Predicting Forex Currency Fluctuations Using a Novel Bio-inspired Modular Neural Network

Christos Bormpotsis

Staffordshire University
School of Computing and Digital Technologies

A Thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

December, 2023

"Faye and Saki are the two amazing souls who have loved and supported me unconditionally. I am forever grateful for the values and lessons you instilled. Thank you for being my pillars of strength and the best parents anyone could ask for."

Abstract

This thesis explores the intricate interplay of rational choice theory (RCT), brain modularity, and artificial neural networks (ANNs) for modelling and forecasting hourly rate fluctuations in the foreign exchange (Forex) market. While RCT traditionally models human decision-making by emphasising self-interest and rational choices, this study extends its scope to encompass emotions, recognising their significant impact on investor decisions. Recent advances in neuroscience, particularly in understanding the cognitive and emotional processes associated with decision-making, have inspired computational methods to emulate these processes. ANNs, in particular, have shown promise in simulating neuroscience findings and translating them into effective models for financial market dynamics.

However, their monolithic architectures of ANNs, characterised by fixed structures, pose challenges in adaptability and flexibility when faced with data perturbations, limiting overall performance. To address these limitations, this thesis proposes a Modular Convolutional orthogonal Recurrent Neural Network with Monte Carlo dropout-ANN (MCoRNNMCD-ANN) inspired by recent neuroscience findings.

A comprehensive literature review contextualises the challenges associated with monolithic architectures, leading to the identification of neural network structures that could enhance predictions of Forex price fluctuations, such as in the most prominently traded currencies, the EUR/GBP pairing. The proposed MCoRNNMCD-ANN is thoroughly evaluated through a detailed comparative analysis against state-of-the-art techniques, such as BiCuDNNLSTM, CNN–LSTM, LSTM–GRU, CLSTM, and ensemble modelling and single-monolithic CNN and RNN models. Results indicate that the MCoRNNMCD-ANN outperforms competitors. For instance, reducing prediction errors in test sets from 19.70% to an impressive 195.51%, measured by objective evaluation metrics like a mean square error.

This innovative neurobiologically-inspired model not only capitalises on modularity but also integrates partial transfer learning to improve forecasting accuracy in anticipating Forex price fluctuations when less data occurs in the EUR/USD currency pair. The proposed bio-inspired modular approach, incorporating transfer learning in a similar task, brings advantages such as robust forecasts and enhanced generalisation performance, especially valuable in domains where prior knowledge guides modular learning processes. The proposed model presents a promising avenue for advancing predictive modelling in Forex predictions by incorporating transfer learning principles.

Acknowledgment

I am writing to express my most profound appreciation and gratitude to my supervisors, Prof. Dr Mohamed Sedky and Dr Asma Patel, for their outstanding guidance, support, and mentorship throughout my work under their supervision. Their unwavering commitment to my growth and development has played a pivotal role in shaping my academic and professional journey. I am truly fortunate to have had supervisors with exceptional expertise in their respective fields. Their wealth of knowledge and willingness to share and impart valuable insights have broadened my academic horizons. In addition, their guidance has expanded my intellectual capabilities and inspired me to explore new avenues of research and innovation.

Furthermore, I am deeply grateful for their constant encouragement and belief in my abilities. Their constructive feedback and thoughtful guidance have pushed me to surpass my limitations and strive for excellence. Their mentorship has instilled in me a sense of perseverance and determination, empowering me to tackle challenges with confidence and resilience. In addition to their professional expertise, I am grateful for their unwavering support on a personal level. Their approachability and willingness to lend an ear during professional and personal matters have created a nurturing and conducive environment for growth. Similarly, their genuine care and understanding have given me the strength and encouragement to navigate through the ups and downs of my PhD journey.

I would also like to thank the research team from Staffordshire University and colleagues who collaborated and shared their knowledge and experiences. The collective efforts and camaraderie have made this journey productive, enjoyable, and memorable.

Last, I would like to express my heartfelt appreciation to my parents and my lovely wife, Marianna, for their unwavering support, understanding, and encouragement throughout this endeavour. Their love, efforts and belief in my abilities have been the driving force behind my accomplishments.

Publications

Bormpotsis, Sedky, and Patel (2023). Predicting Forex Currency Fluctuations Using a Novel Bio-Inspired Modular Neural Network. *Big Data and Cognitive Computing*, 7(3), 152.

Contents

List of Figures

List of Tables

1. Introduction

The chapter introduces the background of Forex and interdisciplinary fields such as AI and RCT, the research motivation, the aim, objectives, and research questions to be achieved. It reviews the methods used in this thesis to discover and evaluate deep learning approaches for Forex price fluctuations. Finally, the chapter deduces the dissertation structure.

1.1 Background

The foreign exchange (Forex) market, a global and highly liquid financial market for currency exchange, plays a critical position in international trade and investment. Its continuous operation and substantial trading volume make it an attractive choice for investors, leading to a growing number of individuals transitioning from the stock market to Forex. It substantially influences contemporary international economies concerning economic expansion, global interest rates, and financial equilibrium (Mai, Chen, Zou, & Li, 2018).

Researchers emphasised that due to the substantial magnitude of daily transactions, investors and financial institutions possess the potential to yield significant returns by accurately speculating and signifying fluctuations in Forex exchange rates (Hayward, 2018). Computational advancements, such as Artificial Intelligence (AI) and its machine and deep learning subfields, are utilised in the stock and Forex markets by providing traders with new ways to scrutinise market data and seek to find potentially profitable trading options (Berradi, Lazaar, Mahboub, & Omara, 2020; Ray, Khandelwal, & Baranidharan, 2018). However, recent AI tendencies have revealed that the synergy between neuroscience, machine, and deep learning is necessary for more informed and better-comprehended decision-making (Russin, O'Reilly, & Bengio, 2020).

Likewise, the convergence of neuroscience and economic theories, notably Rational Choice Theory (RCT), could present a compelling avenue for advancing our understanding and potentially enhancing the predictive capabilities of AI models in the complicated domain of Forex trading price directions (Awunyo-Vitor, 2018; Pujara, Wolf, Baskaya, & Koenigs, 2015). RCT serves as a diverse framework for understanding societal dynamics rooted in the assumption of individual rationality. While offering descriptive clarity to societal issues, this theoretical approach, particularly those previously poorly characterised, often reveals the inherent indeterminacies in social relations and individual perspectives. The indeterminacies RCT identifies extend to social choice, emphasising the complexity of societal decision-making processes. The fundamental tenets of RCT, revolving around individual rationality and the pursuit of optimal outcomes, align seamlessly with utilitarianism to promote people's wealth. RCT provides a comprehensive lens through which to analyse societal behaviour and decision-making by focusing on states of affairs and considering factors that motivate individuals. RCT in financial markets, influencing investors' economic decision-making processes, constitutes a multifaceted cognitive phenomenon intertwined with rational self-interest. Individuals navigate diverse financial conditions in this intricate landscape to derive optimal net benefits (Arnott & Gao, 2019). Furthermore, RCT illuminates how investors assimilate information, exhibit demeanours across various social and economic contexts—notably financial markets like Forex—and formulate trading strategies (Buskens, 2015). Nevertheless, while RCT underscores the centrality of rationality in decision-making, it is imperative to recognise that emotions influence investors' choices (Lerner, Li, Valdesolo, & Kassam, 2015).

Moreover, contemporary insights from neuroscience have contributed to explicating decision-making processes by elucidating the complex connections between rational deliberation and emotional responses mediated by distinct brain regions, such as the insular and prefrontal cortex (Lamm & Singer, 2010; Rilling & Sanfey, 2011). This emerging understanding highlights the interplay between cognitive rationality and affective elements, providing a more nuanced comprehension of how economic reasoning is constructed. Recent studies indicated that behavioural facilitation in the human brain regions, such as the amygdala and hippocampus, is related to emotions and memory retrieval (Eichenbaum, 2004; LaBar & Cabeza, 2006; Olsen, Moses, Riggs, & Ryan, 2012; Phelps & LeDoux, 2005; Roozendaal, McEwen, & Chattarji, 2009). The amygdala and hippocampus correlate with cortical areas, such as the frontal and temporal lobes, including brain parts like the striatum, insular, and prefrontal cortex (Pizzo et al., 2019). Current neuroscientific investigations imply that these parts of the brain are accountable for the individuals' procedural learning, reasoning, and emotions and are likely crucial for decision-making under financial risk conditions (Grossmann, 2013; Loued-Khenissi, Pfeuffer, Einhäuser, & Preuschoff, 2020; McEwen et al., 2015; Price & Drevets, 2010; Ruissen, Overgaauw, & De Bruijn, 2018; Tsukiura, Shigemune, Nouchi, Kambara, & Kawashima, 2013).

AI algorithms, such as Artificial Neural Networks (ANNs), have emerged as a powerful, innovative mechanism for simulating brain functions, such as self-intuition and Natural Language Processing (NLP) linked with emotions, to comprehend information processing and evaluate the possible contingencies to arrive at optimal decision prospects (Abiodun et al., 2018; Fermin, Friston, & Yamawaki, 2022). NLP techniques can be applied to financial textual data to analyse sentiment. Sentiment analysis can help gauge the collective mood of traders and investors, which, combined with economic indicators such as closing, can better anticipate price market movements (Jing, Wu, & Wang, 2021). Hence, traders and institutions increasingly use social media analytics tools to track and analyse trends on platforms like Twitter to help traders make informed decisions (Herrera, Constantino, Su, & Naranpanawa, 2022; C. Wang, Shen, & Li, 2022). Moreover, different ANN types, such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks, have been employed against traditional methods, such as Auto-regressive Integrated Moving Average (ARIMA) and Support Vector Regression (SVR), in contemplating the future price direction applied to a non-stationary time series (Bathla, 2020; M, E.A., Menon, & K.P., 2018).

While ANNs and their refined techniques, including CNN and LSTM, have demonstrated the capacity to recognise patterns and trends in financial markets, investigating these methodologies reveals a crucial research problem. The inherent monolithic architectures of these models pose substantial challenges, complicating the pursuit of more effective pattern recognition in the complex dynamics of financial markets defined below:

- Limited scalability and lack of flexibility: Monolithic architectures may be more challenging to scale because they are not easily divided into shorter, independent modules that can be developed and added to the architecture as needed (Amer & Maul, 2019);
- Difficulty understanding and modifying the architecture: Monolithic architectures can be challenging to understand, maintain, and change, primarily as their size becomes more extensive. Thus, updating the architecture as data or market conditions can be challenging (Ali, Sarwar, Sharma, & Suri, 2019);
- Increased risk of failure: Because monolithic architectures are complex to understand and modify, there is an increased risk of failure when making changes to the architecture. Hence, fixing it can be computationally costly and time-consuming (Yarushev & Averkin, 2018).

1.2 Research Motivation

The motivation behind this thesis is to address the challenges and limitations of the monolithic architectures mentioned above, adopting a bio-inspired approach that draws inspiration from the modularity in the human brain's neural pathways. Towards this direction, recent neuroscience-informed studies have shown that designing computational models mimicking brain areas such as the prefrontal cortex and anterior insula splitting a complex task into modules can execute the task better by presenting adequate performance (Lydon-Staley, Ciric, Satterthwaite, & Bassett, 2019; Morrone, Weber, Huynh, Luo, & Cornell, 2020; K.-J. Wang & Zheng, 2019).

To this end, this thesis seeks to design a novel neuro-informed Modular Neural Network (MNN) utilising a CNN integrating RNNs represented in Chapter 4, attempting to simulate the brain topological modularity, where separate neurons receive inputs from other cells across different levels of hierarchical process memory (Shi et al., 2018; Tzilivaki et al., 2019). Thus, the proposed bio-inspired MNN could be applied during high-risk financial decision-making to handle price Forex fluctuations better against the monolithic architectures.

Furthermore, convolutional RNNs stability could be significantly improved by utilising orthogonal kernel weight initialisation, tackling vanishing issues coupled with Monte Carlo dropout to capture the uncertainty of the MNN, enhancing further the outcomes of the network (Arjovsky, Shah, & Bengio, 2015; Duan & Wang, 2016; Sadr, Gante, Champagne, Falcao, & Sousa, 2022). Finally, the prior knowledge acquired from the proposed MNN of this thesis is partially transferred to a relevant task under data scarcity, enhancing the generalisation performance of ANNs (Cao, Jia, Ding, & Ding, 2021; Ding, Ding, Zhao, Cao, & Jia, 2022; He et al., 2020; B. Li & Rangarajan, 2022).

In essence, the synthesis of neuroscience and economic theories such as RCT discussed above holds the potential to enrich the development of bio-inspired AI models tailored for Forex trading. By bridging the interval between the cognitive and economic aspects of decision-making, this interdisciplinary approach may pave the way for more robust and suitable AI systems capable of navigating the intricate landscape of financial markets.

1.3 Research Aim

This thesis aims to refine and optimise ANNs prediction accuracy in the Forex market by proposing a novel modular neural network model. Furthermore, the thesis delves into the intricate details of the MNN model, drawing inspiration from the neuroscience dynamics of emotions and rationality in human decision-making, mainly as interconnected with brain regions such as the insular and prefrontal cortex.

1.4 Research Objectives

For the fulfilment of the research aim of this thesis, the objectives are:

- To conduct a comprehensive interdisciplinary literature review on neuroscience advancements and computational models to shed light on Forex forecasting.
- To propose a novel bio-inspired modular neural network in an effort to better predict price fluctuations in exchange rates and investigate the potential benefits of modularity in this context.
- To explore the advantages of incorporating a new adaptative mechanism consisting of Monte Carlo dropout and orthogonal kernel initialisation into recurrent layers within a convolutional modular network, replacing the standard pooling layer of a typical and conventional CNN for predicting Forex price fluctuations. Assess their impact on prediction error, uncertainty quantification, and the optimisation process.
- To conduct a comparative analysis between the proposed modular neural network, monolithic single, ensemble and state-of-the-art hybrid architectures, evaluating their performance in prediction error and their ability to capture complex Forex market dynamics.
- To investigate how modular neural network architectures can leverage knowledge gained from the proposed model to enhance the generalisation capabilities of ANNs in Forex predictions under data scarcity conditions, exploring techniques such as modular partial transfer learning.

1.5 Research Questions

This thesis addresses three research questions (RQs) corresponding to the above-mentioned objectives. These RQs are the following:

- **RQ1**: How do bio-inspired modular neural network architectures outperform monolithic ANN architectures in predicting price fluctuations in exchange rates, considering decreasing prediction error and the ability to capture complex market dynamics?
- **RQ2**: What are the potential benefits of incorporating Monte Carlo dropout and orthogonal weight initialisation methods within modular neural network architectures for predicting Forex price fluctuations regarding improved prediction error, uncertainty quantification, and optimisation process enhancement?
- **RQ3**: How can modular neural network architectures leverage knowledge gained from the proposed model to improve the performance and generalisation capabilities of ANNs in Forex predictions, particularly in data scarcity scenarios in a similar task?

1.6 Research Design

The thesis adopts a positivist paradigm, a philosophical approach associated with a deductive research method and a quantitative data analysis procedure. The goal is to propose a novel bio-inspired modular neural network that may contribute to the anticipation and knowledge of exchange rate price fluctuations against state-of-the-art hybrid, ensemble and single monolithic architectures. Positivism generally relies on the belief that understanding can be derived from observable facts and their analysis as is correlated with the deductive method (Park, Konge, & Artino, 2020). In a deductive approach, scholars start with a theory or formulate research questions and then collect and analyse data to test or confirm the idea. Furthermore, quantitative methods involve the collection and analysis of numerical data. This could include statistical studies to identify patterns or relationships. Hence, the research philosophy conducted in this thesis employs a positivist philosophical paradigm, traditionally associated with a deductive approach and quantitative method that can be used to comprehend and describe the aspects of financial decision market behaviour (Saunders, Lewis, & Thornhill, 2019).

Furthermore, an interdisciplinary literature review (Chapter 2) is conducted to identify the most relevant evidence in the context of recent neuroscience findings to inform the development of the proposed modular network (Chapter 4). In addition, related works in the machine and deep learning applications in the context of financial market predictions were also reviewed, identifying the most relevant for comparison with the proposed model. Finally, the use of the quantitative methodology in this thesis involves the collection of scientific data that is precise and based on measurement. The data was collected from the Yahoo Finance API and the Twitter Streaming API for the EUR/GBP Forex currency pair for two years (2018-2019).

As mentioned above, this thesis proposed a bio-inspired modular neural network compared with state-of-the-art and monolithic architectures. Additionally, this thesis employs three objective evaluation metrics, namely Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), and Mean Squared Logarithmic Error (MSLE), to provide evidence of the proposed model performance compared to the benchmark models given in Chapter 5.

Finally, this thesis investigates the EUR/USD exchange rate for one year (2020). It collects data from the same APIs mentioned above, transferring the acquired knowledge from the previous anticipation of the EUR/GBP pair to the EUR/USD price fluctuations to a simple modular ANN. MSE, MAPE, and MSLE have also been used to evaluate the performance of simple modular ANNs with and without partial transfer learning. The thesis research outcomes yielded price movement predictions for EUR/GBP and the EUR/USD exchange rates. Figure 1.1 illustrates the research design of this thesis.

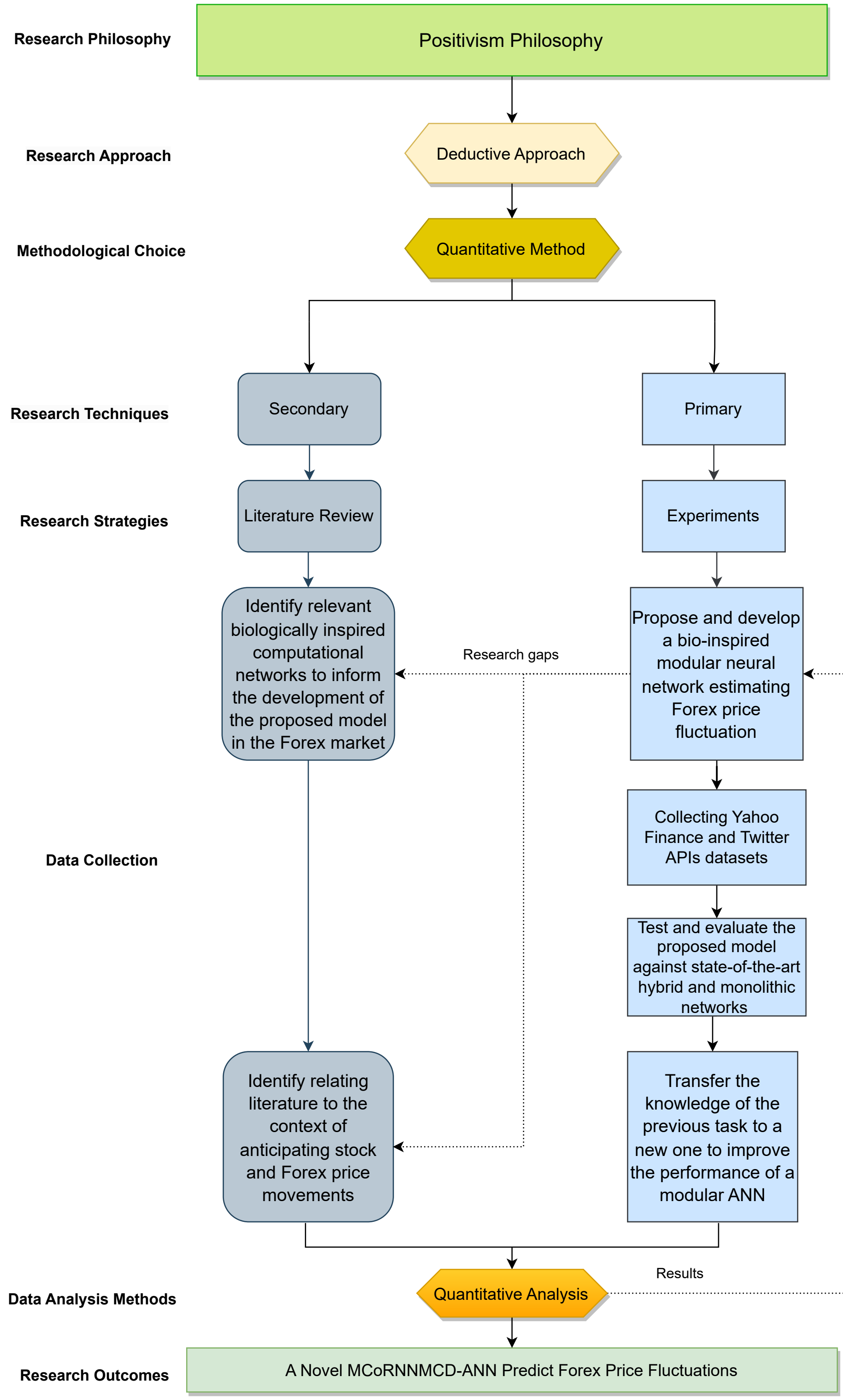


Figure 1.1: Research Design

To manage the research questions, the research methodology of this thesis is composed as follows:

- First, this thesis examined how sentiments in conjunction with the RCT, used to model human decision-making linked with specific brain regions, can be simulated with the help of recent computational intelligence-informed neuroscience findings. Moreover, this thesis explores if the proposed modular neural network inspired by the brain can outperform monolithic and state-of-the-art ANN architectures in forecasting price fluctuations in Forex to address the research issues associated with RQ1.

- Second, this thesis has set a mechanism based on the potential benefits of incorporating Monte Carlo dropout and orthogonal kernel initialisation into recurrent layers within the proposed modular convolutional network, replacing the standard pooling layers of a typical CNN to address the research issues associated with RQ2 aiming to enhance the model predictability performance in Forex price fluctuations.

- Finally, this thesis studied the application of modular partial transfer learning techniques to suggest practical solutions to address the research issues associated with RQ3 to improve the reliability and robustness of modular ANN prediction models in the Forex market.

1.6.1 Modelling

As already discussed, potentially enriching the Forex predictive models may require a combination of rationality and emotional awareness, as market trends influence investors' decisions in the face of anticipating potential losses or gains. Therefore, this thesis established a framework consisting of mental models for investor rationality and emotions based on deductive and probabilistic inferences—conceptual models to comprehend mental models from the determinants of the rationality and sentiments of investor decision-making. Furthermore, AI systems endeavour to simulate conceptual models in complex financial markets such as Forex in the light of more useful predictive outcomes. Consequently, design potentially more accurate computational models to enhance the predictability and robustness of price movements in the Forex market.

To better comprehend the decision process of investors, the functions of the ventromedial prefrontal cortex (vmPFC), a part of the prefrontal cortex in the mammalian brain and anterior insula (AIC), could be incorporated into a mental model creation for Forex trading divided into subtasks. For the first task, using the vmPFC as a basis for rational thinking involved in decision-making can create a guideline that a mental model would follow. The second task using the anterior insula as a basis for sentiment processing associated with emotional awareness and empathy can comprise emotional intellect in the mental model. It is noteworthy that the vmPFC and the AIC have been investigated in Chapter 2 in detail, as it is crucial to understand their functions in financial decision-making according to recent neuroscience findings.

While the mental model provides a general understanding of how the brain works, the conceptual model clarifies the different inputs, outputs, and constraints governing the model. Therefore, the conceptual model can be used after the mental model to help develop a more structured approach to formalise the mental model, defining the various components and establishing a framework for analysis and development. For instance, the conceptual model can translate the mental model, including two subtasks separating inputs. First, rational decision-making that tracks exchange rate prices representing the vmPFC to drive determinations based on market trends; second, employing techniques such as sentiment analysis and accommodating its decision-making representing the AIC accordingly. These representations will consist of the inputs in the machine model.

