, L., Jamieson, K., Rostamizadeh, A., Gonina, K., Hardt, M., Recht, B., and Talwalkar, A. Massively parallel hyperparameter tuning, 2018. Available from https://openreview.net/forum? id=S1Y7OOlRZ [Accessed on 11 August 2019]

Li, X., Chen, S., Hu, X. and Yang, J. (2018) Understanding the Disharmony between Dropout and Batch Normalization by Variance Shift. arXiv preprint *arXiv.* Available from https://arxiv.org/abs/1801.05134 [Accessed 17th July 2019].

Li, X. & Tam, V. (2017) A comparative study of a recurrent neural network and support vector machine for predicting price movements of stocks of different volatilites. *Proceedings of the 2017 IEEE Symposium Series on Computational Intelligence.* Honolulu, HI, USA. 27 Nov - 1 Dec 2017. IEEE. 1-8.

Lior, R. (2010) Ensemble-based classifiers. *Artificial Intelligence Review.* 33 (2), 1-39.

Little, W. A. (1974). The existence of persistent states in the brain. *Mathematical Biosciences.* 19 (2), 145–164.

Lo, A. W., Mamaysky, H. and Wang, J. (2000) Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. *National Bureau of Economic Research, Inc, NBER Working Papers.* 7613.

Lopez de Prado, M. (2018) Advances in financial machine learning. John Wiley & Sons.

Lopez de Prado, M and Lewis, M.J. (2019) Detection of false investment strategies using unsupervised learning methods. *Quantitative Finance*. 19 (9), 1555-1565.

Lu, T., Shiu, Y-M and Liu, T-C. (2012) Profitable candlestick trading strategies – The evidence from a new perspective. *Review of Financial Economics*. 21 (2), 63-68.

Lu, T. (2014) The profitability of candlestick charting in the Taiwan stock market. *Pacific-Basin Financial Journal*. 26, 65-78.

Lu, T.H., Chen, Y.C. & Hsu Y.C. (2015) Trend definition or holding strategy: What determines the profitability of candlestick charting? *Journal of Banking & Finance*. 61, 172-183.

Lui, Y.-H. and Mole, D. (1998) The use of fundamental and technical analyses by foreign exchange dealers: Hong Kong evidence. *Journal of International Money and Finance*. 17 (3), 535–545.

Masters, D. and Luschi, C. (2018) Revisiting small batch training for deep neural networks. *CoRR*, abs/1804.07612.

Madan, I., Saluja, S. & Zhao, A. (2015) Automated Bitcoin trading via machine learning algorithms. *Tech. Notes in Stanford Univ.*

Maknickas, A. & Maknickiene, N. (2015) Investment support system using the EVOLINO recurrent neural network ensemble. *In 2015 7th International Joint Conference on Computational Intelligence*. Lisbon, Portugal. 12-14 November 2015. IEEE. 138-145.

Mallqui, D.C.A. & Fernandes, R.A.S. (2019) Predicting the direction, maximum, minimum and closing prices of daily Bitcoin exchange rate using machine learning techniques. *Applied Soft Computing*. 75, 596-606.

Mann, A.D. & Gorse, D. (2017) A new methodology to exploit predictive power in (open, high, low, close) data. *Proceedings of the 26th International Conference on Artificial Neural Networks*. Alghero, Italy, 11-14 September 2017. Heidelberg: Springer. 495-502.

Mann, A.D. & Gorse, D. (2017) Deep Candlestick Mining. *Proceedings of the 24th International Conference on Neural Information Processing*. Guangzhou, China, 14-18 November 2017. Heidelberg: Springer. 913-921.

Marshall, B., Young, M. & Rose, L. (2006) Candlestick technical trading strategies: Can they create value for investors? *Journal of Banking & Finance*. 30 (1), 2303-2323.

Marshall, B., Young, M. & Cahan, R. (2008) Are candlestick technical trading strategies profitable in the Japanese equity market? *Review of Quantitative Finance and Accounting*. 31 (2), 191–207.

Matsui, Y., Ogaki, K., Yamasaki, T., & Aizawa, K. (2017) PQk-means: Billion-scale Clustering for Product-quantized Codes. ACM International Conference on Multimedia.