AI modelling converts the conceptual model's outputs as inputs into the proposed modular neural network that can be trained on the yielded Forex data from Yahoo Finance (exchange rate prices) and Twitter streaming (sentiment analysis) APIs. This procedure involves selecting appropriate algorithms, such as CNN and RNN, as discussed and analysed further in Chapter 2. Moreover, the simulation of the conceptual models from the AI machine system demonstrates proof of concept of how the proposed modular model of this thesis supports learning about the rational thinking and emotional processing of investors' mental states, trying to address the complex non-linear problem of predicting Forex price fluctuations. Finally, for a proof of concept implementation of the proposed modular neural network's performance, this thesis conducts a comparative analysis with state-of-the-art monolithic neural networks evaluated based on objective metrics such as MSE, MAPE, and MSLE, as presented in Chapter 5. The state-of-the-art was chosen from Chapter 2 as the benchmark for comparison with the proposed model of this thesis because they were successful models in the field at the time of research and provided comprehensive information to replicate their designs and parameters.

Furthermore, establishing baseline models is essential for effectively segmenting the input domain in Forex predictions. These baselines serve as a reference point, enabling synthesising the proposed bio-inspired modular architecture. Drawing motivation from the literature presented in Chapter 2, the modular architecture is strategically designed, leveraging insights from established baselines to enhance the overall robustness and reliability of the predictive models. This approach could ensure a comprehensive understanding of the input space, laying the groundwork for more sophisticated and accurate predictive models in the dynamic context of Forex market analysis. Finally, the Modular Partial Transfer Learning technique is implemented in a relevant task under data scarcity using pre-training modules of the proposed model to enhance the generalisation abilities of ANNs in Forex. Figure 1.2 illustrates the modelling of this thesis.

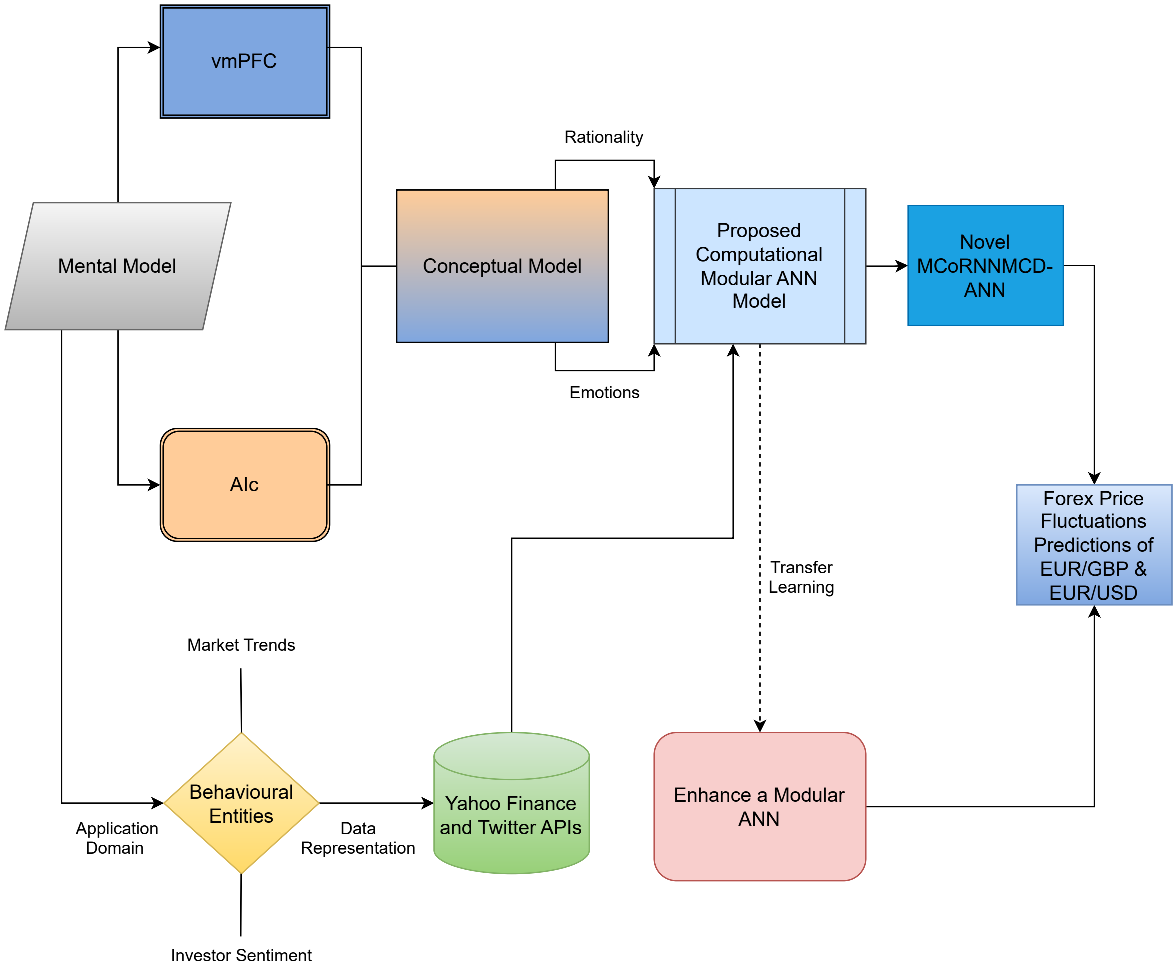


Figure 1.2: Modelling

1.6.2 Decoding Forex Market Dynamics

Financial markets are non-linear and subject to fluctuations due to many events, such as political events, investor behaviour, and central banks, that can directly or indirectly influence market trends. The impact of such events can be considered actively or passively over time, and there are several cases where market predictions may fail. Nevertheless, sentiments are a major driving force in the stock market, and historical stock market data can provide traces of the effect of the public mood at a particular time. However, evaluating the factors and their intensity can be challenging. Therefore, neuroscience-informed methods integrated into the machine and deep learning complementing techniques may enhance prediction performance. Moreover, hyperparameters and neural network architectures can be crucial in deriving model complexity. Identifying such complexity for various computational models can be a research direction to learn inherent market patterns for reliable forecasting.

Based on recent neuroscience findings and brain modularity, this thesis aims to present a novel modular neural network that could effectively predict price movement in the Forex market. The Forex market is highly dynamic, and various economic and social factors can influence it. Modular architectures can model these complex dependencies by allowing the design and training of separate modules to different model factors that may impact the forex rates. For example, one module can model economic indicators such as the closing price of exchange rates, and another specialises in modelling sentiment scores extracted from Twitter. Consequently, a more robust and accurate neural network incorporating these modules' output into a final decision model could enhance Forex predictions.

Additionally, modular architectures can adjust more to changes in the forex market because individual modules can be added, removed, or modified to handle changing market conditions. This process can help the neural model stay up-to-date and factual even as the market evolves, overcoming the monolithic ANNs issues to address the Forex data. Finally, inspired by how the brain restores memory and uses past information to pass into a new task aiming to solve an issue, this thesis employs the transfer learning technique. Hence, the knowledge acquired from the predictions of price fluctuations of the EUR/GBP rate by the proposed model was assigned to a modular ANN to enhance the predictions of the price movements of EUR/USD using less data than the previous task.

1.7 Contributions

In this thesis, to address the limitations of the monolithic ANN architectures yielding improved predictions in Forex price fluctuations, presenting the three main contributions listed below:

- A novel modular neural network architecture inspired by rational choice theory and cognitive neuroscience to model human decision-making in Forex price fluctuation predictions. By incorporating modularity, rationality, and emotions, this thesis aims to represent a significant novelty, pushing the boundaries of existing knowledge and providing new insights into predictive modelling in the Forex domain.
- A new adaptation mechanism consists of Monte Carlo dropout and orthogonal kernel initialisation, incorporating it into a recurrent layer within a convolutional modular network, replacing the traditional pooling layers. This novelty allows for the adaptive width of the recurrent layers

to dynamically adjust its capacity based on the complexities of the Forex data.

- A novel approach that utilises partial transfer learning in the context of Forex prediction that mitigates the adverse effects of data scarcity by effectively using information from previous comparable tasks. The newly modular partially transferred knowledge could help capture complex dynamics, improve prediction errors, and address the challenges posed by data scarcity.

1.8 Thesis layout

The rest of this thesis is arranged as follows:

- Chapter 2 describes the RCT extended with emotions to model human decision-making associated with brain regions that could be formulated with the power of computational intelligence. Furthermore, it investigates the relevant studies in neuroscience and AI, providing an overview of related works in the machine and deep learning model applied to the predictions of financial markets.

- Chapter 3 presents the foundation of ANNs and other innovative techniques such as Monte Carlo dropout, orthogonal initialisation and gradient optimisers.

- Chapter 4 presents the proposed novel neuroscience-informed modular architecture for Forex market predictions, incorporating Monte Carlo dropout and orthogonal weight initialisation, utilising the modular partial transfer learning technique to a relevant data-limited task.

- Chapter 5 presents the set-up of the proposed bio-inspired modular network parameters and benchmark models, comparing their performance based on their hourly closing price and sentiment scores inputs.

- Chapter 6 presents the thesis conclusion by summarising the main contributions of this thesis, discussing the limitations of the work and glancing at promising routes for future research.

2. Incorporating Rational Choice Theory With Neuroscience and AI Systems

This chapter serves as a vast resource for understanding the integration of Rational Choice Theory with recent neuroscience findings simulating AI Systems in the context of Forex. The thesis advocates for adopting a semi-systematic review, positioning it as a nuanced middle ground that transcends the rigidity of a fully systematic approach while retaining the flexibility of a narrative approach in the broad business research subject (Snyder, 2019). This recommendation is particularly pertinent in complex interdisciplinary fields, such as the intersection of AI with Forex predictions, where a diverse array of research traditions encompass statistics, cognitive science, and computing. Semi-systematic reviews and narratives have also proved sufficient to understand better complex areas like NLP and the evidence supporting ML use across clinical trials (Weissler et al., 2021; T. Zhang, Schoene, Ji, & Ananiadou, 2022). A critical analysis was performed to foresee Forex hourly price fluctuations, selecting pertinent sources from Yahoo Finance and Twitter Streaming APIs for the EUR/GBP currency pair.

Moreover, this thesis considered 19,920 recovered bibliographic records, focusing on renowned databases, including Scopus (n = 14,238) and IEEE Xplore (n = 5682). Motivated by this study's objective to revise the monolithic computational model, aiming to enhance the potential of neural networks for more accurate predictions, the following targeted keyword searches were employed, focusing on topics such as: "brain modularity", "financial decisions under risk", "biologically inspired machine", "rational choice theory for finance", "transfer learning process in the brain", "transfer learning in nlp", "transfer learning in financial markets", "transfer learning in risk analysis", "machine learning for Forex/stock predictions", "deep learning for Forex/stock price predictions", "social media analysis for Forex/stock predictions", "NLP for finance", "neuroeconomics", "artificial neural networks mimic brain", "Twitter sentiment analysis for Forex/stock markets", "CNN for Forex/stock predictions", and "RNN for Forex/stock predictions".

The studies were reviewed by removing duplication (n=3791) using the RefWorks citation manager, scanning titles, authors, and DOI, and retaining 16,129 reports. After the screening process, by using exclusion criteria, such as non-English language usage, titles, and the abstract being irrelevant to the thesis's aim, a subset of 967 studies were chosen. Behind the full-text screening, a more thorough examination of the abstracts and the content of the remaining studies applied, and 367 reports were selected as eligible.

The final inclusion criteria resulted in the choice of 160 studies that were identified as the most relevant investigations to this thesis's aim of enhancing the potential of neural networks for more accurate predictions in the context of potentially better anticipating Forex hourly price fluctuations. These inclusion criteria have been formulated as below:

- Bio-inspiration in Computational Models: Studies must showcase a consideration or incorporation of bio-inspired elements and how they can be applied in computational models.
- Relevance to Financial Predictions Using Neural Networks: Studies must demonstrate direct relevance to applying neural networks for more accurate predictions in financial markets, specifically focusing on Forex hourly price fluctuations.
- Incorporation of NLP Methods in Financial Research: Selected Studies should incorporate NLP methods, particularly in financial research and predictions, including sentiment analysis, social media analysis, and machine learning for Forex/stock predictions.
- Transfer Learning in Financial Predictions: Studies should address the application of transfer learning processes in the brain, particularly in financial markets.
- Methodology Contribution and Peer-Reviewed Validation: Studies must contribute to the methodology, introducing new approaches, models, or techniques relevant to enhancing the potential of neural networks in financial predictions. Additionally, the selected articles must be peer-reviewed reports from reputable publishers, ensuring the utilisation of reliable and high-quality sources.

Figure 2.1 illustrates the study selection process that has been followed.

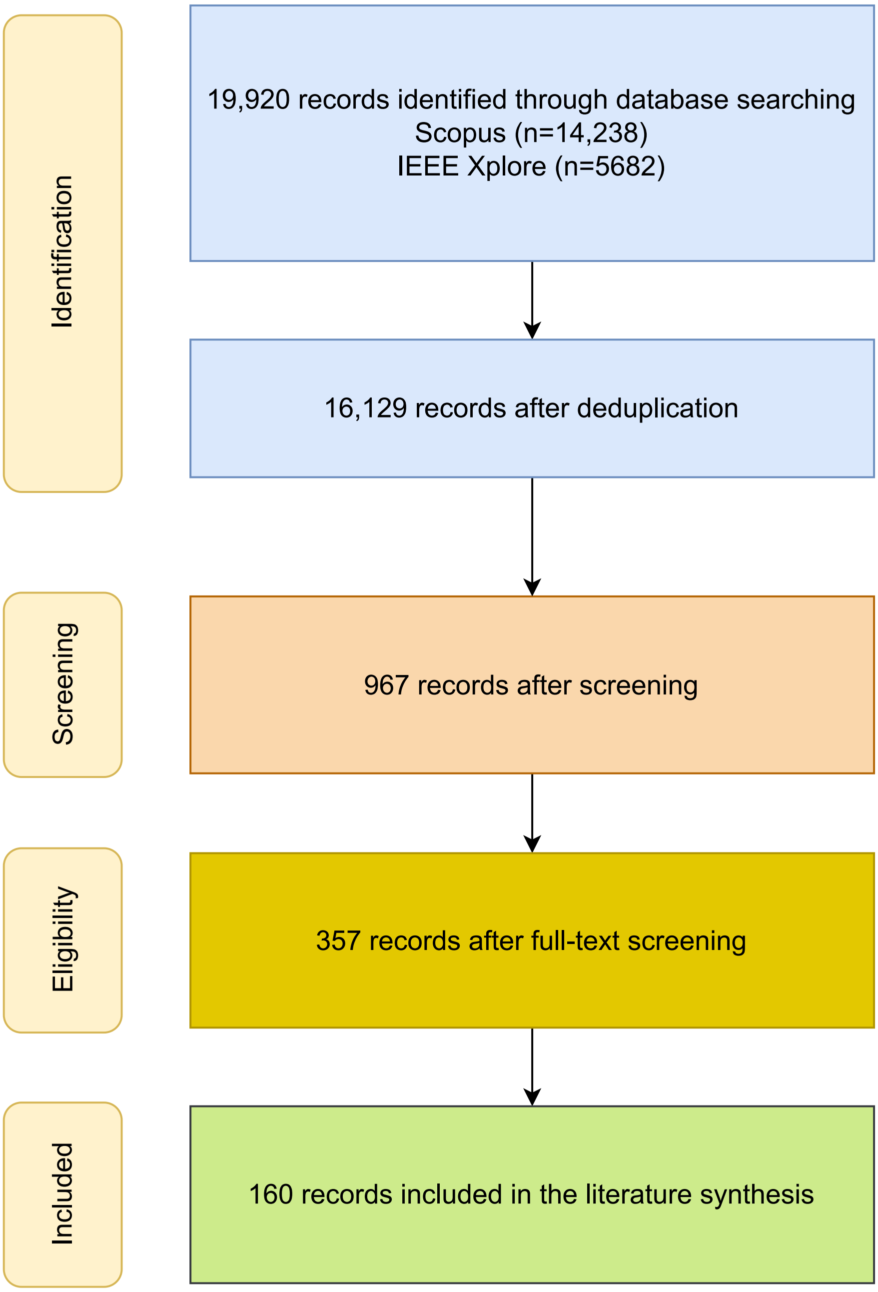


Figure 2.1: The Scopus and IEEE Xplore databases were investigated. The flowchart details justifications for excluding studies from the data extraction and quality assessment.

2.1 Rational choice theory (RCT)

Rational Choice Theory (RCT), a foundational concept in neoclassical economics, asserts that individuals make decisions based on rational assessments of their preferences and self-interest to optimise outcomes. Applied broadly in diverse contexts, including financial markets, RCT analyses consumer behaviour, market dynamics, game theory, and economic phenomena with the premise that individuals act in a manner consistent with maximising their utility (Zey, 2015).

Despite its wide-ranging applicability, RCT has faced criticism for its inherent limitations. Detractors argue that the theory's framework overlooks external factors influencing decision outcomes, particularly emotions, which play a significant role in financial decision-making (Lerner et al., 2015). In the broader economic context, RCT's reception indicates the evolving landscape in economic thought. Criticism directed towards RCT aligns with a broader discourse challenging the dominance of neoclassical economics and this discourse, emphasising the limitations of selfishness and equilibrium, thereby promoting the diversification of economic theories. Behavioural economics, new institutional economics, and other emerging perspectives exemplify this shift, signalling a departure from neoclassical hegemony (Colander, Holt, & Rosser Jr, 2004).

However, proponents assert that RCT remains a valuable tool, offering a reasonable basis for understanding economic decisions (Awunyo-Vitor, 2018). In connecting RCT with neuroscience, recent studies shed light on the neural mechanisms underlying this theory, notably emphasising the roles of the prefrontal cortex (Kroker et al., n.d.; Livet, 2010). Furthermore, Cecchi et al. (2022) propose a correlation between heightened broadband gamma activity (BGA) in the ventromedial prefrontal cortex (vmPFC) and dorsal anterior insula (daIns) and fluctuations in mood. Elevated BGA in these regions is linked to heightened or diminished moods, thereby influencing risk-taking behaviours by amplifying the significance attributed to potential gains or losses in decision-making processes. Furthermore, recent studies have shown that the vmPFC is activated during decision-making tasks involving uncertainty, such as those in the financial markets (Dennison, Sazhin, & Smith, 2022).

This intersection of RCT with neuroscience highlights the intricate relationship between cognitive processes and economic decision-making. The prefrontal cortex's involvement in mood regulation and risk perception underscores the multifaceted nature of decision-making, acknowledging the impact of emotions and neural states (De Martino, Kumaran, Seymour, & Dolan, 2006).

The AIC is another brain region activated during decision-making in financial markets to be involved in the experience of emotions, such as ambiguity (M. Srivastava, Sharma, Srivastava, & Kumaran, 2020). AIC has also been found to be activated during the experience of regret, which can influence future decision-making. For example, if an investor experiences regret after a poor investment decision, this can activate the anterior insula, reducing risk-taking behaviour in the future. Additionally, vmPFC connectivity changes are associated with increased risk-taking after total sleep deprivation (Y. Wang, Dai, Shao, Wang, & Zhou, 2022).

The vmPFC and the AIC can be simulated by AI systems to analyse the brain procedures, yielding more helpful information for its functionality in decision-making (R. Li & Zhang, 2020; J. Zhang, Yin, Chen, & Nichele, 2020). Consequently, this information can assist in modelling how decision-making applies to Forex market anticipation. AI-based models can also incorporate regret-based learning, considering past mistakes and modifying future decisions accordingly (Fan, Zheng, & Li, 2022). Furthermore, AI subsets, such as machine learning (ML) and deep learning (DL) models, can improve investors' decisions in the Forex market by providing them with predictive models and identifying patterns and relationships in data (Bag, Gupta, Kumar, & Sivarajah, 2021). These models can help investors make more informed decisions by providing estimates of future market movements and identifying patterns that are not immediately obvious to human investors.

Similarly, deep learning networks can be beneficial in the context of natural language processing (NLP), indicating topic-level representations of sentences in brain regions such as the medial prefrontal cortex using CNNs and RNN algorithms (Acunzo et al., 2022; V. & Bhattacharyya, 2022). By capturing their complex relationships, these models can learn meaningful representations of words, phrases, and sentences. This ability is crucial for language understanding and sentiment analysis in the Forex market. Figure 2.2 illustrates the CNN models of the sentences and the average word embeddings read by humans experiencing fMRI.

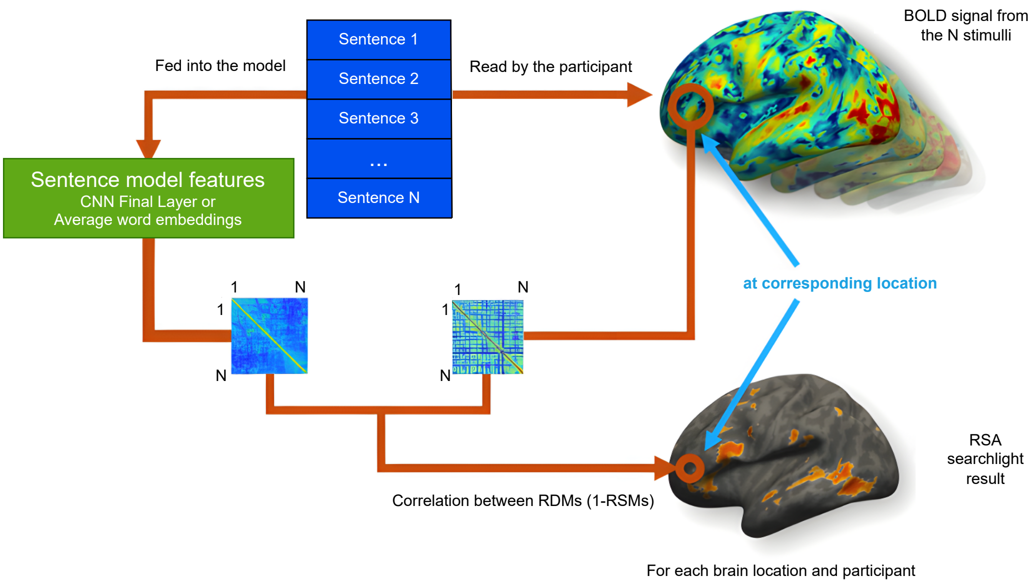


Figure 2.2: CNN models the sentences and the average word embeddings read by humans experiencing fMRI [Acunzo et al. (2022)].

In conclusion, AI methods can attempt to denote the vmPFC and the AIC in the decision-making of Forex investors by incorporating similar decision-making processes and considering emotional factors (Christov-Moore, Reggente, Douglas, Feusner, & Iacoboni, 2020; Qi, Khushi, & Poon, 2020). For instance, AI-based models incorporating deep learning capabilities can attempt to mimic the behaviour of the vmPFC in decision-making. In addition, computational models that consider emotional factors such as stress or fear can attempt to mimic the behaviour of the AIC. Accordingly, by simulating the behaviour of vmPFC and AIC, AI-based models can provide potentially more accurate predictions and drive better decisions in the Forex market. Furthermore, it is worth noting that recent investigations have demonstrated that cognitive biases and emotions often influence human decision-making in financial markets rather than only rational self-interest (Ahmad, 2020; Ishfaq, Nazir, Qamar, & Usman, 2020). Therefore, this has directed to the developing of AI techniques such as NLP that enhance investor decision-making in financial markets, incorporating emotions in the form of sentiment analysis Ren (2021).

As already been discussed, computational methods such as ANNs, CNN and RNN models are utilised to model brain regions such as the vmPFC and AIC to better understand rational behaviour in financial decision-making. Moreover, CNN and RNN are applied in time series predictions to estimate the future of Forex, as discussed below. However, computational models cannot fully replicate the complexity of the human brain, instead some of its functions, and there are still many challenges and limitations to be addressed and investigated.

2.2 Time Series Predictions

Forecasting using time series refers to using data sequences to measure the same thing over some time. Time series can be described as a series of numeric values, each with its time stamp represented by a group of labelled dimensions. For instance, K. Chen et al. (2019) utilised a short-term electric load forecasting model based on deep residual networks, focusing on capturing uncertainty within the models. Therefore, Monte Carlo dropout was implemented in their models to obtain the predictive probability, enabling probabilistic load forecasting. Their findings revealed that incorporating MCD, their proposed model achieves accurate results and demonstrates its effectiveness, outperforming existing models in the field.

Moreover, collecting time series data enables researchers to investigate the past, observe the present, and anticipate the future in financial markets (Zanc, Cioara, & Anghel, 2019). For instance, the data can be analysed using forecast approaches where the last hourly value can be utilised to estimate the value of the following hour by the exact value. The standard prediction models for time series include linear or non-linear approaches. Due to the intricacy of the Forex market, selecting a suitable method is essential in obtaining appropriate forecast results when predicting Forex time series (Shiao, Chakraborty, Chen, Hua Li, & Chen, 2019). Deep learning presents a unique computational learning technique that has received growing attention for financial forecasting, considering the difficulties in predicting the Forex changes (Aryal, Nadarajah, Kasthurirathna, Rupasinghe, & Jayawardena, 2019). In the below sections, substantial research has been conducted using time series-based approaches to explore how different financial elements affect Forex price fluctuations from computational economic and neuroscience points of view.

2.2.1 Neuroeconomics, Computational Knowledge and Complexity of the Forex Market

The assortment of neuroeconomics and the computational basis knowledge is a multi-interdisciplinary field combining computer science, economics, and neuroscience to understand decision-making processes better (Rangel, Camerer, & Montague, 2008). The Forex market is a complex dynamic system and the highest liquid financial market in the world, making it a valuable subject for study in this field. For example, the concept of bounded rationality could be an essential element of the forex market (Jiang, Li, Mai, & Tian, 2022). This notion refers to the idea that individuals have limited thinking abilities and therefore do not always make the best rational decisions (Sáenz-Royo, Chiclana, & Herrera-Viedma, 2022). Instead, they rely on heuristics or mental shortcuts, such as rules of thumb or gut instincts, to make decisions. Likewise, this idea explains why traders may make decisions that deviate from a purely rational analysis of market conditions. For example, a trader may be influenced by his emotions, such as sadness or happiness, or social interactions with other traders. Another influential factor in the complexity of the forex market is the presence of high levels of uncertainty (Hilbert & Darmon, 2020). This tension could result from modifications in the European Central Bank (ECB) and the Bank of England (BoE) policies. For example, if the ECB were to raise interest rates, it could lead to an appreciation of the Euro instead of the British Pound as investors seek higher returns on their investments. Conversely, if the BoE were to lower interest rates, it could lead to a depreciation of the Pound against the Euro.

The EUR/GBP currency pair representing is an interesting case for analysis due to the unique economic and political factors that influence its value for the impact of investors' sentiment on exchange rates (Alvarez-Diez, Baixauli-Soler, & Belda-Ruiz, 2019). For example, an event that has affected the sentiment towards the EUR/GBP pair was the United Kingdom's determination to leave the European Union, commonly referred to as Brexit. The uncertainty surrounding the terms of the UK's exit has caused fluctuations in the value of the British Pound. For instance, in the months directed up to the UK's original exit date in March 2019, the Pound declined significantly against the Euro as investors became wary of the potential negative consequences of Brexit. The forenamed factors are often difficult to predict and can affect the trader's financial decision-making processes.

Hence, there is a need for state-of-the-art computational systems to deal with specific characteristics of the problems that arise in the Forex market, such as high complexity, nonlinearity, strong presence of noise, uncertainty and changes in investor behaviour. Recent studies (Zhong & Enke, 2019) have shown that the usefulness of machine learning algorithms is becoming increasingly high in various industries, including financial market investment. This popularity is because machine learning models do not need to make presumptions about the data and can often yield more precise results than traditional econometric and statistical models. For instance, ANNs can analyse multiple data variables without assumptions, making them a powerful forecasting tool. Similarly, deep neural network models are more effective than traditional linear models in predicting financial markets (Nikou, Mansourfar, & Bagherzadeh, 2019). For instance, LSTMs can identify non-linear relationships between the dynamics of stock prices and the order book, which reflects the visible supply and demand for a stock. The superiority of neural networks over linear models can be attributed to their ability to accurately capture and represent these nonlinearities (Sirignano & Cont, 2019).

Other studies (Sim, Kim, & Ahn, 2019) sought to predict the S&P 500 index using the closing price and by choosing nine technical indicators: SMA, EMA, ROC, MACD, Fast %K, Slow %D, Upper Band, and Lower Band as predictors. In addition, a comparison between CNN, ANN and SVM has been conducted to decide which models are more accurate. The results demonstrated the effectiveness of CNN without using technical indicators compared to ANN and SVM models. Moreover, they concluded that technical indicators were not suitable as input features, as they presented similar behaviour to the moving pattern of the closing price, which led to poor performance. Likewise, Lanbouri and Achchab (2020) presented an LSTM network based on technical indicators to forecast the price of Amazon stock on high-frequency. During the evaluation of the LSTM performance, they conducted two experiments. The first experiment was conducted without technical indicators, employing the Open, High, Low, and Close (OHLC) prices and Volume. In the second one, they used five technical indicators, the EMA12, EMA25, MACD, Bollinger Up and Bollinger Down, the

OHLC prices, and the Volume. Their findings indicated that the closing price was predicted accurately without using technical indicators as input features. In the next section, the significance of the closing price will be examined in the context of Forex predictions.

2.2.2 Closing Price and Time Frame

One of the most critical factors defining the success of an investment in the Forex market is the ability to predict future market movements. The closing price is the best indicator to anticipate Forex markets (C.-C. Chen, Kuo, Kuo, & Chou, 2015). In addition, the closing price is considered the most helpful indicator to predict Forex markets because it is the price at which the call nears and reflects the market's general presumption. Furthermore, it is less susceptible to manipulation than other indicators, such as the opening price, because it reflects the prevailing market view at the end of the trading day.

The time frame of one hour works most suitable for anticipating financial markets because it is a shorter time frame than daily or yearly forecasting (Almasri & Arslan, 2018). Moreover, this shorter time frame allows for more accurate predictions because it could capture the volatility and uncertainty of the market. Besides, the hourly time frame is less affected by long-term trends, such as interest rate changes, which can significantly impact daily or weekly predictions. Likewise, the hourly time frame allows for more frequent predictions and is more responsive to short-term market fluctuations. ML-based models can also be used to predict the direction of currency prices rather than conventional statistical tools. As a result, these models can help investors to identify opportunities and manage risk more effectively (Khedr, Arif, El-Bannany, Alhashmi, & Sreedharan, 2021).

In conclusion, the closing price is the best indicator to utilise to anticipate Forex markets because it reflects the market's overall sentiment and is less susceptible to manipulation than other indicators. The time frame of one hour (or hourly forecasting) works best for anticipating Forex markets because it captures the volatility and uncertainty of the market and is more responsive to short-term market fluctuations. More recently, in terms of uncertainty estimation, researchers exploit Monte Carlo dropout layers by validating the effectiveness of the method of stock price prediction using convolutions and recurrent neurons (Alghamdi, Alotaibi, & Rajgopal, 2021).

AI and machine learning models can enhance the prediction of Forex markets by providing predictive models, identifying patterns, and reducing the impact of cognitive biases and emotions on investment decisions. Combining these methods can help investors create more knowledgeable choices and improve their performance in the Forex market. In contrast to the closing price, the brain's functionality under specific economic conditions and the computational representation of different types of ANNs influenced by recent neuroscience findings will be discussed in the next section. The main reveal is the importance of using computational models in different domains that can be applied to Forex markets, potentially boosting better anticipation of the Forex market.

2.3 Brain Structure and Computational Representation

Studying brain structure and its computational representation is a complex and multidisciplinary field that draws on neuroscience, cognitive psychology, and computer science knowledge. This field aims to understand how the brain processes information and how this information can be illustrated algorithmically (Kriegeskorte & Douglas, 2018). One of the critical areas of investigation in this field is neural networks, which are computational models inspired by the brain's structure and function (Kriegeskorte, 2015). Neural networks contain artificial neurons trained to perform specific tasks by adjusting the strengths of the connections between them. Another vital area of study is cognitive architectures, which are computational models that attempt to replicate the human brain's cognitive processes (Prieto et al., 2016). These models incorporate knowledge from neuroscience to represent different aspects of human cognition, including image recognition, natural language processing, and decision-making. The study of brain structure and computational representation is a rapidly growing domain that can furnish new understandings into how the brain operates and lead to the development of machine learning technologies to improve the predictive performance of investors' decision-making in Forex.

2.3.1 Brain Behaviour under Financial Risk

Frydman and Camerer (2016) highlighted the significance of financial decisions in shaping people's lives and the cognitive and neural procedures that influence these decisions. For example, households make decisions that violate sound financial principles, leading to underdiversified stock holdings. In addition, investors tend to over-extrapolate from past returns and trade too often, while top corporate managers, who are highly educated, make decisions that are affected by overconfidence and personal history. Principles from cognitive science can explain these behaviours by enabling more comprehensive and better collaboration between mathematical modelling, cognitive and neural measures, and observed behaviour in the interdisciplinary study of financial decision-making.

Tanaka, Yamamoto, and Haruno (2017) suggested that economic inequality can critically affect human mood states. Furthermore, the parts of the brain, such as the amygdala and the hippocampus, play a critical role in how people are affected by economic inequality and influence how individuals make decisions in risky situations. Yu, Mamerow, Lei, Fang, and Mata (2016) examined how age affects naturalistic risk-taking behaviour to specify neural procedures in which activity is linked. The regions observed were the bilateral striatum anterior insula and ventromedial prefrontal cortex areas associated with decision-making with emotion and value. The researchers uncovered that while these areas showed similar activities in the younger and older studied groups, the vmPFC showed different activities depending on the age when making decisions under risks due to changes in their neural processing over time.

Moreover, the vmPFC is a region of the brain that is involved in a variety of different psychological functions. These functions include decision-making, regulation of emotions, and recognition of facial expressions (Hiser & Koenigs, 2018). In the same direction, Peng et al. (2020) aimed to understand the connections between brain areas involved in a risk-taking task and to detect how age differences between participants affected these connections while playing the Balloon Analogue Risk Task. The brain activity data demonstrated that young and older adults had similar neural patterns when making decisions. However, their analysis also revealed that the connection between the vmPFC, dorsolateral prefrontal cortex, and anterior insula was more robust in older adults, meaning they could better make higher-quality decisions. Figure 2.3 shows the area of model nodes displayed in a 3D semi-transparent brain, and the four models exhibit different link directions between the VMPFC and DLPFC. VMPFC, the ventromedial prefrontal cortex; DLPFC, the dorsal lateral prefrontal cortex; AI

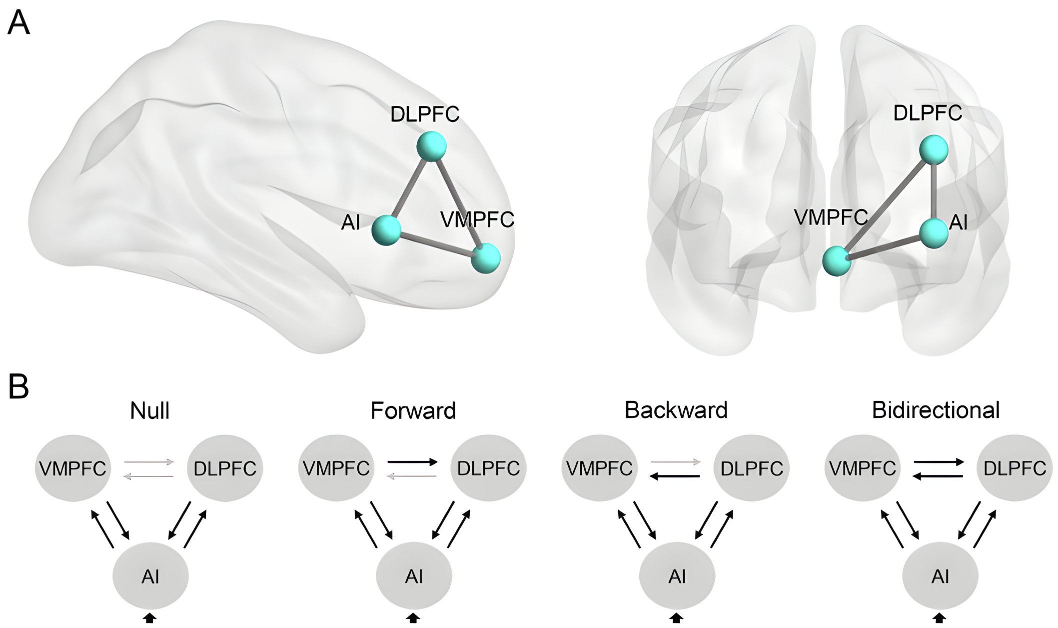


Figure 2.3: (A) The area of model nodes showed in a 3D semi-transparent brain. (B) Four models exhibit different link directions between the VMPFC and DLPFC. VMPFC, the ventromedial prefrontal cortex; DLPFC, the dorsal lateral prefrontal cortex; AI [Peng et al. (2020)].

Leong, Pestilli, Wu, Samanez-Larkin, and Knutson (2016) explored how individuals may display different attitudes towards risk, such as being unusually attracted to gambles with a slight chance of winning large amounts. Using a functional imaging technique indicated that the nucleus accumbens (NAcc) and anterior insula (AIns) areas in the brain are likely connected to exhibiting different risks. Furthermore, this connection correlated with an individual's preference for positively skewed gambles influences a person's behaviour in making uncertain decisions, meaning the person may weigh the risks and bonuses before determining. More recently, Leong et al. (2021) found that the AIns-NAcc tract was correlated with lower activity in the NAcc during anticipation of no incentives. While followed up with adolescents two years later, they could see a more robust connection of the AIns-NAcc tract was associated with greater motivation.

Lau, Ozono, Kuratomi, Komiya, and Murayama (2020) explored people's interest in gambling through curiosity or a reward, and their brain activity was monitored. Their experiments showed that when people accepted gambling, their brain activity was higher in a particular area of their brain called the ventral striatum than when they rejected it. Moreover, the activity spread to another part of the brain called the brain dorsal striatum when they made a decision. The neural substrates in different brain regions considering trust and risk in digital financial decision-making have also been investigated (Carbo-Valverde, Lacomba-Arias, Lagos-García, Rodriguez-Fernandez, & Verdejo-Román, 2020), exploring the relationship between neural responses, trust, risk, and financial digitalisation decisions. They utilised functional magnetic resonance imaging (fMRI) to investigate whether brain responses to the safety associated with digitalised and non-digitalised financial transactions can explain differences in adopting digital financial channels. Their study revealed that high and low-frequency users of digital financial services show differences in brain procedure and brain structure. Besides, high-frequency users of digital financial channels demonstrate improved brain activation functions in regions correlated to emotional processing to the trust game.

This section investigated the influence of brain behaviour under financial risk. In the next section, the effect of neuroscience on computational brain systems, with a focus on understanding cognitive operations and behaviour, will be discussed. Still, computational neuroscience findings are vital in understanding how the nervous system processes information as conceivable connotations in financial decision-making.

2.3.2 Brain Computational Systems Influenced by Neuroscience Findings

In recent years, neuroscience has made significant strides in comprehending the intricacies of the brain and how it processes information. Computational intelligence methods are typically designed to mimic specific aspects of intelligence observed in biological systems, such as brain function, species evolution, and the social behaviour of populations. In addition, computational neuroscience aims to understand how the nervous system processes information to provide rise to cognitive operation and behaviour, and models play a crucial role in achieving this goal. For example, deep neural networks (DNNs) have recently dominated several domains of artificial intelligence and excel at forecasting neural responses to novel sensory stimuli with high accuracy (Kietzmann, McClure, & Kriegeskorte, 2017). In this direction, CNNs have shown remarkable results incorporating Monte Carlo dropout. For instance, Dürr et al. (2018) recognised the exceptional performance of deep CNN in classifying image-based phenotypes. However, real-world high-content screening (HCS) experiments revealed that prior knowledge of all potential phenotypes could be impractical. Therefore, they utilised a Monte Carlo dropout to establish uncertainty measures for each phenotype prediction. They conclude that employing the MCD method leads to a significant enhancement in model accuracy. Recently, Tousignant, Lemaître, Precup, Arnold, and Arbel (2019) presented a deep 3D CNN with parallel convolutional pathways for predicting future disability progression in patients with Multiple Sclerosis, leveraging multi-modal brain Magnetic Resonance Images (MRI)data. Their framework demonstrates substantial predictive capabilities. Further, they utilised Monte Carlo (MC) dropout to quantify the uncertainty of the network in its output. Their findings show the effectiveness of Monte Carlo dropout and suggest that clinicians use the associated uncertainty estimates to consider which scans require additional examination.

Furthermore, to leverage the neuroscience information, researchers examined which algorithms could better represent the regions of the brain and hence simulate their functions to the machines in an attempt to tackle different tasks. For example, Lun, Yu, Chen, Wang, and Hou (2020) propose a deep CNN structure for selecting the electrode pairs' raw electroencephalography (EEG) signals over the motor cortex region as hybrid samples without preprocessing or artificial feature extraction operations. In their proposed structure, a 5-layer CNN is applied to learn EEG features, a 4-layer max pooling is employed to reduce dimensionality, and a fully connected (FC) layer is utilised for classification. The motor imagination (MI) tasks are divided into four types and are used to test the accuracy of their proposed approach. The results demonstrated that the globally averaged accuracy on group-level classification could achieve 97.28%, the area under the receiver operating characteristic (ROC) curve 0.997, and the electrode pair on ten subjects dataset is FC3-FC4, with the highest accuracy of 98.61%. Their proposed approach supplies a novel idea for facilitating the creation of Brain-computer interface (BCI) methods, accelerating the clinical application procedure. BCIs allow individuals to communicate by decoding neural activity from brain areas, such as the motor cortex and translating it into text in real-time. For instance, recently, Willett, Avansino, Hochberg, Henderson, and Shenoy (2021) developed an intracortical BCI that decodes attempted handwriting movements from neural activity in the motor cortex. They used an RNN decoding approach to train the BCI, and their study participant, who had a spinal cord injury, could type at a swiftness of 90 textures per minute with 94.1% raw accuracy. These typing speeds exceeded those reported for any other BCI and were analogous to typical smartphone typing speeds of people in the participant's age group. In addition, they could decode complete handwritten sentences in real-time, enabling the participant with tetraplegia to communicate by attempting to handwrite their intended message.

Shafiei et al. (2020) introduced an objective method for evaluating mental health in cancer patients using a combination of CNN-LSTM models. They used data recorded by TobiiPro eyeglasses from cancer patients and healthy individuals while viewing artworks in an in-house gallery. Their proposed CNN-LSTM was used to objectify evaluation and categorise the mental health metrics of individuals. Results show that the model had a classification accuracy of 93.81%, 94.76%, and 95.00% for hope, anxiety, and mental well-being metrics, respectively. The researchers propose that their proposed model could be employed for home-based mental health monitoring for patients after oncologic surgery. Future work would include further verification of their model and examining the consequences of specific surgery types. Figure 2.4 displays the mental health evaluation using a CNN-LSTM model.

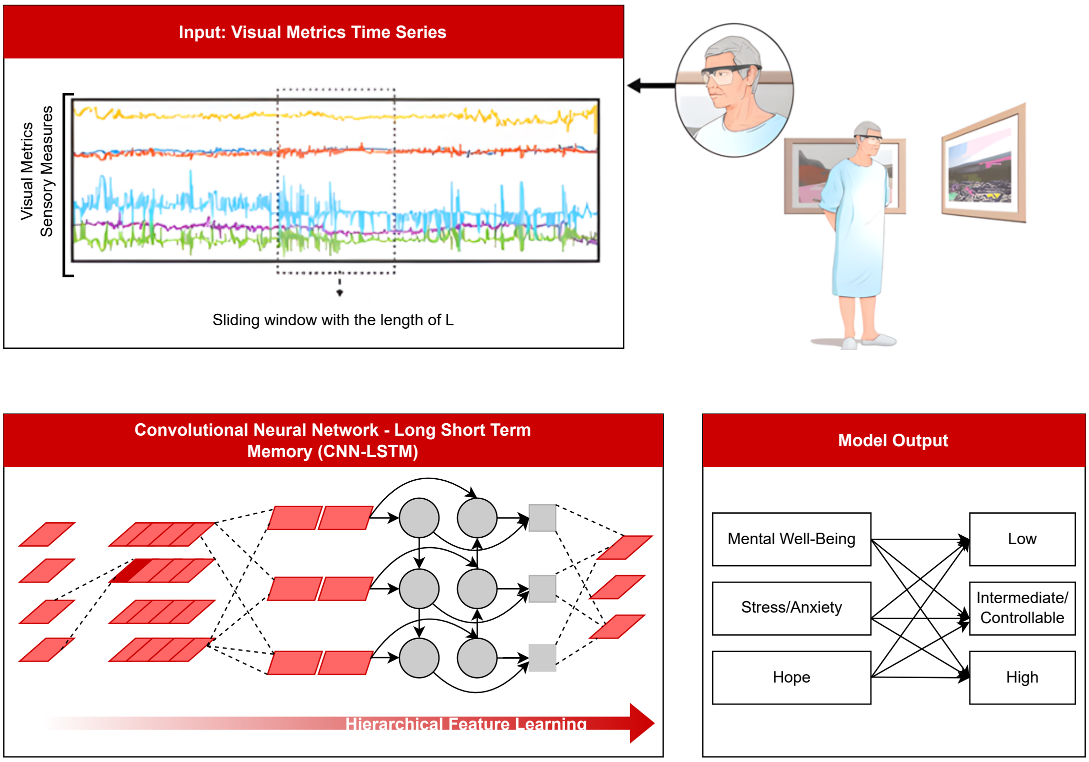


Figure 2.4: Mental health evaluation using a CNN-LSTM model [Shafiei et al. (2020)].

Alamia, Gauducheau, Paisios, and VanRullen (2020) examined which of the neural network architectures, namely feedforward and RNNs, is more appropriate to match human behaviour in artificial grammar learning as an essential characteristic of language accession. The outcomes of their study revealed that both methods could learn grammar after the sequential training such as humans accomplish. However, RNNs perform closer to humans than feedforward ones, independently of the grammar complexity level. Hence, they suggested that recurrent models are tighter to the human cognitive knowledge underlying language. Additionally, they supported the hypothesis that recurrent architectures best model explicit understanding. However, they also noted that further studies are needed to investigate ecological grammar to corroborate the conclusion further.