Mazza, P. (2013) Price Dynamics and Market Liquidity: An Intraday Event Study on Euronext. *The Quarterly Review of Economics and Finance*. 56 (1), 139-153.

McNally, S., Roche, J. & Caton, S. (2018) Predicting the Price of Bitcoin Using Machine Learning. *In 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing*. Cambridge, UK. 21-23 March 2018. IEEE. 339-343.

Menkhoff, L., & Taylor, M. P. (2007). The obstinate passion of foreign exchange professionals: Technical analysis. *Journal of Economic Literature.* 45 (4), 936–972.

Morris, G. (1992) *Candlepower: Advanced Candlestick Pattern Recognition and Filtering Techniques for Trading Stocks and Futures.* Chicago, Irwin Professional Publishing.

Morris, G. (1995) *Candlestick Charting Explained: Timeless Techniques for Trading Stocks and Futures.* Second ed. New York, McGraw-Hill Trade.

Morris, K. (1972) Mathematical thought from ancient to modern times. *Oxford University Press.* 2.

Mourdoukoutas, P. (2018) How To Find The Next Ripple, Ethereum, Monero and EOS - Try The Kimchi Premium. Available from: https://www.forbes.com/sites/panosmourdoukoutas/2018/04/29/how-to-find-the-next-ripple-ethereummonero-and-eos-try-the-kimchi-premium/. [Accessed: 11th August 2021].

Nakamoto, S. (2008) Bitcoin: A peer-to-peer electronic cash system. *Consulted.*

Naranjo, R., Arroyo, J. & Santos, M. (2018) Fuzzy modeling of stock trading with fuzzy candlesticks. *Expert Systems with Applications.* 93, 15-27.

Neely, C. J., Weller, P. A. and Ulrich, J. M. (2009) The adaptive markets hypothesis: Evidence from the foreign exchange market. *Journal of Financial and Quantitative Analysis.* 44 (2), 467–488.

Neftci, S. N. (1991) Naive trading rules in financial markets and wiener-kolmogorov prediction theory: A study of technical analysis. *The Journal of Business.* 64 (4), 549.

Nelson, D.M.Q., Pereira, A.C.M. & de Oliveira, R.A. (2017) Stock market's price movement prediction with LSTM neural networks. *In Proceedings of the 2017 International Joint Conference on Neural Networks*. Anchorage, AK, USA. 14-19 May. IEEE. 1419-1426.

Olson, D. (2004) Have trading rule profits in the currency markets declined over time? *Journal of banking & Finance*. 28 (1), 85–105.

Park, C.-H. and Irwin, S. H. (2004) The profitability of technical analysis: A review. *University of Illinois at Urbana-Champaign, Department of Agricultural and Consumer Economics.* AgMAS Project Research Reports 37487.

Phillips RC, Gorse D (2018) Cryptocurrency price drivers: Wavelet coherence analysis revisited. *PLoS ONE* 13 (4). Available from. https://doi.org/10.1371/journal.pone.0195200 [Accessed 31st August 2019]

Picasso Ratto, A., Merello, S., Oneto, L., Ma, Y., Malandri, L. & Cambria, E. (2018) Ensemble of Technical Analysis and Machine Learning for Market Trend Prediction. *In 2018 IEEE Symposium Series on Computational Intelligence*. Bangalore, India. 18-21 November 2018. IEEE. 2090-2096.

Prado, Hércules & Ferneda, Edilson & C.R. Morais, Luis & Luiz, Alfredo & Matsura, Eduardo (2013) On the Effectiveness of Candlestick Chart Analysis for the Brazilian Stock Market. *Procedia Computer Science*. 22, 1136-1145.

Prasaddas, S. & Padhy, S. (2012) Support Vector Machines for Prediction of Futures Prices in Indian Stock Market. *International Journal of Computer Applications*. 41 (3), 22–26.

Rasmussen, C.E. & Williams, K.C.I. (2006) Gaussian processes for machine learning. *MIT Press*. ISBN 0-262-18253-X.

Rather, A.M., Agarwal, A. & Sastry, V. (2015) Recurrent neural network and a hybrid model for prediction of stock returns. *Expert Systems with Applications.* 42 (6). 3234-3241.

Raudys, S. & Zliobaite (2006) The Multi-Agent System for Prediction of Financial Time Series. *In 8th International Conference on Artificial Intelligence and Soft Computing.* Zakopane, Poland. 25-29 June 2006. Berlin: Springer. 653-662.