Y. Wang and Sun (2021) investigated the formation of an innate RNN in the native networks of the mammalian brain. The researchers uncovered that the unidirectional connections form an RNN from three basic units: input units arriving from emotion regions, a hidden unit in the medial prefrontal cortex (mPFC), and output units encountered at the somatic motor cortex (sMO). Furthermore, they provided evidence that the neurons from the basal lateral amygdala (BLA) and the insular cortex (IC) project to the mPFC motor-cortex-projecting (MP) neurons, which form a local self-feedback loop and target central projecting neurons of the sMO. Their study also notes that the innate RNN may lack long-term memory-storage capability due to the vanishing issue and may not be involved in complex sequential movements. Finally, their study concludes that further research is needed to explore the local connectivity components linking pyramidal tract-type corticostriatal (PT-CStr) neurons implicated in motivated behaviour with different types of interneurons to comprehend the innate RNN composition fully. Figure 2.5 illustrates the neurons in the sentiment regions, which activate the MP neurons in the mPFC.

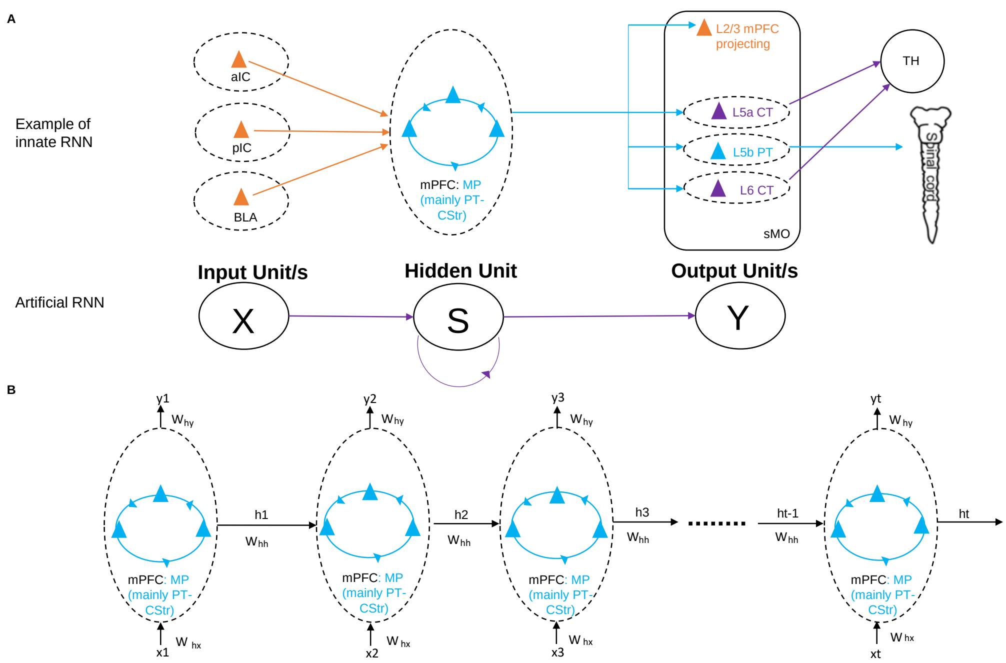


Figure 2.5: (A) The neurons in the sentiment regions (BLA, aIC, and pIC), which trigger (orange arrow) the MP (mainly PT-CStr) neurons in the mPFC. Excited MP neurons exaggerate the signal by triggering themselves. They further communicate the signal to the downstream sMO neurons. In the sMO, excited L5a CT neurons and L6 CT neurons also link (purple arrow) the TH (thalamus); excited L5b PT-CSpi neurons innervate (light blue arrow) the spinal cord. Each dashed circle symbolises one unit. Bottom: a simple model of RNN. (B) An example of an innate RNN unfolded by timeline. x, input; y, output; w, weight; h, hidden. [Y. Wang and Sun (2021)].

By understanding the neural processes in the brain that underlie decision-making, researchers can potentially design computational models that more accurately replicate human behaviour, as presented above. This experience can be especially useful in finance, where accurately predicting market movements and investor behaviour can be challenging. Furthermore, this knowledge can inform and influence the development of ANNs to model Forex time series data. However, there are limitations to conventional ANNs when it comes to modelling time series data. One such area for improvement is the inability to effectively handle long-term dependencies or patterns over an extensive time frame. Recent neuroscience findings have shed light on how the brain can handle long-term dependencies to inform the development of ANNs for modelling time series data. One such approach is the benefit of recurrent neural networks designed to process sequential data. However, due to the highly complex economic environments such as the Forex market, monolithic neural network architectures such as single CNNs and RNNs can be challenging to modify or adapt once trained since any network changes can affect the entire system. Moreover, even a small change to the data can affect the entire monolithic architecture. Thus, neural architectures inspired by the brain could suggest better performance capabilities than monolithic neural networks. For instance, modularity in the brain is responsible for optimising the information capability in the brain's neural paths. Therefore, the brain's modularity allows humans and animals to gain further knowledge without forgetting earlier obtained knowledge. The following section will examine the application of modular neural networks and their usefulness against monolithic architectures from computational neuroscience and financial views.

2.3.3 Brain Modularity

The human brain network is modular, dividing cognitive tasks such as visual perception or language processing into sub-tasks, aiming to perform them easier (Bertolero, Yeo, Bassett, & D'Esposito, 2018; Sporns & Betzel, 2016). Kelkar and Medaglia (2018) revealed that brain modularity is crucial to the human brain's functioning and evolution. The human brain's modular architecture balances specialised and integrated processing, providing functional robustness and adaptability. Short-range connections facilitate local processing, while a few long-range connections are maintained to enhance interregional communication and information transfer. Long-distance links are also disproportionately obeyed along the brain's lengthiest axis, enabling transmission from the anterior frontal cortex to the visual cortex. Hence, evolutionary selection has favoured economic, modular brain networks that supply vigorous, flexible, and adaptive processing throughout the lifespan and over evolutionary time. As a result, the human brain's modularity is a pervasive, dynamic, and essential phenomenon for cognitive functions and overall survival. Recent studies suggested that brain network modularity alterations in cognitive performance prosecuting working memory and reasoning processes are essential in understanding individual divergences in learning (Baniqued, Gallen, Kranz, Kramer, & D'Esposito, 2019; Chaddock-Heyman et al., 2021). Cognitive abilities are indicated in the vmPFC and the AIC, including emotions and the ability to plan under risk process(Adolphs, 1999; (Bud) Craig, 2009; Christopoulos, Tobler, Bossaerts, Dolan, & Schultz, 2009; Falk et al., 2018; Mohr, Biele, & Heekeren, 2010; Nieuwenhuis & Takashima, 2011); observing high modular variability in the insular regions. The existence of von Economo neurons defines the insula (VENs) (Butti, Santos, Uppal, & Hof, 2013), which are accelerated communicating circuits within the brain (X. Chen et al., 2021). Moreover, it is likely that the VENs relay information related to decision-making or awareness (Allman et al., 2010; Cauda et al., 2013). Researchers suggested that the cortical brain regions vary fundamentally in their position, having a specific contribution to economic choices, which is mainly determined by the inputs of each region, (Yoo & Hayden, 2018). The modular approach to operating neuroanatomy of economic decision-making confirms the actions of economic choices, such as comparing values, in the regional architecture of the brain (Padoa-Schioppa, 2011; Rangel & Hare, 2010).

Findings from neuroscience have shown that brain modularity is crucial in learning and decision-making. Therefore, this study proposes a novel MCoRNNMCD-ANN model, inspired by recent neuroscience discoveries, which enables accurate and robust predictions of challenging price movements in the Forex market. Sharwood Smith (2019) presents a comparison of two frameworks for understanding the relationship between cognition and the brain: the neuroemergentist framework proposed by Hernandez et al. (2019) and the Modular Cognition Framework (MCF), also known as Modular Online Growth and Use of Language (Truscott & Smith, 2004). The comparison aims to demonstrate that the evidence for emergentist and dynamic traits in cognitive development and processing can also be understood as the result of a modular mind with a stable set of independent systems that are in place at birth. The MCF approach highlights the collaboration between modular systems and how they adapt to solve a constant flow of tasks. The MCF perspective on bilingual development and behaviour is also discussed, specifically how more than one language can coexist in one mind and the cognitive benefits of bilingualism. The commentary concludes by suggesting that emphasising the contrasts between these frameworks may obscure possible similarities and that further research is needed to integrate these perspectives better to understand the mind's and brain's functioning.

Fermin et al. (2022) revealed that the insula, a region of the brain that receives a vast amount of interoceptive information from visceral organs, has specialised modules that form parallel networks with the prefrontal cortex and striatum to form higher-order interoceptive representations. These representations are compelled and context-dependent to support habitual, model-based and exploratory control of visceral organs and physiological processes. Their Insula Hierarchical Modular Adaptive Interoception Control (IMAC) model suggests that the insula's parallel connections with PFC sub-regions, the striatum, and neuromodulatory input from the dopaminergic system play a role in the emergence of unconscious emotions, conscious feelings and the rise of visceral dysfunctions observed in depression.

The study of parallel feature extraction is discussed further in the next section, as the structure of the proposed MCoRNNMCD-ANN has been influenced and inspired by the recent neuroscience findings, as also referred to in section 1.2 of this thesis.

2.3.4 Parallel Feature Extraction and Computational Representation of Brain Modularity

Parallel feature extraction procedure has been utilised in many scientific fields, such as neuroscience and natural language processing. This procedure is a critical phase in the machine learning process, as it involves selecting the most informative and relevant features from the data, which can improve the performance of a machine learning model. Additionally, parallel feature extraction can enable more sophisticated and computationally intensive feature extraction techniques, leading to more accurate predictions (Farsi, Amayri, Bouguila, & Eicker, 2021).

One way to achieve parallel feature extraction is by using a modular neural network, which can be represented as a multi-head model. A multi-head model consists of multiple parallel branches, each responsible for extracting a specific feature from the input data. This architecture allows the model to extract multiple features simultaneously rather than sequentially. Additionally, each branch can be optimised independently, receiving inputs and improving the model's overall performance. Various computational models, such as CNNs, have been utilised to model parallel feature techniques. For example, Dai and Jonnagaddala (2018) examined the effectiveness of CNNs in forecasting psychiatric conditions. Their study saw that CNNs could provide accurate results in classifying the symptom severity, even without the need for advanced feature engineering or feature selection. They have also explored the impact of parallel CNN architectures with different filter region sizes, improving performance. More specifically, the error of CNN was decreased by 0.004 and, with an MAE of 0.644, surpassed that of NBM and C4.5 with an MAE of 0.660 and 0.716, respectively. Finally, their analysis emphasised the challenges of parameter optimisation and obtaining high-quality embeddings as critical factors that affect the performance of CNNs. Future research will address these challenges and investigate better representation technologies to capture the syntax and semantics of the content.

CNNs can form a computational method to model cortical representation and organisation for spatial visual processing. However, Shi et al. (2018) investigated how to overcome the limitation of CNNs to understand better the brain processes temporal information in vision. They found that CNN can be extended with RNN, including recurrent connections that allow spatial representations to be remembered and accumulated over time. The RNN also agreeably predicted cortical responses to natural movie stimuli than the traditional CNNs, particularly in areas of the visual cortex responsible for processing temporal information. Additionally, the RNN ware able to improve performance in action recognition while maintaining its ability in image recognition, making it a valuable model for developing brain-inspired AI systems that can continuously learn and expand capabilities. Finally, the researchers suggested that this approach could help understand and model other perceptual or cognitive systems beyond vision, such as natural language processing, memory, and planning. Figure 2.6 illustrates the architectural structure of the RNN and CNN layers. It is worth noting that the findings from this analysis regarding the CNN weakness to understand better the brain processes temporal information inspired the proposed MCoRNNMCD-ANN to integrate the RNN model into the CNN. However, in the MCoRNNMCD-ANN, the pooling layers have replaced the RNN, as discussed in Chapter 4 and by using natural language processing as the researcher suggested.

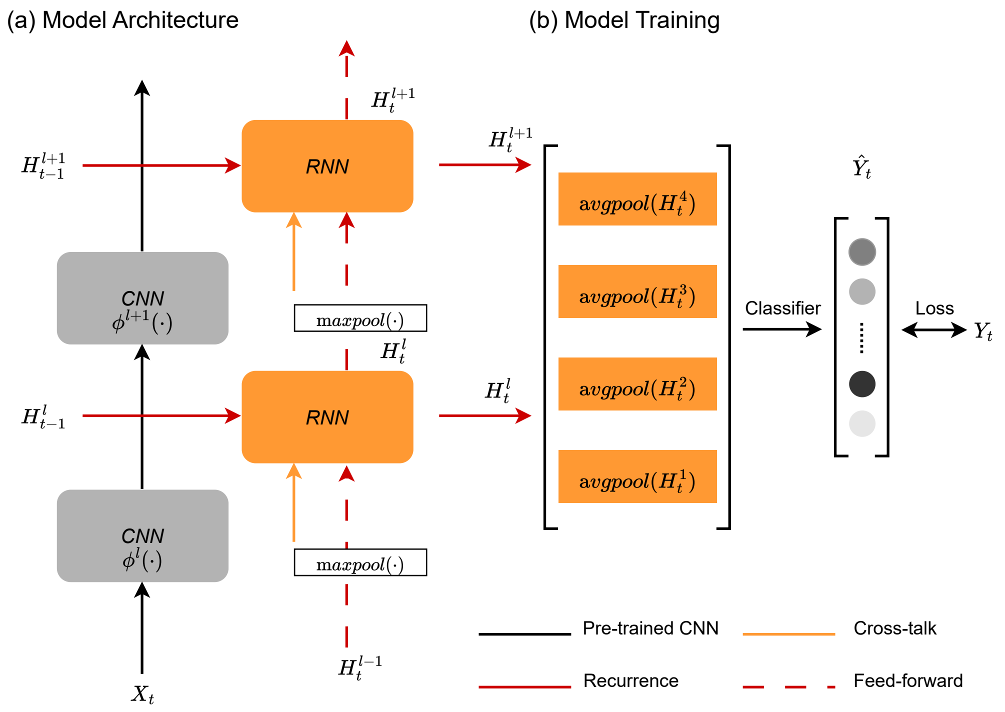


Figure 2.6: (a) The architectural structure of the RNN. (b) The model training process. The grey blocks show the CNN layers; the orange blocks show the RNN layers [Shi et al. (2018)].

Yao, Zhang, Zhou, and Liu (2019) developed a deep learning model for image classification that combines two types of neural networks. More specifically, their model uses a parallel system that combines a CNN and an RNN for image feature extraction and a unique perceptron attention mechanism to unite the features from both networks. Their proposed model also utilised a switchable normalisation method and targeted dropout regularisation technique to improve its performance and robustness. Their findings have shown that their suggested method outperforms current state-of-the-art methods based on CNN, demonstrating the benefits of using a parallel structure.

Neuroscientists have also examined computational brain modularity to explain brain functionalities. For example, Tzilivaki et al. (2019) investigated that complex, non-linear dendritic computations necessitate the development of a new theory of interneuron arithmetic. Using thorough biophysical models, they foresaw that the dendrites of FS basket cells in both the hippocampus and the prefrontal cortex are supralinear and sublinear. Furthermore, they compared a Linear ANN, in which the input from all dendrites is linearly merged at the cell body, and a two-layer modular ANN, in which the input is fed into two parallel, separated hidden layers. Despite that, the linear ANN exhibited relatively good performance; the two-layer modular ANN surpassed the respective linear ANN, which failed to illustrate the variance assembled by discrepancies in the input area. Hence, they recommended that a two-layer ANN that deems both dendritic nonlinearities is considerably better for FS basket cells than the currently considered linear point neuron. This two-layer modular ANN inspired this thesis's proposed MCoRNNMCD-ANN design. Figure 2.7 illustrates a linear ANN and the two-layer modular ANN.

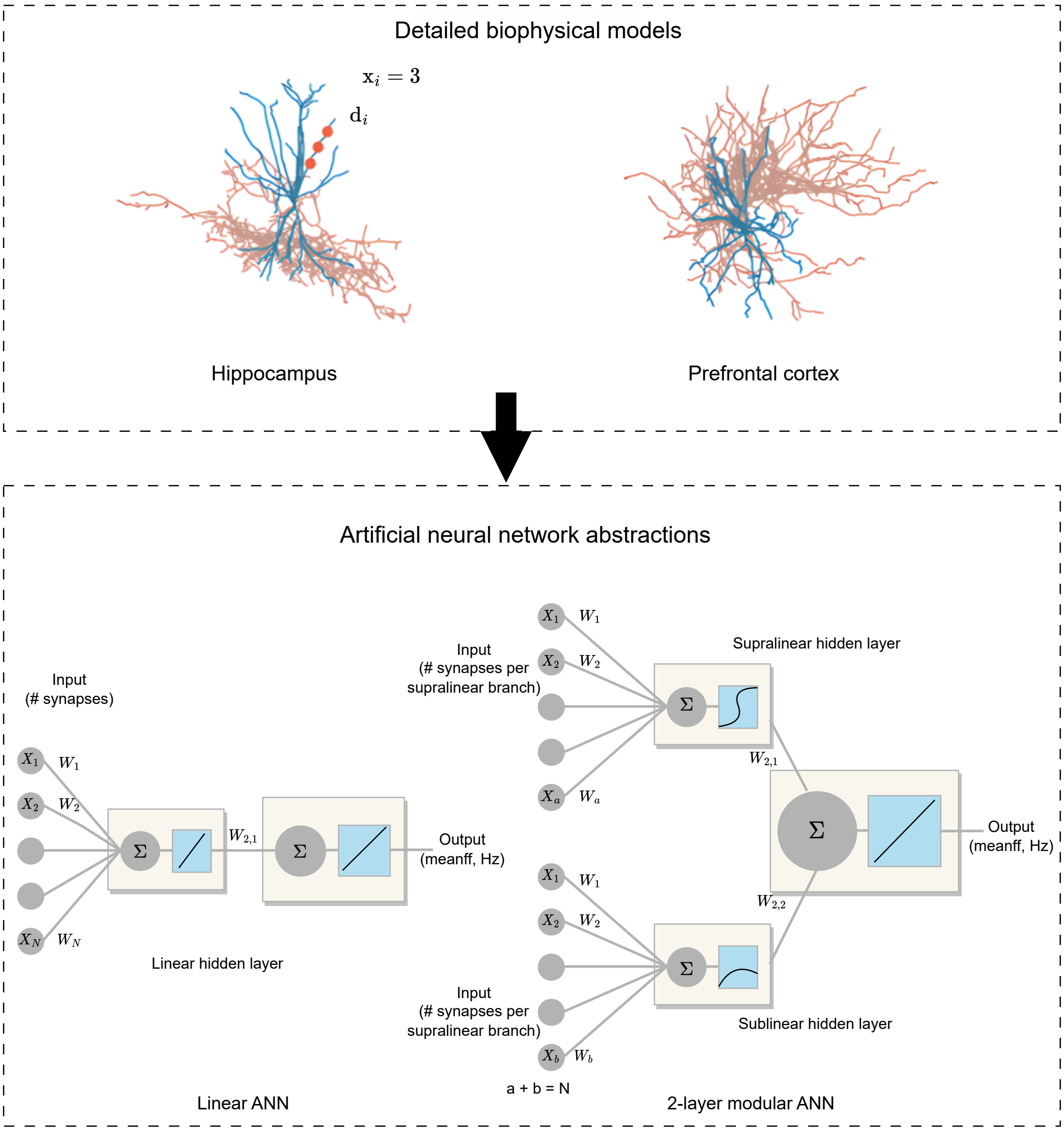


Figure 2.7: A linear ANN and a two-layer modular ANN, in which the input is provided into two parallel, divided hidden layers [Tzilivaki et al. (2019)].

More recently, Flesch, Juechems, Dumbalska, Saxe, and Summerfield (2022) uncovered that the "rich" learning approach, which structures the hidden units to prioritise relevant features over irrelevant ones, results in neural coding patterns consistent with how the human brain processes information. Additionally, they found that these patterns evolve as the task progresses. For example, when they trained deep CNNs on the task using the "rich" learning method, they discovered that it induced structured representations that progressively transformed inputs from a grid-like structure to an orthogonal structure and eventually to a parallel structure. These non-linear, orthogonal and parallel representations demonstrated a vital element of their research, as they suggest that the neural networks can code for multiple, potentially contradicting tasks effectively.

In financial markets, Gradojevic, Gencay, and Kukolj (2009) investigated whether a non-parametric modular neural network (MNN) can value the S&P 500 European call options. The MNN method showed superior out-of-sample pricing performance compared to an array of parametric and non-parametric models, observing that the average mean squared prediction errors (MSPE) can be reduced between 12% and 70%, in comparison to the basic MNN model. In conclusion, they revealed that modularity improves the generalisation properties of standard feed-forward neural networks, suggesting that the performance of MNNs can be further investigated in highly volatile markets. Baek and Kim (2018) proposed a framework, ModAugNet, which consists of two modules, integrating a novel data augmentation technique for stock market index forecasting. Their model consists of a prediction LSTM module and an overfitting prevention LSTM module. Two representative stock market data, the S&P500 and the Korea Composite Stock Price Index 200 (KOSPI200), are used to assess the proposed model's performance. Their findings demonstrate that the proposed ModAugNet-c had a lower test error than a monolithic deep neural network, an RNN and SingleNet, a comparable model without an overfitting prevention LSTM module. For example, compared to SingleNet's corresponding S&P 500 forecasting errors, the test MSE, MAPE and MAE decreased to 54.1%, 35.5%, and 32.7%, respectively. Similarly, compared to SingleNet's corresponding KOSPI200, the forecasting errors decreased to 48%, 23.9%, and 32.7%, respectively. However, a drawback of their study was that they did not view other types of information, such as news, investors' sentiment, and macroeconomic characteristics. Figure 2.8 displays the ModAugNet model.

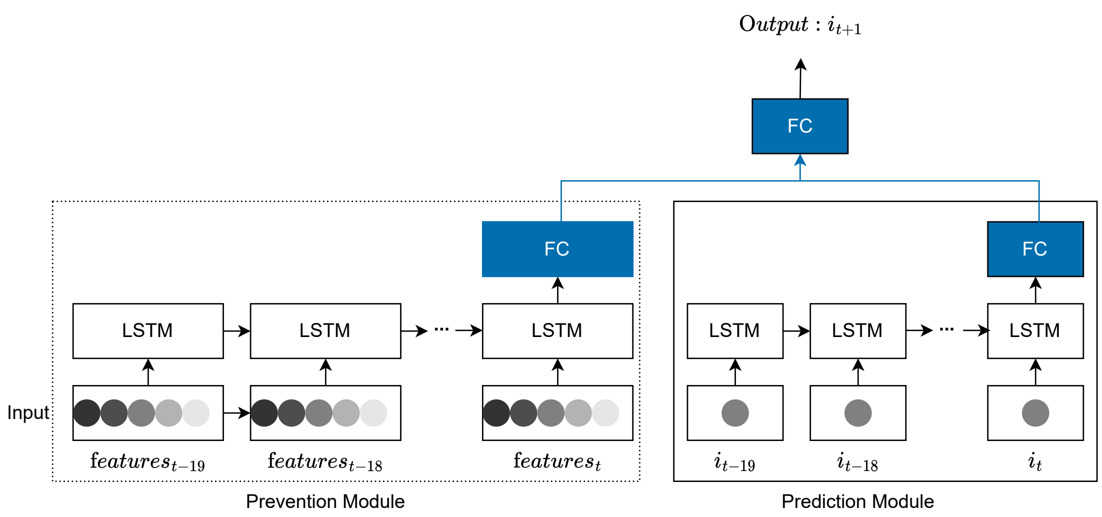


Figure 2.8: ModAugNet model [Baek and Kim (2018)].

Likewise, Lee and Kim (2020) proposed a new stock market forecasting framework called NuNet to improve the accuracy of predictions of S&P500, KOSPI200, and FTSE100 prices. NuNet is an end-to-end integrated neural network comprising two feature extractor modules: a super-high dimensional market information feature extractor and a target index feature extractor. The market feature extractor uses a combination of CNN and ConvLSTM layers to extract features. The target index feature extractor uses two layers of stacked LSTM to learn the temporal patterns of the target index. The outputs of each module are then concatenated for feature fusion and connected to fully connected layers to predict the closing price of the target index. Additionally, a mini-batch technique called trend sampling is suggested to sample more current data during training. Results show that NuNet outperforms all baseline models, including the SingleNet (Baek & Kim, 2018) and the SMA in the three indexes.

Other studies have also investigated the success of applying modular architectures in different domains. For example, Ma, Zhang, Du, Ding, and Sun (2019) proposed a parallel architecture to predict metro ridership in large-scale networks. Their model consists of two modules, a CNN and a bi-directional long short-term memory network (BLSTM), responsible for extracting spatial and temporal features, respectively. The two networks are concatenated and fed into a fully connected network for metro ridership prediction. Their proposed model was tested on the Beijing metro network and outperformed traditional statistical models, deep learning architectures and sequential structures such as ANN, RNN, CNN–BLSTM, SAR, OLS, ARIMA, DMVST-Net, and sequential CNN–BLSTM. As a result, Metro authorities can use their model to effectively allocate resources and improve service in overcrowded areas and daily operational management.

Duthie and Budzynska (2018) presented an approach for analysing the ethos (character of the speaker) in natural language text, particularly in UK parliamentary debates. They design a deep modular recurrent neural network (DMRNN) utilising proven methods from argument mining and sentiment analysis to develop an ethos mining pipeline. The pipeline can extract information about ethos reliably and robustly, with a macro-F1 score of 0.83 for ethotic statement extraction and 0.84 for polarity annotation. The researchers also apply this pipeline to an analysis of the political landscape, called ethos analytics, to find evidence of the strong impact of ethos in comprehending the dynamics of governments. Future work contains further enriching resources and technology for ethos mining and analytics. In textual dialogues, Chatterjee et al. (2019) suggested a novel deep learning-based approach to detect emotions such as happiness, sadness, and anger. Their system combines both semantic and sentiment-based representations for more accurate emotion detection. The evaluation of their approach revealed that it surpasses conventional machine learning techniques such as Support Vector Machines (SVM), Decision Trees, Naive Bayes and various deep learning models such as CNN and RNN. More specifically, they proposed a deep learning-based model that uses two parallel LSTM layers to learn semantic and sentiment feature representations and encode sequential patterns in the user utterance. These two feature replicas are then concatenated and passed to a fully connected network, which models relations between these features and outputs likelihoods per emotion class. Their approach could enhance the performance of real-world chatbots and other textual-dialogue-based applications.

This section explored brain modularity's applicability in neuroscience and the financial markets. An effort is made to emphasise the importance of modularity in designing and learning in ANNs. Moreover, using a multi-head model in the Forex market can be particularly beneficial as it allows the model to extract multiple features simultaneously, such as economic indicators and sentiment analysis. As a result, these features can be combined to predict future market movements accurately. In the next section, other techniques, such as transfer learning, can improve learning from the knowledge that computational models yielded from the financial predictions domain by transferring information in a related domain. Furthermore, this information transfer could enhance the performance of simple forms of ANNs that train with fewer data in the related field of financial predictions, such as Forex.

2.4 Transfer Learning

Transfer learning refers to the capability of a system, be it a biological brain or a computational model, to leverage knowledge gained from one domain to improve performance in a related field. For example, the brain can extract underlying patterns, principles, or representations from one task and apply them to another, enhancing learning performance.

In computational models, transfer learning is also used to improve the generalisation performance of another model on relevant tasks with fewer data. By leveraging knowledge learned from the pre-training phase, the model can form with a better initial understanding of the problem, requiring less training data to execute competently. Transfer learning in computational models often involves transferring learned weights, representations, or feature extractors from one model to another. In the following two sections, 2.4.1 and 2.4.2, the application of the transfer learning method from a biological and computational matter is presented.

2.4.1 Transfer Learning Process in the Brain

Researchers showed that the hippocampus and vmPFC support the appearance of conceptual knowledge and its impact on choice behaviour, utilising this knowledge to solve complex decision problems. Moreover, they revealed that the ability to transfer prior knowledge into novel tasks is a shaping characteristic of human intelligence (Kumaran, Summerfield, Hassabis, & Maguire, 2009). More recently, studies showed that across rodents and humans during new learning, prior knowledge enhances cortical activation and cortico–cortical functional connectivity (Bein, Trzewik, & Maril, 2019). Furthermore, Bein, Reggev, and Maril (2020) proposed that when the information is supported by prior knowledge, learning novel associations leads to asymmetric cortical effects, a concept related to the human cognitive function (Kong et al., 2018). This concept is a consequence of the asymmetry we likely have in language and many other higher cognitive specialisations. The mnemonic processes in our brains build long-term knowledge and, more specifically, how different phases of memory formation, such as encoding, consolidation, retrieval, and reconsolidation (Van Kesteren & Meeter, 2020). Consolidation of memories is the advanced transfer of memories into a state where they stay stable over a long-lasting time interval (Luboeinski & Tetzlaff, 2021).

Utilising the transfer learning technique to a computational model could simplify the efforts to develop a model from scratch by applying the algorithms within one domain to another. Furthermore, a model trained on a task with much data, applied on another task with less data available, has recently led to significant advancements in machine learning and fields such as natural language processing. In the below section, the computational approaches of transfer learning explore the potential enrichment in forecasting the performance of computational models in cognitive processes and financial markets.

2.4.2 Computational Transfer Learning Approaches

Transfer learning (TL) is an approach that allows the use of data from unrelated or partially related tasks to improve the performance of a model in a new task. For example, in Forex and AI, TL can be used to forecast future currency trends by leveraging prior knowledge from historical currency data.

Predicting currency trends is challenging, as it involves analysing vast historical data and identifying patterns that can indicate future trends. In addition, traditional machine learning algorithms, such as linear regression, need help to capture the complexity of currency markets. Deep learning has been presented to be a decisive tool for analysing time-series data. Computational models, such as CNNs and RNNs, can explore large amounts of historical data and identify patterns that indicate future trends. These models are also able to adapt to changes in the market and can improve their predictions over time. However, even with deep learning, predicting currency trends still needs to be improved. One of the main problems of deep learning in this context is the need for extensive data portions, which is expensive and time-consuming. By leveraging prior knowledge from historical currency and sentiment data, TL can improve the performance of deep learning models and reduce the need for supplemental data.

In TL, a model trained on a source task can transfer its knowledge to perform a target task with fewer data. This process can be designed by adjusting the model's parameters to suit the target task's characteristics better. In conclusion, TL, in combination with deep learning, can be a robust toolset for predicting currency trends. TL allows for the reuse of prior knowledge from historical currency data and can improve the performance of deep learning models. Thus, it can lead to more accurate predictions and a reduced need for additional data. Below will be presented as either computational primarily utilised in neuroscience and finance to understand better how TL can enhance generalisation performance.

H. Li, Parikh, and He (2018) aimed to improve the early diagnosis of neurological disorders by using advanced machine learning algorithms and resting-state functional Magnetic Resonance Imaging (fMRI) data with autism spectrum disorder (ASD) classification as the target task. The researchers proposed a deep transfer learning neural network (DTL-NN) framework that utilises previously collected healthy subject data from existing databases to enhance the classification of brain functional connectivity patterns for new disease tasks. The DTL-NN approach was compared to traditional deep neural networks, and support vector machine models demonstrated adequate performance. Furthermore, the DTL-NN approach consistently improved across multiple data and consistently identified discriminatory functional connectivity patterns associated with ASD. The authors finally indicated that their approach could help improve diagnosis for rare neurological disorders where investigating large cohorts is challenging.

M. Chen et al. (2020) proposed a method for the early identification of cognitive deficits in very preterm infants by analysing brain structural connectome data from diffusion tensor imaging (DTI) scans taken at term-equivalent age. The researchers utilised deep CNN and the transfer learning technique to achieve this. Their analysis included 110 infants and found that the proposed transfer learning enhanced the convolutional neural network (TL-CNN) model. Moreover, their TL-CNN performed better than other models, such as LR, SVM, DNN, CNN and TL-DNN, predicting cognitive assessment scores for both cognitive deficit classification and continuous cognitive score prediction tasks. The consequences of their study suggested that deep learning models may be an effective method to predict later neurodevelopmental effects in very preterm infants at term-equivalent age and help determine the brain regions that are most indicative of cognitive deficit. Figure 2.9 shows the transfer learning CNN (TL-CNN) model.

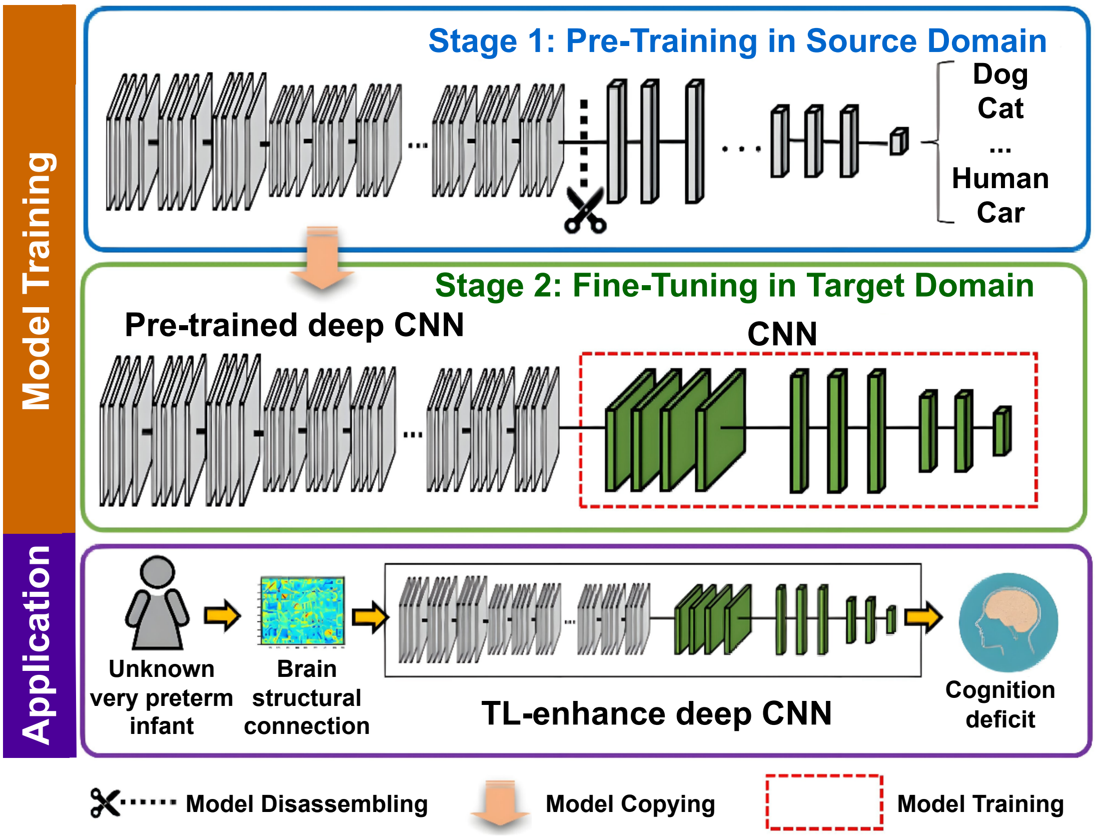


Figure 2.9: transfer learning CNN (TL-CNN) model to forecast cognitive deficiencies utilising brain structural connectome data acquired at a period in very preterm infants [M. Chen et al. (2020)].

Kraus and Feuerriegel (2017) decision examined deep neural networks for financial decision support, specifically regarding stock market predictions based on company disclosures. The researchers found that deep learning methods, such as RNNs, are better equipped to handle the complexity and ambiguity of natural language in financial disclosures. Moreover, RNNs performed better than traditional machine learning approaches such as Ridge regression, Lasso, Elastic net, Random forest, SVM, AdaBoost and Gradient boosting. Their study also explores transfer learning, showing the significance where the network is pre-trained on a different corpus, which results in higher directional accuracy in predicting stock price movements in response to financial disclosures. The study's results highlight the potential of deep learning for financial decision support and suggest that it could be a valuable tool for investors and automated traders. However, the study also notes that the configuration of deep neural networks is challenging and requires extensive parameter tuning for promising results.

Merello, Ratto, Oneto, and Cambria (2019) compared feed-forward neural networks (FFNNs) by applying regression and classification approaches to forecast the prices for ten companies from the Nasdaq Index using stock and textual data. The data was obtained from Yahoo Finance, Nasdaq News and Google Finance, with the time duration between two posterior samples to be selected hourly. In order to manage the issue of data scarcity, they used transfer learning. They concluded that regression access derived better results against classification approaches, and the employment of transfer learning proved effective in heightening the prediction performance.

To address the data scarcity and to choose the optimal features' selection as inputs to anticipate the KOSPI 200 and the S&P 500 prices for five companies, Nguyen and Yoon (2019) used an LSTM based on transfer learning. Historical closing price data retrieved from the Yahoo financial website for six years. This study indicated that deep transfer learning outperformed the baselines of SVM, RF, and KNN models. For future research, Nguyen proposed implementing numerical and sentiment data to enhance the predictability performance of the models.

Cen and Wang (2019) investigated the usefulness of the LSTM model in predicting crude oil price fluctuations. They proposed using a transfer learning technique to improve the conventional application field of LSTM, such as NLP, where a large portion of data is a consensus training accuracy of LSTM. Consequently, transfer learning will increase the accuracy of oil market price prediction. Their study evaluated the predictive ability of their proposed LSTM model on the West Texas Intermediate and Brent crude oil prices by decomposing the time series into intrinsic mode functions using the ensemble empirical mode decomposition method. Results show that their proposed LSTM model can detect the main fluctuation characteristics of crude oil prices for different fluctuation frequency levels. Furthermore, transfer learning enhances performance in terms of forecast accuracy.

This section presented that the acquired knowledge from the computational models is transferred to a related domain, such as in financial predictions, by enhancing their generalisation performance. As already discussed in Chapter 1 and the above sections of Chapter 2, decision-making is affected by the emotions of the investors, and not only the rationality of the decisions plays the final role in their choices. Therefore, the following section will discuss natural language processing techniques considering simulating investors' emotions by presenting many aspects of the sentiment analysis incorporated into the financial market's anticipation.

2.5 Natural Language Processing in Financial Markets

NLP is a field of study that concentrates on the relations between computers and human languages. NLP aims to allow computers to comprehend, interpret, and generate human language. The connectivity between NLP and the brain refers to the effort to understand the workings of the human brain when processing language and to use this understanding to improve NLP algorithms. Research in this area seeks to uncover the underlying computational principles of the brain's language processing capabilities and to replicate them in NLP models. For example, Caucheteux and King (2022) compared deep language models to identify the computational principles behind their ability to generate brain-like representations of sentences. It analysed the brain responses to 400 sentences in 102 subjects using functional magnetic resonance imaging (fMRI) and magnetoencephalography (MEG) and tested how the algorithms mapped onto the brain responses. The outcomes demonstrated that the parallel between the algorithms and the brain primarily relies on their ability to predict words from context. This likeness shows the rise and maintenance of perceptual, lexical, and compositional replicas within each cortical region. Their study concluded that modern language algorithms partially converge towards brain-like solutions, delivering a favourable path for comprehending the foundations of natural language processing. NLP also has a wide range of applications, and one area where it is seeing increasing use is finance and Forex.

NLP has yet to be considerably applied to Forex prediction compared to other approaches such as linear regression, neural networks, and time-series analysis. Regardless, NLP is typically utilised to analyse and process textual data. Since Forex prediction mainly involves analysing numerical time-series data, there may be different choices for this domain (Islam, Hossain, Rahman, Hossain, & Andersson, 2020; Sezer, Gudelek, & Ozbayoglu, 2020). Furthermore, NLP can analyse enormous quantities of unstructured data, such as news articles. This analysis can help professionals extract relevant information and insights that aid decision-making. For example, an NLP-based system might be utilised to analyse news articles about a particular company to determine its sentiment and prospects. A system like this can be especially useful in an industry where information constantly changes. Traditional analysis methods, such as financial statements and performance metrics, may only tell part of the story. In Forex, NLP can also analyse news from social media posts related to currency markets to identify trends and potentially make trading decisions. Given the increasing interconnectedness of global financial markets, it is becoming increasingly important for traders to fast and accurately analyse a wide range of information to make informed decisions. NLP can help facilitate this process by providing a more comprehensive view of relevant information and trends. Recent trends in NLP have shown the success of large language models (LLMs) in various NLP tasks. However, this success is not a guarantee against simpler models.

The performance of a model depends on various factors, such as the task, the size and quality of the dataset, and the specific architecture of the model. For instance, Ezen-Can (2020) investigated implementing a pre-trained BERT model and a simple bidirectional LSTM model for intent classification using a small dataset collected for building chatbots. The author wanted to answer whether using a large language pre-trained model like BERT would perform better than simple models, especially when the dataset is small. The outcomes revealed that the simple bidirectional LSTM model outperformed the pre-trained BERT model in accuracy. The LSTM model also had faster training times compared to the BERT model. The author concluded that the model chosen should be based on the specific task and dataset rather than just choosing the most popular model. The results also indicated that BERT overfitted more than the simple LSTM model for smaller datasets.

NLP techniques are utilised to automatically label input text data based on predefined rules to identify the polarity of user sentiment called rule-based systems. These systems can provide a sentiment score that can express the emotional state of the individuals. Well-known rule-based systems are presented below.

Sentiment Lexicons

Rule-based lexicons are dictionaries that contain rules for identifying specific words or phrases in the text. These lexicons are often utilised for parts-of-speech tagging, named entity recognition, and sentiment analysis. The rules in a rule-based lexicon may be established on regular expressions, word patterns, or other criteria.

Well-known lexicons that are used for sentiment analysis on Twitter:

- SentiWordNet (Baccianella, Esuli, Sebastiani, et al., 2010): SentiWordNet is a lexical resource that assigns a positive or negative sentiment score to each synset (a set of synonyms) in WordNet. It can be employed to determine the sentiment of a given text by summing the sentiment scores of the words in the text.

- AFINN (Nielsen, 2011): AFINN is a lexicon of a list of English words rated for valence with an integer between -5 and 5.
- VADER (Valence Aware Dictionary and sEntiment Reasoner) (Hutto & Gilbert, 2014): VADER is a rule-based sentiment analysis lexicon attuned explicitly to social media sentiments.
- TextBlob (Loria et al., 2018): TextBlob is a Python library providing a simple API for text data. It includes several features for natural language processing, including part-of-speech tagging, noun phrase extraction, sentiment analysis, and more.

NLP techniques such as sentiment analysis in finance and Forex will persist in expanding in the forthcoming years as more and more professionals recognise its value. Furthermore, as NLP technology improves, it will become an increasingly important tool for professionals looking to make knowledgeable decisions in an increasingly complicated and fast-paced world. However, it is essential to note that NLP is not a magic bullet and has to be employed with other breakdown methods, such as ANNs. While NLP can provide valuable insights, it is ultimately up to the user to interpret and act on this information in a way that is appropriate for their specific investment goals and circumstances. The sentiment analysis application in financial markets such as in Forex is presented in the following section.

2.5.1 Sentiment Analysis in Financial Markets

Sentiment analysis in finance uses natural language processing and computational linguistics techniques to identify, extract, and quantify the sentiment expressed in financial texts. Sentiment analysis includes diagnosing financial news articles, earnings call transcripts, social media posts, and other types of financial communication to understand the sentiment of market participants and predict market movements. There are many potential applications for sentiment analysis in finance, including stock market prediction, risk management, and investment decision-making. By understanding the sentiments of market participants, investors and financial institutions can make more knowledgeable decisions and improve their financial performance. Overall, sentiment analysis in finance is a rapidly growing field with significant potential to improve our understanding of financial markets and make better investment decisions. Combining NLP and neural networks can be a potent tool for Forex rate prediction because it allows for integrating a broad scope of data sources.

Sohangir, Petty, and Wang (2018) investigated whether lexicon-based techniques can enhance the performance of sentiment extraction from social media data in financial markets. They used three lexicon-based methods, namely, VADER, SentiWordNet, and TextBlob, to analyse data from StockTwits tweets and compared the results to machine learning methods such as logistic regression, linear SVM, and Naive Bayes for classification. Their study revealed that the lexicon-based approaches outperformed machine learning methods regarding accuracy and that VADER was the best lexicon method for predicting StockTwits users' sentiment. Consequently, utilising lexicon-based methods like VADER can improve the accuracy of sentiment analysis and is faster than the forenamed machine learning techniques.

Other researchers (Seifollahi & Shajari, 2019) proposed an embedded word sense disambiguation (WSD) method for Forex market prediction in an already-designed model that utilises the WordNet and SentiWordNet lexicons. The sentiment analysis applied to the news headlines from marketwatch.com by anticipating the directional movement of the EUR/USD exchange rate. They concluded that detecting many words can hone sentiment analysis. However, one of the challenges of the proposed WSD method was the enormous amount of time needed to execute the algorithm. For future investigation, they proposed to foresee the prices as an alternative to directional movement and the usage of social media platforms.

Souma, Vodenska, and Aoyama (2019) applied deep learning methods by training RNN with LSTM units to forecast the financial news sentiments. They utilised the Thomson Reuters News Archive and the Thomson Reuters Tick History datasets. The global vectors (GloVe) are employed as a word representation method. The polarity, i.e., positive or negative sentiment, has been delineated based on the log return of the ratio between the average price for one minute before and one minute behind the news was acknowledged. They concluded that the model was in place to estimate the positive or negative news. Mohan, Mullapudi, Sammeta, Vijayvergia, and Anastasiu (2019) compared models such as the RNN-LSTM, ARIMA, and Facebook Prophet (FB prophet), aiming to enhance the prediction accuracy of the stock values of the S&P 500. Textual information and stock prices were collected with a web scraper. Their findings indicated that the RNN model outperformed the ARIMA and FB prophet in price anticipations. However, one limitation of their study was that the models could have performed better in events where markets presented high or low volatility.

Sirimevan, Mamalgaha, Jayasekara, Mayuran, and Jayawardena (2019) proposed LSTM-RNN to analyse particular sources and employed lexicons such as TextBlob to extract the sentiment scores. This approach achieved highly accurate predictions for one-day, seven-day, fifteen-day and thirty-day time-frames, providing insight for investors and companies to track their performance in the Dow Jones 30 index. Additionally, beta values for stock market data and macroeconomic variables such as exchange and gold rates will be studied. Likewise, Shahi, Shrestha, Neupane, and Guo (2020) compare the performance of LSTM and GRU models for the Nepal Stock Exchange - Agricultural Development Bank Limited (NEPSE-ADBL) market forecasting, focusing on the impact of comprising financial news sentiments using the VADER lexicon. Their study encountered that both LSTM and GRU are circumstantial in stock anticipation when using only stock market features as input and that integrating sentiment scores improves the performance of both models for stock price predictions. In addition, their study suggests that using LSTM-News and GRU-News models, which require more computation power, yields better results. Finally, as a future direction, they propose a cooperative deep-learning architecture incorporating LSTM-News and GRU-News models that could be used as an expert system to recommend the best forecasting dynamically.