Refenes, A.N., Zapranis, A. & Francis, G. (1994) Stock performance modeling using neural networks: a comparative study with regression models. Neural Networks. 7 (2), 375–388.

Riedmiller, M. & Braun, H. (1993) A direct adaptive method for faster backpropagation learning: The rprop algorithm. *Proceedings of 1993 IEEE International Conference on Neural Networks.* San Francisco, CA, USA. 28 March - 1 April 1993. IEEE. 586-591.

Rousseeuw, P.J (1987) Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics. 20 (1), 53-65.

Schulmeister, S., 2007. The interaction between the aggregate behavior of technical trading systems and stock price dynamics. *Working Paper*, Austrian Institute of Economic Research.

Schwaerzel, R. & Rosen, B. (1997) Improving the accuracy of financial time series prediction using ensemble networks and high order statistics. *In Proceedings of International Conference on Neural Networks.* Houston, TX, USA. 12 June 1997. IEEE. 2045-2050. [1,3 ref interesting]

Sermpinis, G., Stasinakis, C., Theofilatos, K. & Karathanasopoulos, A. (2015) Modeling, forecasting and trading the EUR exchange rates with hybrid rolling genetic algorithms—Support vector regression forecast combinations. *European Journal of Operational Research.* 247 (3), 831-846.

Sermpinis, G., Theofilatos, K., Karathanasopoulos, A., Georgopoulos, E.F. & Dunis, C. (2013) Forecasting foreign exchange rates with adaptive neural networks using radial-basis functions and particle swarm optimization. *European Journal of Operational Research.* 225 (3), 528-540.

Shah, D. & Kang, Z. (2014) Bayesian regression and Bitcoin. *In 52nd Annual Allerton Conference on Communication, Control, and Computing.* Monticello, IL, USA. 30 Sept.-3 Oct. 2014. IEEE. 409-414.

Shahriari, B., Swersky, K., Wang, Z., Adams, R., and de Freitas, N. (2016) Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE.* 104 (1),148–175.

Sharpe, W. (1966). Mutual Fund Performance. *The Journal of Business.* 39 (1), 119-138.

Shen, F., Chao, J. & Zhao, J. (2015) Forecasting exchange rate using deep belief networks and conjugate gradient method. *Neurocomputing.* 167 (C), 243–253.

Shiller, Robert J. (1993) Measuring Asset Values for Cash Settlement in Derivative Markets: Hedonic Repeated Measures Indices and Perpetual Futures. The Journal of Finance. 48 (3), 911–931.

Simon, G., Lendasse, A., Cottrell, M., Fort, J.-C. & Verleysen, M. (2005) Time series forecasting: Obtaining long term trends with self-organizing maps. *Pattern Recognition Letters.* 26 (12), 1795-1808.

Sin, E. & Wang, L. (2017) Bitcoin price prediction using ensembles of neural networks. *In 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery.* Guilin, China. 29-31 July 2017.IEEE. 666–671.

Smeeton, N.C. (1985) Early History of the Kappa Statistic. *Biometrics.* 41, 795.

Smith, C. & Jin, Y. (2014) Evolutionary multi-objective generation of recurrent neural network ensembles for time series prediction. *Neurocomputing.* 143 (1), 302-311.

Snoek, J., Larochelle, H. & Adams, R.P. (2012) Practical Bayesian Optimization of Machine Learning Algorithms. *Proceedings of the 25th International Conference on Neural Information Processing Systems.* Lake Tahoe, Nevada, USA, 3-6 December. ACM. 2951-2959.

Sorjamaa, A., Hao, J., Reyhani, N., Ji, Y. & Lendasse, A. (2007) Methodology for long-term prediction of time series. *Neurocomputing.* 70 (18), 2861-2869.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R. (2014) Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research.* 15 (1), 1929-1958.

Srivastava, N., Mansimov, E. & Salakhutdinov, R. (2015) Unsupervised learning of video representations using LSTMs. *Advances in Neural Information Processing Systems.*

Steinert, L. & Herff, C. (2018) Predicting altcoin returns using social media. *PLoS ONE* 13 (12). Available from https://doi.org/10.1371/journal.pone.0208119 [Accessed 16th August 2019].