Researchers have also used rule-based methods alongside AI to investigate a crucial aspect of fake reviews. For example, using deep learning approaches, Sadiq et al. (2021) proposed a framework for predicting the authenticity of numeric ratings on the Google Play Store. Their method consists of two stages. In the first phase, the polarity of reviews is estimated using sentiment analysis tools such as TextBlob and VADER to construct ground truth. In the second phase, star ratings are anticipated from the text format of reviews after training deep learning models such as CNN, LSTM, RNN, BiLSTM, and GRU on the ground truth obtained by VADER. The framework results are then compared with the actual ratings of the apps on the Google Play Store to find any mismatch between the user reviews and user ratings. The best-performing model, CNN, showed a classification accuracy of 89% against the LSTM, RNN, BiLSTM, and GRU and robust results of 82%, 89% and 86% in terms of precision, recall and f1 score, respectively.

Kumar, De, and Roy (2020) developed a hybrid Recommendation System (RS) for movies to minimise the impact of limitations such as the need for previous user history and habits to perform the task. Their RF consisted of collaborative filtering (CF) and content-based filtering (CBF) jointly with sentiment analysis of tweets from microblogging sites. The purpose of using movie tweets was to understand current trends, public opinion, and user reactions to movies. Experiments conducted using public databases have yielded promising results. After preprocessing, the text extracted from tweets was used for sentiment analysis by applying the VADER rules-based method. The results showed that VADER performance was better than the other methods, such as Naive Bayes, TextBlob, PTWE, and Attention models, which used bidirectional LSTM and IMDB rating.

Twitter is also one of the primary sources in which text data is acquired to extract investors' sentiments. Therefore, the following section will discuss its effect on financial markets like Forex.

2.5.2 Sentiment Analysis Effect of Twitter in Financial Markets

Twitter is an influential social media platform with a reputation as a source of breaking news and real-time information. In recent years, there have been assumptions about whether or not Twitter could affect stock market decisions. While it is difficult to pinpoint any specific factor as the sole driver of market movements, it is worth examining how Twitter could influence the stock market.

Twitter could affect stock market decisions through the dissemination of information. As a real-time platform that allows users to share news and updates, Twitter can quickly spread information about a company or industry (Duz Tan & Tas, 2021). For example, suppose a company announces positive earnings or a new product launch on Twitter. In that case, this information could attract the attention of investors and drive up the company's stock price. On the other hand, if unfavourable information is shared on Twitter, it could drive down the stock price. In addition, Twitter could influence stock market decisions through sentiment analysis. Analysing the sentiment of tweets about a particular company or industry makes it possible to get a sense of how the market reacts to specific events or developments. Investors could use this analysis to decide whether to buy or sell a particular stock.

Regarding the potential influence of Twitter on the Forex market, the platform could serve as a means of communication and information spread (Bianchi, Gómez-Cram, Kind, & Kung, 2023). For example, if a central bank announces a change in monetary policy on Twitter, this could impact the value of the country's currency. Similarly, if a government official tweets about a significant economic development or event, this could affect the Forex market.

The impact of sentiment analysis from various data sources such as Twitter, Google Search Trends, e- News headlines have been investigated in an effort for researchers to understand the financial markets better. In this domain, Aasi, Imtiaz, Qadeer, Singarajah, and Kashef (2021) suggested a Multivariate Multistep output long-short term memory (MMLSTM) to investigate the impact of sentiment analysis on the prediction of Apple Inc stock, utilising the daily closing price, Twitter data, news headline, and Google trends. Furthermore, they compared the MMLSTM model against the ARIMA, Random Forest and other LSTM models, outperforming them. Their findings revealed that the proposed MMLSTM method improved the MSE by 65% compared to the ARIMA and Random Forest models. Moreover, Cavalli and Amoretti (2021) suggested a novel approach based on the one-dimensional CNN (1D-CNN) to predict the bitcoin trend, utilising multiple information, such as daily historical values, financial indicators and oscillators, and sentiment data. In addition, they employed the Valence Aware Dictionary and sEntiment Reasoner (VADER) to obtain the sentiment score from Twitter. Their findings showed that the proposed 1D-CNN model outperformed the other methods of the baseline (Greaves & Au, 2015), logistic regression, SVM, neural network, random forest, LSTM, and the eXtreme Gradient Boosting (XGBoost). More specifically, their model accomplished the highest accuracy of 74.2% compared to the accuracy of 53.4%, 66%, 65.3%, 55.1%, 57.4%, 67.2%, 48.3% attained from the other algorithms, respectively.

Moreover, one of the primary motivations for the usage and popularity of Twitter sentiment analysis in market prediction is the vast amount of data available on the platform. The Twitter platform has over 330 million monthly active users; on average, 500 million tweets are sent daily. Hence, it provides a large and diverse sample of data that can be employed to investigate investors' sentiment. Furthermore, research in this area has shown a correlation between the sentiment of tweets posted by individual investors and market movements. For example, a study by Tabari, Seyeditabari, Peddi, Hadzikadic, and Zadrozny (2019) examined the sentiment analysis of Twitter by analysing a dataset of 11,000 tweets for Apple, Facebook, and Amazon. The data was retrieved from the Twitter API and labelled using Amazon Mechanical Turk, comparing different neural networks, such as a shallow CNN and a shallow LSTM, against the baseline SVM model. The results showed that the LSTM network outperformed the other methods. Their findings indicated a significant link between tweets and stock returns. Likewise, Aggarwal, Gupta, Garg, and Goel (2019) analysed various factors and parameters from social media such as Twitter and historical prices from Poloniex and datahub.io; they investigated the effect of gold price on the price of bitcoin. In their study, several deep learning models such as CNN, LSTM and GRU are utilised to anticipate the price of bitcoin. The results were that the deep learning models are adequate for Bitcoin price forecasting.

Integrating symbolic methods like lexica, such as VADER, as inputs in neural models like RNNs empowers AI systems to capture explicit rule-based reasoning and complex pattern recognition. This fusion of approaches could contribute to designing more robust AI systems. For example, neural models, such as RNNs, effectively capture sequential dependencies and context over time. Furthermore, by combining symbolic inputs, which provide domain-specific information such as from Twitter and linguistic knowledge, the neural models could better grasp the context in which the data is presented. This enhanced context understanding enables AI systems to perform more accurate cognitive tasks like sentiment analysis, emotion recognition, and language understanding.

The above discussion offered a piece of knowledge on the power of a combination of NLP rule-based systems and neural models that could create more robust AI models in financial markets like Forex. Moreover, it has been revealed that sentiment analysis is crucial in financial predictions, while one of the influential data sources is Twitter. However, the Forex market is highly complicated and influenced by many factors. Therefore, while Twitter could be a valuable source of information for forex traders, it is better to be something other than the whole seed of information on which trading decisions but instead to be combined with other financial data, such as the closing price of an exchange rate. Consequently, researchers have been increasingly interested in finding new ways to predict stock market movements. A pathway could be combining historical stock data and social mood from user-generated content on sources such as Twitter and web news to predict stock prices more accurately.

Even though the historical stock data and social mood can be combined and utilised in ANNs, their monolithic architecture, widely used in price predictions, still presents challenges in high-complexity volatile markets. Hence, monolithic network architectures still need to be investigated in an overview of their applications in financial markets. These challenges will be presented in the next section and identified by aiming to fill them out with the proposed MCoRNNMCD-ANN developed in this thesis. Finally, benchmark state-of-the-art models will be utilised from section 2.6 to be compared with the proposed model. This comparison will validate if the MCoRNNMCD-ANN can outperform the monolithic networks in predicting price movements in Forex markets, as presented in Chapter 5.

2.6 Overview of Machine and Deep Learning Financial Predictive Models

In order to predict challenging financial markets' fluctuations and accurately forecast them, researchers have proposed several machines and deep learning methods such as the CNNs, the variants of RNNs, namely the GRU and LSTM and their hybrid architectures. For instance, Galeshchuk and Mukherjee (2017) suggested a CNN for predicting the price change direction in the Forex market. They utilised the daily closing rates of EUR/USD, GBP/USD, and USD/JPY currency pairs. Moreover, they compared the results of CNN with baseline models such as the majority class (MC), autoregressive integrated moving average (ARIMA), exponential smoothing (ETS), ANN and SVM. Their findings showed that the baseline models and SVM yielded an accuracy of around 65%, while their suggested CNN model had an accuracy of about 75%. Deep learning architectures such as the LSTMs were recommended for future investigation in Forex. Likewise, Shiao et al. (2019) employed the support vector Regression (SVR) and the RNN with LSTM to capture the dynamics of Forex data using the closing price of the USD/JPY exchange rate. The results indicated that their suggested RNN model outperformed the SVR model with an RMSE of 0.0816, which achieved an RMSE of 0.1398, respectively. Maneejuk and Srichaikul (2021) investigated which ARIMA, ANN, RNN, LSTM, and support vector machines (SVM) models presented better performance to the Forex market predictions. They used the daily closing price of five currencies: the Japanese yen, Great Britain Pound, Euro, Swiss franc, and the Canadian dollar for six years. Each model's performance was evaluated using the RMSE, MAE, MAPE and Theil U. Their findings showed that the ANN outperformed the other models in predicting the CHF/USD currency pair. On the other hand, the LSTM obtains better results than the other methods in predicting EUR/USD, GBP/USD, CAD/USD, and JPY/USD rates. For instance, the LSTM achieved the MAE of 0.0300 in the prediction of the EUR/USD compared to the MAE of 0.0435, 0.0319, 0.0853, 0.0560 obtained from the ARIMA, ANN, RNN, LSTM, and SVM models, respectively.

McNally, Roche, and Caton (2018) have also utilised RNNs and LSTM networks to predict the price of Bitcoin in USD for six years. The neural networks were compared to the Autoregressive Integrated Moving Average (ARIMA) method. Their findings unveiled that the LSTM surpassed the RNN and the ARIMA models by an accuracy of 52%. The RNN and the ARIMA achieved an accuracy of 50.25% and 50.05%, respectively. Moreover, Hossain, Karim, Thulasiram, Bruce, and Wang (2018) suggested a model based on deep learning to forecast the stock price of the Standard & Poor's 500 (S&P500) from 1950 to 2016, combining LSTM and GRU networks, comparing to a multilayer perceptron (MLP), CNN, RNN, Average Ensemble, Hand-weighted Ensemble and Blended Ensemble. Their findings revealed that the LSTM-GRU model surpassed the other methods, achieving a mean square error (MSE) of 0.00098, with the other models accomplishing MSE of 0.26, 0.2491. 0.2498. 0.23, 0.23, and 0.226, respectively. Similarly, Althelaya, El-Alfy, and Mohammed (2018) investigated long-short-term memory (LSTM) architectures to forecast the closing prices of S&P500 for eight years. Their findings showed that the Bidirectional LSTM (BLSTM) was the most appropriate model, outperforming the MLP-ANN, the LSTM and the stacked LSTM (SLSTM) models, achieving the lowest error in the short- and long-term predictions. For example, the BLSTM achieved a mean absolute error (MAE) of 0.00736 in the short-term forecasts compared to MAE of 0.03202, 0.01398 and 0.00987 for the MLP-ANN, LSTM and SLSTM, respectively. Ojo, Owolawi, Mphahlele, and Adisa (2019) aimed to predict the behaviour of the stock market using a stacked LSTM network model on the American Stock Exchange. The data used in their study was historical stock market data from the NASDAQ Composite (IXIC) from January 2009 to July 2019. Their model's predictions were compared to the actual stock market behaviour to evaluate its accuracy, revealing that the stacked LSTM network model could predict stock market behaviour with an accuracy of 53.6%. While this accuracy rate is low, it is still significant, considering the complexity of forecasting stock markets. However, their study also noted that predicting the stock market is complicated, and numerous characteristics, such as political events and global news, could influence its behaviour. Lu, Li, Li, Sun, and Wang (2020) proposed a predicting technique for stock prices employing a combination of CNN and LSTM, which utilises the memory function of LSTM to analyse relationships among time series data and the feature extraction capabilities of CNN. Their CNN-LSTM model uses opening, highest, lowest and closing prices, volume, turnover, ups and downs, and change as input and extracts features from the previous ten days of data. Their method is compared to other forecasting models such as LSTM, MLP, CNN, RNN, and CNN-RNN. The results showed that their CNN-LSTM outperformed the other models by presenting an MAE of 27.564 in contrast to MLP with 37.584, CNN with 30.138, RNN with 29.916, LSTM with 28.712 and CNN-RNN with 28.285. They concluded that their proposed CNN-LSTM could provide a reliable reference for investors' investment decisions. However, their model still needs to improve, as it only considers the effect of stock price data on closing prices rather than combining sentiment analysis and national policies into the predictions. Alonso-Monsalve, Suárez-Cetrulo, Cervantes, and Quintana (2020) investigated using CNNs and convolutional LSTM networks as an alternative to conventional MLP and radial basis function (RBF) ANNs for anticipating the price movements of cryptocurrency exchange rates utilising high frequencies. Their study compared the performance of these four different network architectures on six popular cryptocurrencies: Bitcoin, Dash, Ether, Litecoin, Monero, and Ripple. Results showed that convolutional LSTM networks outperformed all other models significantly, with the CNNs also providing good results, particularly for Bitcoin, Ether, and Litecoin. The study concludes that CNNs and convolutional LSTM networks are suitable for predicting the trend of cryptocurrency exchange rates using technical indicators. However, the study is limited to short-term trend prediction, and more research is needed to address limitations for practical application in trading settings.

Researchers have further investigated the influence of ANNs in forecasting prices. For example, Al-Sulaiman (2022) examined the feasibility of predicting future stock processes based on substantial price changes by constructing a deep-forward neural network model with three and four hidden layers. Their proposed model outperformed other techniques, such as linear regression, multiplayer perception, and SVM. Ajoku et al. (2021) explored the usage of ANN in predicting daily stock market prices from various banks. Their research introduced an efficient forecasting model using an Ensemble ANN consisting of three ANNs with three hidden layers utilising the ensemble averaging (EA) theory. Their ensemble model (EM) outperformed traditional MLP neural networks to forecast banks' closing value, presenting fewer errors, 2 in total, against 3, 3.29, 4, and 2.71 from MLPs, respectively. For future research, they proposed utilising other factors, such as macroeconomic and human psychological characteristics, as input variables. Figure 2.10 illustrates their proposed EM.

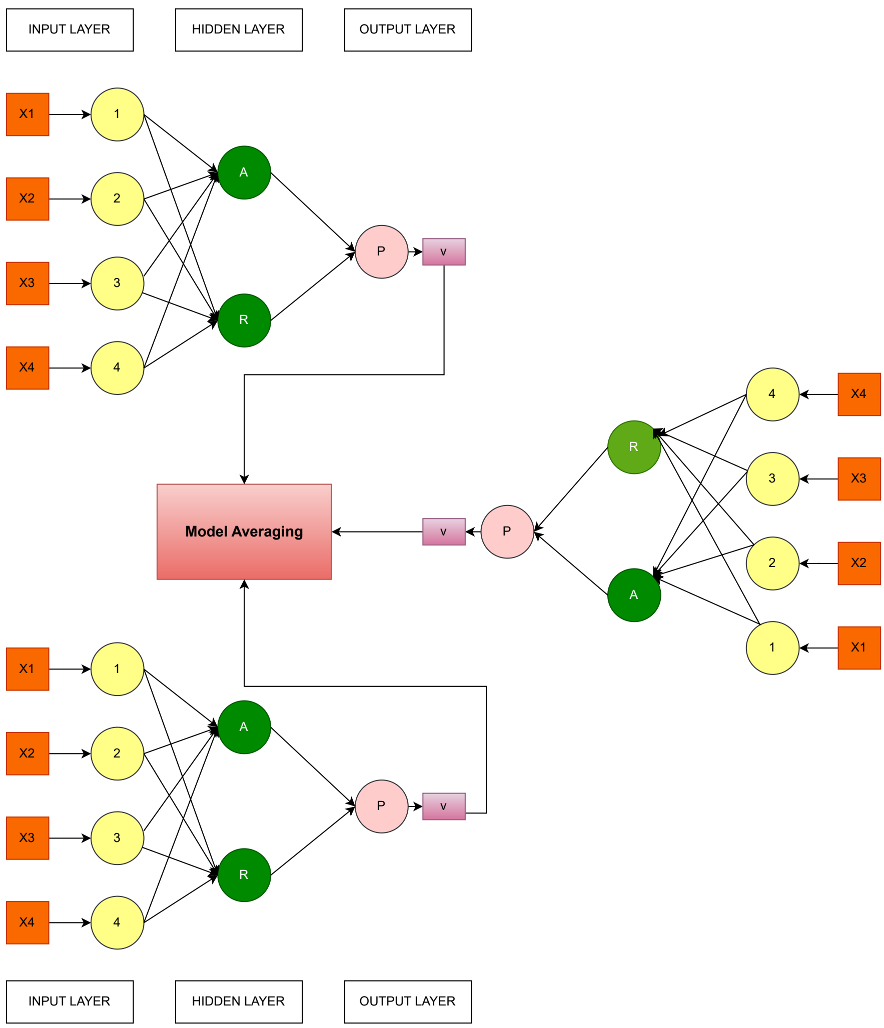


Figure 2.10: Architecture of the Ensemble Model [Ajoku et al. (2021)].

Recently, Y. Zhang, Chu, and Shen (2021) argued that LSTM is a better model than other ANNs for processing complex financial time series data to predict stock price movements. To improve the forecast accuracy, they use investor attention proxies, such as the Baidu index's search volume and news count, and traditional stock variables like price, volume, and technical indexes as inputs for their LSTM model. The results indicated that the LSTM model with the added investor attention proxies outperforms other models, such as ANNs, regarding prediction accuracy and time efficiency. Hence, the authors conclude that the LSTM model with attention proxies is an optimistic method for anticipating stock prices. S. Yang (2021) suggested a model-based deep learning framework using LSTM as a benchmark and GRU models to predict financial indicators values. The financial indicators, such as return on tangible assets (ROTA), price-to-sales ratio (PTSR), and price-earnings ratio (PER) of two stock companies, Mondelez International and Hormel Food Corp, were utilised to verify their proposed framework. Their framework evaluates economic conditions and determines the degree of financial risks for organisations. Their study employs data from the USA stock market but suggests that using data from other markets, such as European and Asian, could improve the results' validity. Their study found that the LSTM and GRU models accurately predict the chosen financial indicators. They suggest incorporating text information from financial websites and news channels in future works could improve prediction accuracy. More recently, Kanwal et al. (2022) proposed a hybrid deep learning technique forecasting the prices of Crude Oil, Crude Oil (CL=F1) and Global X DAX Germany ETF (DAX) for the individual stock item; DAX Performance-Index (GDAXI) and Hang Seng Index (HSI). Their Bidirectional Cuda Deep Neural Network Long Short-Term Memory that compounds BiLSTM Neural Networks and a one-dimensional CNN (BiCuDNNLSTM-1dCNN) compared against the LSTM deep neural network (LSTM-DNN), the LSTM–CNN, the Cuda Deep Neural Network Long Short-Term Memory (CuDNNLSTM), and the LSTM. The results from their study showed that the BiCuDNNLSTM-1dCNN outperformed the other models, validating the outcomes by using the RMSE and MAE metrics; for instance, in the DAX predictions, the BiCuDNNLSTM-1dCNN achieved MAE of 0.566; while the LSTM-DNN, the LSTM–CNN, the CuDNNLSTM, and the LSTM achieved MAE of 0.991, 3.694, 2.729, 4.349 in the test dataset, respectively. Features such as sentiment information have not been exploited in their study. Figure 2.11 illustrates their proposed BiCuDNNLSTM-1dCNN.

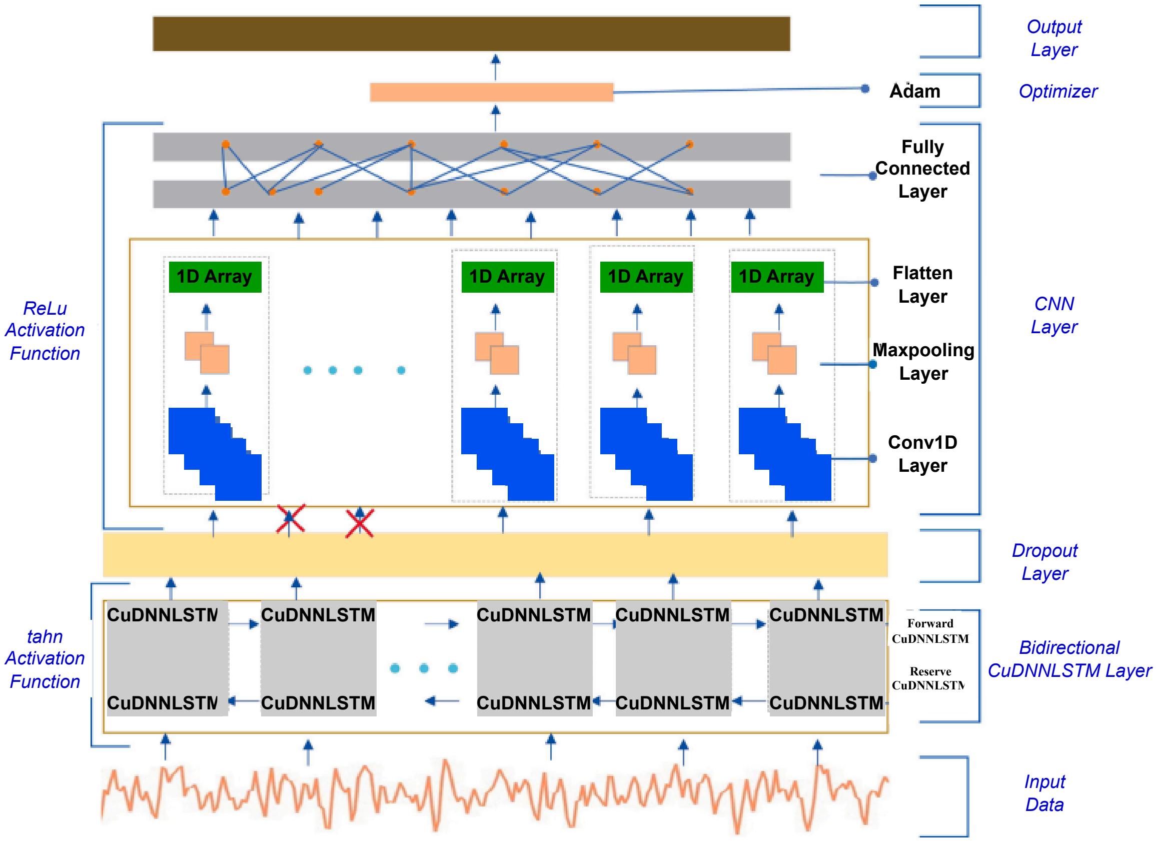


Figure 2.11: Architecture of the BiCuDNNLSTM-1dCNN [Kanwal et al. (2022)].

Zhou, Zhou, and Wang (2022) proposed a hybrid stock forecasting model (FS-CNN-BGRU) that combines Feature Selection (FS), CNN, and Bidirectional Gated Recurrent Unit (BGRU) to predict stock performance. Their model was evaluated and compared to single models such as the CNN, LSTM, and GRU and mixed models like the CNN-LSTM and CNN-GRU. The results show that the FS-CNN-BGRU model outperforms all other models regarding prediction error (MAPE) and R2. Their proposed FS-CNN-BGRU model achieved an error of 1.4325%, which is lower than that of CNN-GRU at 1.6354%, CNN-LSTM is 1.6426%, that GRU is 1.8332%, LSTM is 1.8654%, and that of CNN is 2.0601%. Finally, researchers proposed that there is still space for progress. Pokhrel et al. (2022) compared the performance of three deep learning models, LSTM, GRU, and CNN, in predicting the next day's closing price of the Nepal Stock Exchange (NEPSE) index. The study uses fundamental market data, macroeconomic data, technical indicators, and financial text data of Nepal's stock market. Their models' performances are compared using standard assessment metrics like Root Mean Square Error (RMSE), MAPE, and Correlation Coefficient (R). Their results indicated that the LSTM model architecture provides a superior fit with The smallest RMSE 10.4660 MAPE 0.6488 and with R score 0.9874 in contrast to the GRU with RMSE 12.0706, MAPE 0.7350, R 0.9839, and the CNN with RMSE

13.6554, GRU 0.8424, R 0.9782. Their study also suggested that the LSTM model with 30 neurons was the supreme conqueror, followed by GRU with 50 neurons and CNN with 30 neurons. Finally, they proposed developing hybrid predictive models, implementing hybrid optimisation algorithms, and comprising other media sentiments in the model development methodology for future work. Undoubtedly, it is evident that the CNNs as a financial forecasting model in regression analysis is considered one of the most potent techniques in academic settings that consistently outperformed other state-of-the-art models, including different ANNs, LSTMs, and fuzzy-based approaches (Kirisci & Cagcag Yolcu, 2022).

The studies above have yielded noteworthy achievements in forecasting financial markets. However, scholars have underscored the untapped potential for exploration within the domain of time series models, like LSTM and GRU, in Forex predictions. Renowned for their adeptness in capturing long-term dependencies within time-series data, these models offer a promising avenue for enhancing the accuracy of Forex forecasts. Likewise, within the Forex market context, researchers have mainly highlighted the investigation of Modular Neural Networks (MNNs) as a unique approach, alongside the rising trend of NLP, which has yet to be extensively explored in the forecast of price fluctuations (Islam et al., 2020; Sezer et al., 2020). However, the challenge of developing MNN architectures is that it can take time to design and train the individual modules to lead to an effective combination in the final network decision; therefore, more research is needed to determine their effectiveness and practicality in this field.

Regarding structural elements and parameters of models, it has been observed that while dropout is widely employed in stock predictions, the utilisation of Monte Carlo dropout (MCD) relevant to uncertainty quantification in Forex predictions has been relatively limited in an effort to improve predictions. MCD, utilising dropout during the inference phase, provides the model not only a point estimate but also an estimate of uncertainty associated with each prediction (Miok, Nguyen-Doan, Škrlj, Zaharie, & Robnik-Šikonja, 2019; Zhu & Laptev, 2017). Hence, the MCD used has yet to be thoroughly investigated for possible contributing to better Forex predictions. Additionally, non-orthogonal weight initialisation methods are commonly used in stock prediction models. However, the potential benefits of orthogonal weight initialisation methods in addressing issues like the vanishing gradient problem and improving the optimisation process in Forex prediction models have yet to be extensively examined (Duan & Wang, 2016; Saxe, McClelland, & Ganguli, 2013).

Finally, Table 2.1 encapsulates the state-of-the-art models, highlighting research gaps and future opportunities pertinent to this thesis. The overarching purpose is to contribute to advancing more accurate, robust, and adaptable models, thereby enabling a potentially better understanding in anticipation of the intricacies inherent in the Forex market, ultimately benefiting financial decision-makers.

Table 2.1: Existing work in deep learning models and future research directions.

Authors	Model	Type	Gaps and Future Research
Kanwal et al. (2022)	BiCuDNNLSTM	Hybrid	Feature extraction from multivariate datasets
Lu et al. (2020)	CNN–LSTM	Hybrid	Integration of emotional factors
Hossain et al. (2018)	LSTM–GRU	Hybrid	Application in domains such as Forex
Alonso-Monsalve et al. (2020)	CLSTM	Hybrid	Window size and network structure adjustment
Ajoku et al. (2021)	EM	EA	Consider macroeconomic factors and news
Pokhrel et al. (2022)	2D-CNN	Single	Sentiment incorporation
Pokhrel et al. (2022)	GRU	Single	Sentiment incorporation
Pokhrel et al. (2022)	LSTM	Single	Sentiment incorporation
Islam et al. (2020)	Various	Various	Investigation of modular architectures

2.7 Critical Analysis

The multidisciplinary review within this study, incorporating recent neuroscience and financial market insights, underscores the ongoing need to enhance machine and deep learning methods. It also highlights the importance of modular design as a solution to the challenges posed by monolithic architectures (Islam et al., 2020). Monolithic neural networks often suffer from catastrophic forgetting when learning new skills, altering their previously acquired knowledge. This study advocates for neural networks inspired by the modular organisation of human and animal brains, capable of integrating new knowledge without erasing existing knowledge—a fundamental consideration (Ellefsen, Mouret, & Clune, 2015). In addition, the direction of examining investors' sentiment combined with economic indicators like closing prices is a promising trend requiring further investigation (Sezer et al., 2020).

In the realm of computational models, recent studies highlight the significance of techniques like orthogonal initialisation and MCD, which improved the performance of ANNs (Duan & Wang, 2016; Miok et al., 2019). These techniques diverge from models relying solely on default weights and conventional dropout methods frequently implied in exploring financial predictive models from the literature, conceivably by enhancing predictive performance. Simultaneously, primary data plays a pivotal role in this research, offering a direct path to its aim of forecasting the hourly closing price of EUR/GBP, which is integral to financial analysis (Barcellos & Zamarro, 2021). These data, meticulously gathered from Yahoo Finance (closing prices) and Twitter (sentiments) APIs, seamlessly align with the study's context (Nobata, Tetreault, Thomas, Mehdad, & Chang, 2016; J. Yang & Counts, 2010). Beyond introducing and comparing baseline models to optimally partition the data, these sources enable a comprehensive assessment of state-of-the-art hybrid, ensemble and single monolithic architectures selected from the literature, which were relevant to this study's aim and feasible for replication.

2.7.1 Baseline Models

The significance of baselines is crucial in this study as they were created to address research gaps, such as the limited utilisation of MCD and the orthogonal kernel initialisation, reducing overfitting and potentially enriching the Forex market's anticipation. These new models provide a starting point for further analysis. They could help researchers identify areas for improvement as an essential tool in designing possible more accurate predictive models discussed further in Chapters 4 and 5. Moreover, baselines are vital for effectively partitioning the input domain in the context of Forex predictions. This partitioning, in turn, optimally allocates inputs, thereby enhancing task performance. This importance is substantiated by primary research leveraging closing prices and sentiment scores from Yahoo Finance and Twitter Streaming APIs as inputs aggregated based on hourly rates within 2018–2019. Tables 2.2, 2.3, and 2.4 present the test performance of the baseline models in anticipating the EUR/GBP hourly closing price based on the MSE, MAE, and Mean Squared Logarithmic Error (MSLE) objective evaluation metrics.

Table 2.2: Baseline models performance metrics in closing prices (CP) of EUR/GBP.

Model	Variables	Metrics	Train	Valid	Test	R^2	Time Duration
CoRNNMCD	CP	MSE	6.7184×10^{-5}	6.2938×10^{-5}	$\mathbf{5.8785 \times 10^{-5}}$	0.99	2:30
		MAE	0.00549	0.00538	**0.00529**		
		MSLE	3.1801×10^{-5}	2.79863×10^{-5}	$\mathbf{2.6419 \times 10^{-5}}$		
CoRNN	CP	MSE	6.8642×10^{-5}	6.3699×10^{-5}	5.9447×10^{-5}	0.99	1:27
		MAE	0.00551	0.00541	0.00532		
		MSLE	3.2413×10^{-5}	2.8352×10^{-5}	2.6773×10^{-5}		
CoGRUMCD	CP	MSE	6.9541×10^{-5}	6.4196×10^{-5}	5.9700×10^{-5}	0.99	7:38
		MAE	0.00551	0.00542	0.00532		
		MSLE	3.3150×10^{-5}	2.8718×10^{-5}	2.7013×10^{-5}		
CoGRU	CP	MSE	6.9597×10^{-5}	6.4106×10^{-5}	5.9565×10^{-5}	0.99	3:59
		MAE	0.00552	0.00541	0.00532		
		MSLE	3.3222×10^{-5}	2.8713×10^{-5}	2.6914×10^{-5}		
1D-CNN	CP	MSE	0.00012	0.00011	0.00011	0.99	**0:36**
		MAE	0.00763	0.00757	0.00747		
		MSLE	5.3941×10^{-5}	4.800×10^{-5}	4.6293×10^{-5}		

Table 2.3: Baseline models performance metrics in sentiment scores of EUR/GBP.

Model	Variables	Metrics	Train	Valid	Test	R^2	Time Duration
CoRNNMCD	Sentiment	MSE	0.00079	0.00076	0.00067	0.62	2:21
		MAE	0.01535	0.01512	0.01456		
		MSLE	0.00031	0.00027	0.00026		
CoRNN	Sentiment	MSE	0.00077	0.00076	0.00066	0.63	1:24
		MAE	0.01504	0.01489	0.01428		
		MSLE	0.00029	0.00027	0.00025		
CoGRUMCD	Sentiment	MSE	0.00076	0.00074	**0.00065**	0.64	6:09
		MAE	0.01465	0.01439	**0.01394**		
		MSLE	0.00029	0.00026	**0.00024**		
CoGRU	Sentiment	MSE	0.00077	0.00075	0.00066	0.63	3:47
		MAE	0.01497	0.01479	0.01418		
		MSLE	0.00030	0.00027	0.00025		
1D-CNN	Sentiment	MSE	0.00085	0.00083	0.00073	0.61	**0:36**
		MAE	0.01661	0.01644	0.01591		
		MSLE	0.00034	0.00031	0.00028		

Table 2.4: Baseline models performance metrics in closing prices in CRNN and sentiment scores in CGRU of EUR/GBP.

Model	Variables	Metrics	Train	Valid	Test	R^2	Time Duration
CRNN	CP	MSE	6.8795×10^{-5}	6.4113×10^{-5}	5.9921×10^{-5}	0.99	1:42
		MAE	0.00530	0.00544	0.00535		
		MSLE	3.2451×10^{-5}	2.8487×10^{-5}	2.6988×10^{-5}		
CGRU	Sentiment	MSE	0.00079	0.00076	0.00068	0.63	3:10
		MAE	0.01577	0.01545	0.01519		
		MSLE	0.00030	0.00027	0.00026		

The CoRNNMCD and the CoGRUMCD performed better than the other baseline models in Tables 2.2 and 2.3, presenting less error in the MSE, MAE, and MSLE test sets for the closing prices and sentiment scores, respectively. Moreover, these two baselines will be used to develop the proposed Modular Neural Network model. For instance, CoRNNMCD in closing prices (Table 2.2) demonstrated fewer errors in the test sets, decreasing the MSE by 1.12%, 1.54%, 1.32%, and 60.68% for the CoRNN, CoGRUMCD, CoGRU, and 1D–CNN, respectively. Likewise, sentiment scores (Table 2.3) presented better arrangement in CoGRUMCD with fewer errors in test sets by decreasing the MSE by 3%, 1.52%, 1.52%, and 11.59% for the CoRNNMCD, CoRNN, CoGRU, and 1D–CNN, respectively. The typical 1D-CNNs did not employ the orthogonal RNNs coupled with MCD instead of pooling layers and were used as a baseline, showing less execution time in closing prices and sentiments. However, 1D–CNNs MSE was significantly higher than the other baselines and performed worse. Also, it has been observed that using MCD could increase baseline computational time. Nevertheless, the MCD application significantly improved performance in the selected baselines.

All representatives' R-squared in Table 2.2 was high (R^2), meaning the models can fit well with the datasets. However, in Tables 2.3 and 2.4 for the CGRU, the models' more moderate R^2 value has been observed. On the other hand, a high R-squared does not mean a correlation with objective evaluations such as the MSE, which can be very useful for comparing the models to provide a more comprehensive evaluation of the predictions. Finally, Table 2.4 shows that the best-performed CoRNNMCD and CoGRUMCD significantly outperformed the convolutional RNN (CRNN) and convolutional GRU (CGRU), presenting 1.92% and 4.5% lower MSE in the test set. These results prove the efficiency of the suggested adaptive mechanism of this thesis, consisting of MCD and orthogonal kernel initialisation against the models that did not imply it, like the CRNN and CGRU.

2.7.2 Hybrid and Ensemble Benchmark Models

The choice of hybrid algorithms for this study prioritised adopting the most current, reputable, and state-of-the-art techniques available, which can be replicated as well, according to the provided information by the authors. This focus on the most recent and state-of-the-art models ensures that the study is grounded in the latest developments and contributes to advancing understanding in the forecast of hourly EUR/GBP price fluctuations.

Table 2.5 shows that the CNN–LSTM performed better than the other models, presenting more inconsequential errors in the MSE, MAE, and MSLE test sets.

Table 2.5: Hybrid models' performance metrics receive closing prices and sentiment scores of EUR/GBP.

Model	Metrics	Train	Valid	Test	R^2	Time Duration
BiCuDNNLSTM	MSE	0.00013	0.00015	0.00015	0.99	1:19
	MAE	0.00848	0.00874	0.00866		
	MSLE	6.4001×10^{-5}	6.6003×10^{-5}	6.9725×10^{-5}		
CNN-LSTM	MSE	6.4471×10^{-5}	5.9427×10^{-5}	$\mathbf{7.005 \times 10^{-5}}$	0.99	1:41
	MAE	0.00538	0.00536	**0.00552**		
	MSLE	3.0273×10^{-5}	2.6574×10^{-5}	$\mathbf{3.2089 \times 10^{-5}}$		
LSTM-GRU	MSE	0.00017	0.00014	0.00014	0.99	3:03
	MAE	0.00935	0.00901	0.00887		
	MSLE	7.8049×10^{-5}	6.3934×10^{-5}	6.4782×10^{-5}		
CLSTM	MSE	0.00501	0.00514	0.00507	0.79	2:18
	MAE	0.05722	0.05764	0.05743		
	MSLE	0.00229	0.00231	0.00234		
Ensemble	MSE	0.00325	0.00337	0.00320	-0.48	**0:22**
	MAE	0.03716	0.03761	0.03665		
	MSLE	0.00149	0.00152	0.00146		

For example, CNN–LSTM demonstrated fewer errors in the test sets, decreasing the MSE by 72.66%, 66.61%, 194.55%, and 191.43% for the BiCuDNNLSTM, LSTM–GRU, CLSTM, and ensemble learning respectively. The ensemble learning approach presented less execution time due to the less intricate architecture of the ANNs, but its MSE was significantly higher than the CNN–LSTM. It is worth mentioning that the BiCuDNNLSTM is running in GPU based on CUDA utilisation, which can boost the speed of training time of deep learning models. Moreover, factors such as the time steps of each hybrid model can affect its execution time, as discussed in Chapter 5. Finally, the hybrid models presented a high R^2, with the CLSTM showing a moderate R^2 value and the ensemble learning technique showing a negative R^2. The negative R^2 could indicate a fundamental flaw in the chosen ensemble learning or between the ensemble model and the underlying structure of the data.

2.7.3 Single Benchmark Models

Likewise, the choice of algorithms for this study strongly emphasised selecting the most recent single methods employed for the possible hourly price fluctuation forecast in the EUR/GBP.

Table 2.6 revealed that the GRU performed better than the other models, presenting less error in the MSE, MAE, and MSLE test sets.

Table 2.6: Single models' performance metrics receive closing prices and sentiment scores of EUR/GBP.

Model	Metrics	Train	Valid	Test	R^2	Time Duration
2D-CNN	MSE	0.00012	0.00011	0.00012	0.99	0:51
	MAE	0.00760	0.00746	0.00765		
	MSLE	5.3407×10^{-5}	4.8839×10^{-5}	5.9233×10^{-5}		
GRU	MSE	8.7491×10^{-5}	7.8116×10^{-5}	$\mathbf{9.1750 \times 10^{-5}}$	0.99	0:47
	MAE	0.00628	0.00595	**0.00614**		
	MSLE	4.0481×10^{-5}	3.5261×10^{-5}	$\mathbf{4.5704 \times 10^{-5}}$		
LSTM	MSE	0.00732	0.00723	0.00736	0.77	**0:10**
	MAE	0.04911	0.04911	0.04922		
	MSLE	0.00334	0.00327	0.00338		

For instance, GRU exhibited fewer errors in the test sets, decreasing the MSE by 26.68% and 195.01% for the 2D–CNN and LSTM, respectively. LSTM presented less execution time, implying 30 neurons and an Adam optimiser that can obtain a faster convergence rate. However, the MSE of LSTM was considerably higher than the GRU. The single models also presented a high R^2, with the LSTM presenting a more moderate R^2 value.

2.7.4 Discussion

This chapter explored the integration of RCT modelling with contemporary neuroscience insights and possible simulation by AI systems, specifically within the intricate domain of financial predictions in the Forex market. Moreover, the critical scrutiny of computational models influencing financial landscapes lays a robust groundwork, elucidating the intricate interplay among RCT, neuroscience, and AI systems. Within this comprehensive exploration, the analysis identified pivotal research gaps in Forex prediction. A compelling imperative emerges for investigating modular neural network architectures, addressing the underutilisation of Monte Carlo dropout and orthogonal weight initialisation methods.

On the other hand, to strategically overcome challenges arising from limited data in Forex predictions, the thesis proposes techniques such as partial transfer learning to fortify generalisation performance amidst data scarcity. Significantly, this goes beyond establishing benchmarks; the goal is the genuine enhancement of generalisation performance in ANNs. As proposed in forthcoming chapters (Chapter 4 and Chapter 5), the deliberate use of insights from the proposed bio-inspired modular ANN is envisioned to catapult predictive capabilities beyond those afforded by a simpler modular ANN, which lacks the enriching insights of the advanced model. It is important to note that transfer learning is a pivotal component of this thesis; it is not used solely as a benchmark but as an integral element amplifying the thesis's contribution to advancements in financial predictions.

The subsequent chapter will thoroughly examine the computational properties inherent in neural networks, emphasising specific architectural considerations relevant to Forex prediction. This in-depth study spans the intricacies of Monte Carlo dropout, orthogonality, and pertinent optimisers, culminating in a comprehensive framework that addresses the identified research gaps and propels the field of Forex prediction models into new realms of sophistication.

Simultaneously, informed by the review of bio-inspired models, this thesis strategically incorporates highly effective models such as CNN and RNNs. This selection establishes a foundation for further exploration of novel models in the financial domain proposed in Chapter 4. As an additional purpose, the application and potential impact of transfer learning in enhancing the models' predictive capabilities receiving less data in another similar task in the Forex market could enrich the discourse on financial modelling.

3. Computational Properties of Neural Networks

This chapter will examine the computational foundation properties of neural networks, including their architecture. Furthermore, it will delve into the specific computational properties of Monte Carlo dropout and orthogonal weight initialisation methods. These techniques have shown promise in various domains and have the potential to significantly enhance the performance of neural network models in Forex prediction. In addition to Monte Carlo dropout and orthogonal weight initialisation, different optimisation algorithms are explored in training neural networks.

3.1 ANN Architecture

An ANN is a feedforward artificial neural network of multiple layers of interconnected neurons. The equation for a primary ANN network is denoted as,

$$y = f\left(\sum_{i=1}^{n} w_i x_i + b\right), \tag{3.1}$$

where y is the outcome of the neural network, f is the activation function (e.g., sigmoid, ReLU), w_i is the weight for the input x_i, and b is the bias term. 3.1 represents a single neuron in the network. An artificial neural network would have multiple layers of neurons, each with weights and biases. The outcome of one layer is utilised as the input for the next layer.

Another version with more layers of the ANN or a Multilayer Perceptron (MLP) can be represented by:

$$y = f_l\left(f_{l-1}\left(\dots f_2\left(f_1\left(\sum_{i=1}^{n} w_{1i} x_i + b_1\right)\right)\dots\right)\right), \tag{3.2}$$

where, l is the number of layers, f_i is the activation function for the -th layer, w_i is the weight for the input x_i, and b_i is the bias term for the -th layer.

3.2 represents the output of the ANN network, a function of multiple layers of neurons with different activation functions, weights, and biases. In the following sections of this chapter, different types of ANN such as CNNs and RNNs will be discussed.

3.2 CNN Architecture

CNN is a type of ANN for processing data with a grid pattern, such as images, inspired by the organization of animal visual cortex (Fukushima, 1980). CNN is commonly employed for image and video processing tasks, but it can also be used for processing time series data. A convolutional layer applies 1D filters to the input data. Each filter is a small 1D window that slides over the input data and performs a dot product operation between the filter and the portion of the input data the filter is currently "looking at." The result of this dot product operation is called convolution. Mathematically, the convolution between a 1D filter "w" of length "k" and an input signal "x" of length "n" can be represented as:

$$(x * w)[i] = \sum_{j=1}^{k} x[i+j]w[j] \tag{3.3}$$

Where * denotes the convolution operation, the output is represented by (x * w), and i, j is the multiplied element's index.

After the convolution, a bias term b is added to each element of the convolution computed as:

$$y[i] = f((x * w)[i] + b) \tag{3.4}$$

Where f is the activation function, in this case, it could be ReLU, for example, which is illustrated as:

$$(x) = max(0, x) \tag{3.5}$$

A max-pooling layer is then typically applied to the convolutional layer's output to reduce the data's spatial dimensions and make the network more invariant to small translations of the input data. The max pooling layer can be defined as follows:

$$y[i] = \max_{j=0}^{k-1} x[i \times s + j] \tag{3.6}$$

Here, x is the input time series, y is the output time series, i is the output index, and k is the size of the pooling window or the 'pool size'. The operator max returns the maximum value within the pooling window of size k at the i-th position.

After the max pooling, the output of the pooling layer is flattening and can be given as :

$$y[i] = flatten(x[i, j)] \tag{3.7}$$

Where x is the input, which is the output of the convolutional layer or max pooling operation, and y is the flattened output.

Finally, to generate the network's output, the dense layer's output is passed through a softmax activation function, which converts the output of the dense layer into a probability distribution over the possible classes. Then, for a dense layer, the output can be represented mathematically as:

$$y = f(W_y x + b) \tag{3.8}$$

Where x The flattening output is used as input to the dense layer, W_y is the weight matrix, b is the bias term, and f is the activation function. A typical CNN architecture for time series consists of a one-dimensional convolutional layer, a polling layer flattened to a dense layer, which is illustrated in Figure 3.1.