Fischer, T., Krauss, C. & Deinert, A. (2019) Statistical Arbitrage in Cryptocurrency Markets. *Journal of Risk and Financial Management.* 12 (1), 31-46.

Taylor, M. P. & Allen, H. (1992) The use of technical analysis in the foreign exchange market. *Journal of International Money and Finance*. 11 (3), 304–314.

Troiano, L., Villa, E.M. & Loia, V. (2018) Replicating a Trading Strategy by Means of LSTM for Financial Industry Applications. *Transactions on Industrial Informatics*. 14 (7), 3226-3234.

Tsai, C.F. & Quan, Z.Y. (2014) Stock prediction by searching for similarities in candlestick charts. *ACM Transactions on Management Information Systems*. 5 (2), 1-9.

Tse, Thomas. (2020) The Nuances of OKEx Linear and Inverse Futures/Swap and How to Maximize Your Profit. OKEx. Available from. https://www.okex.com/academy/en/the-nuances-of-okex-linear-and-inverse-futures-swap-and-how-to-maximize-your-profit [Accessed 2nd March 2021]

van der Maaten, L.J.P. & Hinton, G.E. (2008) Visualizing High-Dimensional Data Using t-SNE. *Journal of Machine Learning Research*. 9, 2579-2605.

Wang, L. & Gupta, S. (2013) Neural Networks and Wavelet De-Noising for Stock Trading and Prediction. *Time Series Analysis, Modeling and Applications*. Heidelberg, Springer. 229–247.

Williams, R.J., Hinton, G.E., Rumelhart, D.E. (1986) Learning representations by back-propagating errors. *Nature*. 323 (6088), 533–536.

Xie, H., Zhao, X. & Wang, S. (2012) A Comprehensive Look at the Predictive Information in Japanese Candlestick. *Procedia Computer Science*. 9 (1), 1219-1227.

Xu, D., Zhang, Y., Cheng, C., Xu, W. & Zhang, L. (2014) A Neural Network-Based Ensemble Prediction Using PMRS and ECM. *In 2014 47th Hawaii International*

Conference on System Sciences. Waikoloa, HI, USA. 6-9 January 2014. IEEE. 1335-1343.

Yang, B., Gong, Z.J. & Yang W (2017) Stock market index prediction using deep neural network ensemble. *In 36th Chinese Control Conference*. Dalian, China. 26-28 July 2017. IEEE. 3882-3887.

Yang, J., Rao, R., Hong, P. & Ding, P (2016) Ensemble Model for Stock Price Movement Trend Prediction on Different Investing Periods. *In 2016 12th International Conference on Computational Intelligence and Security*. Wuxi, China. 16-19 December 2016. IEEE. 358-361.

Yu, H., Nartea, G.V., Gan, C. & Yao, L.J. (2013) Predictive ability and profitability of simple technical trading rules: Recent evidence from Southeast Asian stock markets. *International Review of Economics & Finance*. 25, 356-371.

Yu, L., Wang, S. & Lai, K.K. (2005) A novel nonlinear ensemble forecasting model incorporating GLAR and ANN for foreign exchange rates. *Computers & Operations Research*. 32 (10), 2523-2541.

Yu, L., Wang, S. & Lai, K.K. (2008) Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. *Energy Economics*. 30 (5), 2623-2635.

Yu, L., Wang, S. & Lai, K.K. (2015) A novel nonlinear ensemble forecasting model incorporating glar and ann for foreign exchange rates. *Computers and Operations Research*. 32 (10), 2523-2541.

Zeng, Y. & Liu X. (2018) A Stock Price Fluctuation Forecast Model Based on LSTM. *In 14th International Conference on Semantics, Knowledge and Grids*. Guangzhou, China. 12-14 September. 261-264.

Zhou, Z.-H., Wu, J. & Tang, W. Ensembling neural networks: Many could be better than all. *Artificial Intelligence*. 137 (2), 239-263.

Zhou, Z.-H., Wu, J.-X., Wu, Y.J. & Chen, S.-F. (2001) Genetic Algorithm based Selective Neural Network Ensemble. *In Proceedings of the 17th International Joint Conference on Artificial Intelligence*. Seattle, WA, USA. 4-10 August 2001. ACM. 797-802.