CNN pseudocode can be described as follows:

Algorithm 1 CNN

procedure CNN(input: x, filter: w, bias: b, pool size: k, stride: s, weight: W_y)
    Convolutional Layer:
    for $i = 1$ to $n - k + 1$ **do**
        $y[i] \leftarrow 0$
        for $j = 1$ to k **do**
            $y[i] += x[i + j] \cdot w[j]$
        end for
        $y[i] += b$
        $y[i] \leftarrow \max(0, y[i])$
    end for
    Max Pooling Layer:
    for $i = 0$ to $\left\lfloor \frac{n}{s} \right\rfloor - 1$ **do**
        $y[i] \leftarrow \max_{j=0}^{k-1} x[i \cdot s + j]$
    end for
    Flatten Layer:
    $y \leftarrow \text{flatten}(x)$
    Dense Layer:
    $y \leftarrow \text{softmax}(W_y \cdot x + b)$
end procedure
CNN(x, w, b, k, s, W_y)

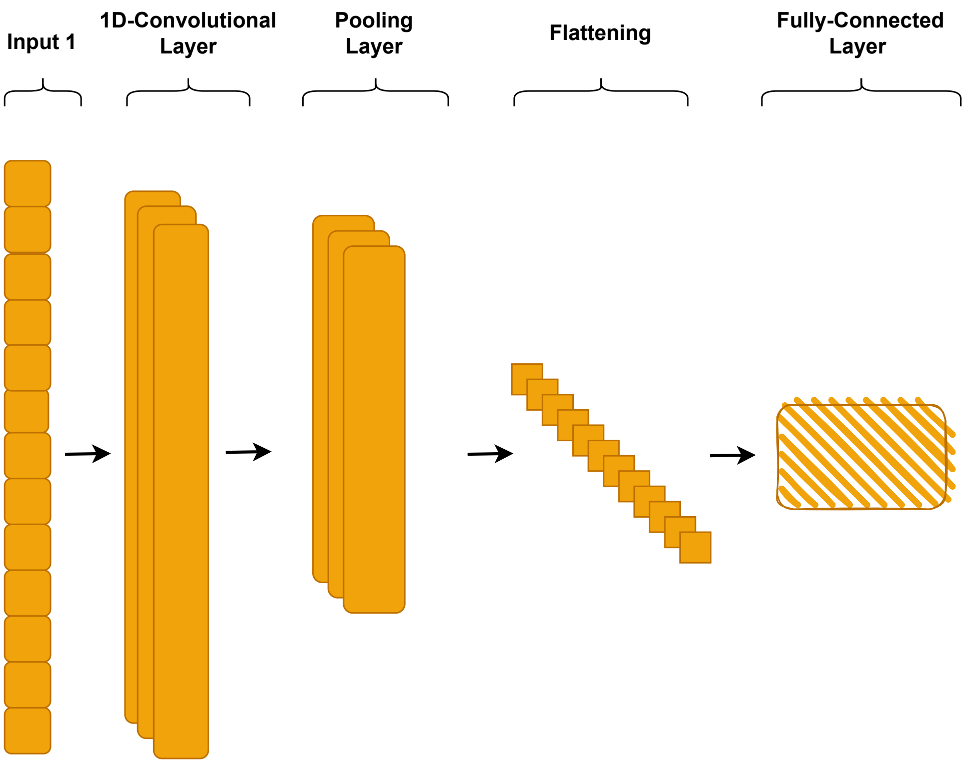


Figure 3.1: CNN Architecture.

3.3 Simple RNN Architecture

An RNN is a neural network well-suited for application in sequential data, such as time series or natural language. An RNN is a set of hidden units that process input data and maintain a state, or memory, that captures information about the past (Hochreiter, 1991). A vector h represents the state of the RNN at any given time step.

At each time step t, the RNN processes an input x_t and updates its hidden state h_t using the following equations:

$$h_t = \tanh(W_{hh}h_{t-1} + W_{xh}x_t + b_h) \tag{3.9}$$

Here, h_t is the hidden state at time step t, h_{t-1} is the hidden state at the previous time step, x_t is the input at time step t, W_{hh}, W_{xh} are weight matrices, and b_h is a bias term.

The output of the RNN at time steps t:

$$y_t = \mathrm{softmax}(W_{hy}h_t + b_y) \tag{3.10}$$

Here, y_t is the output at time step t, h_t is the hidden state at time step t, W_hy is a weight matrix, and b_y is a bias term. RNN is illustrated in Figure 3.2.

RNN pseudocode can be described as follows:

Algorithm 2 RNN

procedure RNN
  initialize h_0
  for each time step t **do**
    $h_t = \tanh(W_{hh}h_{t-1} + W_{xh}x_t + b_h)$
    $y_t = \text{softmax}(W_{hy}h_t + b_y)$
  end for
end procedure

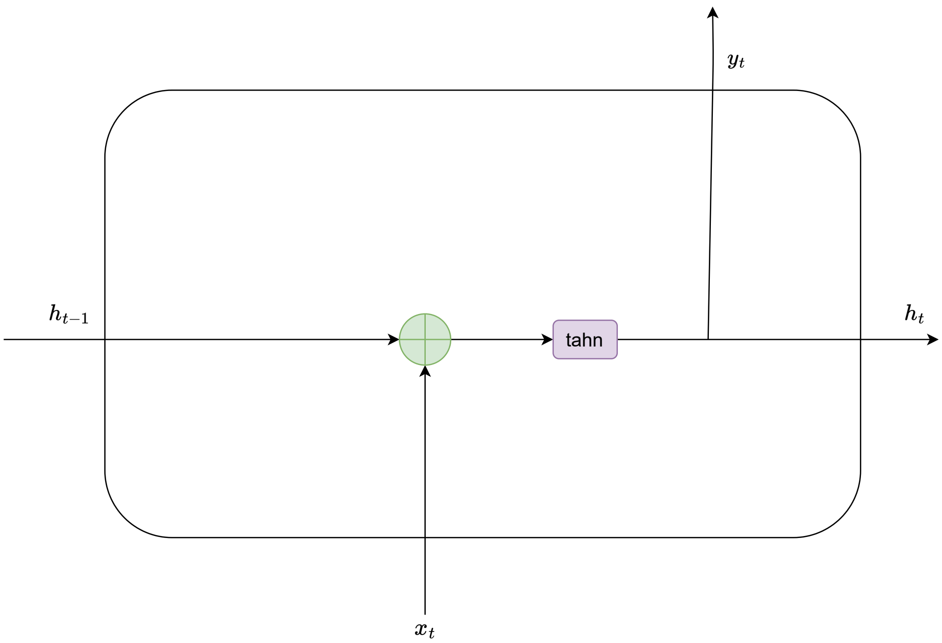


Figure 3.2: RNN Architecture.

To train the RNN, we use backpropagation through time (BPTT), which involves unrolling the RNN in time and applying the chain rule to compute the gradient of the loss function regarding the parameters of the model given as:

$$\frac{\partial L}{\partial p} = \sum_{t=1}^{T} \frac{\partial L}{\partial y_t} \frac{\partial y_t}{\partial p} \tag{3.11}$$

Here, L is the loss function, and $\frac{\partial L}{\partial y_t}$ is the gradient of the loss function concerning the output at time step t. The gradient of the output at time step t for the model parameter p can be computed using the chain rule as follows:

$$\frac{\partial y_t}{\partial p} = \frac{\partial y_t}{\partial h_t} \frac{\partial h_t}{\partial p} \tag{3.12}$$

The gradient of the hidden state at time steps t concerning the model parameter p can be calculated employing the chain rule as follows:

$$\frac{\partial h_t}{\partial p} = \frac{\partial h_t}{\partial h_{t-1}} \frac{\partial h_{t-1}}{\partial p} \tag{3.13}$$

If the gradients of the weights in the network become very small or "vanish" as they are backpropagated through many time steps, it can make it challenging for the network to learn effectively. This problem is known as the vanishing issue. The vanishing issue can occur if the gradients of the hidden states concerning the previous hidden states are small, i.e., $\frac{\partial h_t}{\partial h_{t-1}} \approx 0$. For example, this can happen if the activation function used in the network, such as the hyperbolic tangent function, has small gradients when the input is significant in magnitude. To address the vanishing gradients problem, several techniques have been developed, such as using a different type of activation function (e.g., ReLU), using a different type of RNN architecture (e.g., LSTM or GRU), or using more advanced optimisation algorithms (e.g., Adam).

3.3.1 LSTM Architecture

An LSTM, is an RNN capable of learning long-term dependencies in data (Hochreiter & Schmidhuber, 1997). This capability is in contrast to traditional RNNs, which need help learning long-term dependencies due to the problem of vanishing gradients.

LSTMs achieve this by introducing a few additional mechanisms to the traditional RNN architecture. These include an input gate, an output gate, and a forget gate. The input gate controls data flow into the LSTM cell, the output gate controls data flow out of the cell, and the forget gate controls which information to discard from the cell state. The equations for the LSTM are given below:

Input gate:

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{3.14}$$

Forget gate:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \tag{3.15}$$

Output gate:

$$o_t = \sigma(W_o[ht-1, x_t] + b_o) \tag{3.16}$$

Cell input:

$$\tilde{C}_t = \tanh(W_C[ht-1, x_t] + b_C) \tag{3.17}$$

Cell state:

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}t \tag{3.18}$$

Hidden state:

$$h_t = o_t \odot \tanh(C_t) \tag{3.19}$$

Where: h_t is the hidden state at time t, x_t is the input at period t, f_t, i_t and o_t are the forget, input, and output gates, respectively. $\tilde{C}_t$ is the candidate cell state, C_t is the cell state at time t, W_f, W_i, W_C, W_o are weight matrices and b_f, b_i, b_C, b_o are bias vectors. LSTM is displayed in Figure 3.3.

LSTM pseudocode can be described as follows:

Algorithm 3 LSTM

procedure LSTM
  initialize C_0, h_0
  for each time step t **do**
    $f_t = \sigma(W_f[h_{t-1}, x_t] + b_f)$
    $i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)$
    $\tilde{C}t = \tanh(W_C[ht - 1, x_t] + b_C)$
    $C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}t$
    $o_t = \sigma(W_o[ht - 1, x_t] + b_o)$
    $h_t = o_t \odot \tanh(C_t)$
  end for
end procedure

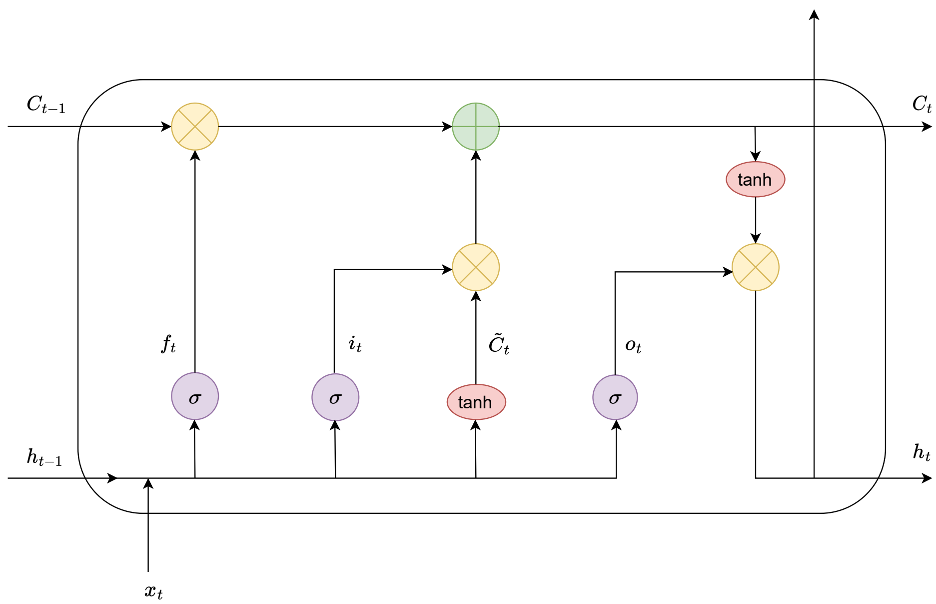


Figure 3.3: LSTM Architecture.

3.3.2 GRU Architecture

The GRU is a type of RNN architecture (Cho et al., 2014). Like other RNNs, a GRU processes sequential input data by iteratively updating a hidden state, allowing it to make decisions based on the entire history of the input seen so far. However, unlike traditional RNNs, which use a simple update function to combine the previous hidden state with the current input, GRUs use a more complex gating mechanism to control the flow of information through the hidden state. Two gates control the update process in a GRU: the "reset gate," denoted by r, and the "update gate," represented by z. These gates are operated to decide how much of the previous hidden state, indicated by h_{t-1}, should be used in the update process and how much of the current input, denoted by x_t, should be used to update the hidden state at the current timestep, indicated by h_t.

The reset gate and update gate are calculated as follows:

$$r_t = \sigma(W_r[h_{t-1}, x_t] + b_r) \tag{3.20}$$

$$z_t = \sigma(W_z[h_{t-1}, x_t] + b_z) \tag{3.21}$$

$$\tilde{h}t = \tanh(W_h[r_t \odot ht - 1, x_t] + b_h) \tag{3.22}$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \tag{3.23}$$

Where: h_t is the hidden state at time t, x_t is the intake at time t, r_t and z_t are the reset and update gates, respectively. $\tilde{h}_t$ is the candidate hidden state, W_r, W_z, W_h are weight matrices, b_r, b_z, b_h are bias vectors. Figure 3.4 shows the GRU.

GRU pseudocode can be described as follows:

Algorithm 4 GRU

procedure GRU
    initialize h_0
    for each time step t **do**
        $r_t = \sigma(W_r[h_{t-1}, x_t] + b_r)$
        $z_t = \sigma(W_z[h_{t-1}, x_t] + b_z)$
        $\tilde{h}t = \tanh(W_h[r_t \odot ht - 1, x_t] + b_h)$
        $h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t$
    end for
end procedure

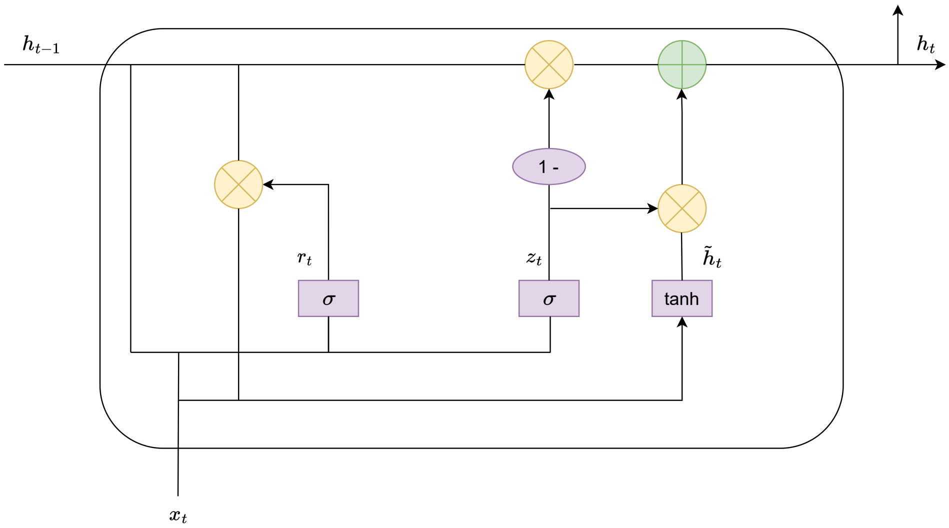


Figure 3.4: GRU Architecture.

3.4 Orthogonal Initialization

In orthogonal kernel initialisation, the input weight matrix is initialised with an orthogonal matrix O, a square matrix whose columns are orthogonal unit vectors. This process ensures that the input to the RNN or GRU is appropriately scaled and decorrelated, aiding in training and improving performance. The theorem of the orthogonal matrix can be stated as follows:

Let W be the weight matrix initialised with an orthogonal matrix O. The goal is to show that the matrix O preserves the orthogonality of W.

The orthogonality of a matrix W can be defined as:

$$W^T W = I \tag{3.24}$$

If we initialise W with an orthogonal matrix O, we have:

$$W = O \times U \tag{3.25}$$

where U is a learned matrix.

Then, compute the transpose of W as:

$$W^T = (O \times U)^T = U^T O^T = U^T O^{-1} \tag{3.26}$$

since O is orthogonal, its inverse is its transpose.

Substituting W^T in the orthogonality definition in Equation 3.24, getting:

$$W^T W = (U^T O^{-1})(O \times U) = U^T U = I \tag{3.27}$$

Therefore:

$$W^T W = I, \tag{3.28}$$

which shows that the matrix O preserves the orthogonality of the weight matrix W.

Orthogonal matrices have some valuable properties, such as preserving the vector's magnitude and the angle between the vectors; this is why it is helpful in RNNs to avoid issues such as vanishing problems.

3.5 Bayesian Decision Theory in Finance

Bayesian decision theory is a formal and mathematical framework used to model decision-making in uncertain situations. It is based on maximising expected utility, which involves considering the probabilities of different outcomes and their associated utility to determine the best course of action.

In finance, Bayesian decision theory finds applications in various areas, such as investment decision-making, valuation of financial assets, and risk modelling. It is beneficial when limited information is available, or outcomes are uncertain. A fundamental equation in Bayesian decision theory is the expected utility equation, which calculates the expected value of an action. It is represented as:

$$E[U] = \sum_{i=1}^{n} P(x_i) u(x_i) \tag{3.29}$$

In 3.29, $E[U]$ represents the expected utility, $P(x_i)$ denotes the probability of outcome x_i, and $u(x_i)$ represents the utility associated with outcome x_i. By summing up the products of the probabilities and utilities for each possible outcome, we obtain the expected utility, which measures an action's overall value or desirability.

Another important equation in Bayesian decision theory is Bayes' rule, which updates the probabilities of different outcomes based on new information. It can be expressed as:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{3.30}$$

Here, $P(A|B)$ denotes the posterior probability of event A given the occurrence of event B. $P(B|A)$ represents the likelihood of event B given event A. $P(A)$ and $P(B)$ direct to the prior probabilities of events A and B, respectively. Bayes' rule allows us to incorporate new evidence or information to update our beliefs about the probabilities of different events.

Bayesian decision theory has numerous applications in finance, as already mentioned, including portfolio optimisation, option pricing, and risk management.

By utilising Bayesian methods, decision-makers can make more informed investment decisions, gain insights into the valuation of financial instruments, and effectively assess and manage financial risk. In an overview, Bayes' rule allows for quantifying probabilities and utilities, enabling better-informed decisions and improved understanding and management of financial risk.

Bayesian Model Uncertainty

To perform parameter inference in a time series dataset with inputs denoted as $x = [x_1, \ldots, x_r]$ and corresponding outputs of the regressor N denoted as $N = [N_1, \ldots, N_r]$, we can utilise Bayes' theorem as follows:

$$P(\omega|x, N) = \frac{P(N|x, \omega)P(\omega)}{P(N|x)} \tag{3.31}$$

In 3.31 the:

- $P(\omega|x, N)$ represents the posterior probability distribution of the parameters ω given the observed inputs x and regressor outputs N. It quantifies our updated knowledge about the parameters after incorporating the observed data.
- $P(N|x, \omega)$ is the likelihood function, which captures the probability of observing the regressor outputs N given the inputs x and the parameters ω. It reflects how well the parameters explain the observed data.
- $P(\omega)$ represents the prior probability distribution, which captures our initial beliefs about the parameters ω before observing any data. It provides a way to incorporate existing knowledge or assumptions about the parameter values.

The denominator $P(N|x)$ is a normalising constant to ensure the posterior distribution is a valid probability density function. To evaluate the integral in the denominator denoted as $\int P(N|x, \omega)P(\omega)d\omega$, represents the marginal likelihood of the observed data given the inputs, the marginal likelihood can rewrite as:

$$P(N|x) = \int P(N|x, \omega)P(\omega)d\omega \tag{3.32}$$

This integral sums up the likelihood of the data for each possible parameter value weighted by the prior probability distribution. It measures how well the parameters explain the observed data, considering all possible parameter values. By dividing the likelihood $P(N|x, \omega)$ by the marginal likelihood $P(N|x)$, we can obtain the normalised posterior probability distribution $P(\omega|x, N)$. Mathematically, this can be expressed as:

$$P(\omega|x, N) = \frac{P(N|x, \omega)P(\omega)}{\int P(N|x, \omega)P(\omega)d\omega} \tag{3.33}$$

The numerator $P(N|x, \omega)P(\omega)$ represents the joint probability of observing the data and the parameters. The normalised posterior distribution is obtained by dividing it by the marginal likelihood $P(N|x)$. This normalisation ensures that the posterior distribution integrates to 1 for a valid probability density function meaning that the probabilities assigned by the distribution cover all possible outcomes and follow the rules of probability. Mathematically, this condition can be expressed as:

$$\int P(\omega|x, N)d\omega = 1 \tag{3.34}$$

Furthermore, it allows us to make inferences about the parameters ω based on the observed data, accounting for our prior beliefs and the likelihood of the data given different parameter values.

The above equations provide a rigorous framework for incorporating prior beliefs, the likelihood of the data, and normalisation to estimate the posterior distribution and make informed inferences about the parameters in light of the observed data.

In computational finance, Bayesian Decision Theory can be effectively combined with Monte Carlo dropout to tackle various decision problems. MCD comes into play by generating samples from the posterior distribution and approximating the updated beliefs after incorporating the observed data. This integrated approach allows financial practitioners to make informed decisions, quantify uncertainties, and assess the expected utility of various outcomes, facilitating risk management, asset allocation, option pricing, and other financial applications. Below MCD is presented.

Monte Carlo Dropout

In MCD, introduced by Gal and Ghahramani (2016), dropout randomly sets a fraction of neurons to zero during training, effectively removing them from the network. This process creates a set of different models, each with a unique configuration of active neurons.

Let us consider the goal of estimating the model's predictive distribution, which represents the uncertainty in the model's predictions. The predictive distribution is $p(y|x,\theta)$, where y is the target variable, x is the input data, and θ represents the model parameters.

Each model can be viewed as a different network instantiation with a specific set of active neurons. To approximate the predictive distribution $p(y|x,\theta)$, we can treat these $f_t(x)$ values as samples from that distribution. For example, let us denote the t-th model as $f_t(x)$, set of predictions $f_1(x), f_2(x), \ldots, f_T(x)$ obtained from multiple forwards passes with dropout enabled representing the prediction made by that model. To compute the expected prediction over the ensemble of models to estimate the predictive distribution that can be expressed as:

$$\mathbb{E}[f(x)] = \int f(x)p(f|x)df, \tag{3.35}$$

where $\mathbb{E}[\cdot]$ denotes the expectation, $p(f|x)$ is the distribution of models given the input x, and df represents the measure over the space of models. In MCD, approximate this integral by sampling from the distribution of models $p(f|x)$ using dropout. Each sampled model provides a prediction $f_t(x)$ based on a particular configuration of active neurons.

By averaging the predictions from these samples, we can estimate the expected prediction:

$$\hat{y} = \frac{1}{T}\sum_{t=1}^{T} f_t(x) \tag{3.36}$$

In the above equation, sum the predictions from all T models, represented by $f_t(x)$, from $t = 1$ to T. Then, diverge the sum by the number of samples T to obtain the average prediction $\hat{y}$.

In summary, the equation $\hat{y} = \frac{1}{T}\sum_{t=1}^{T} f_t(x)$ is a valid approximation of the expected prediction in MCD. By averaging the predictions from multiple models sampled using MCD, the estimation of the predictive distribution, obtain a more accurate prediction, and assess the model's uncertainty.

One of the key advantages of employing Monte Carlo Dropout is its capability to measure model uncertainty, identify anomalous data points and facilitate more informed decision-making in critical systems. Therefore, this thesis applies the MCD in the proposed MCoRNNMCD-ANN, as presented in Chapter 4, to improve the model's performance. However, despite its potential benefits, certain limitations are associated with implementing Monte Carlo Dropout in time series prediction.

3.6 Gradient and Optimizers

The gradient measures the shift in all weights concerning the change in error. Assuming W_h is the weight of a node k in a neural network, then the update rule for adjusting the weight W_{hk} of W_{hk} is calculated as,

$$W_{hk} \leftarrow W_{hk} - n\frac{\partial \varepsilon_t}{\partial W_{hk}}, \tag{3.37}$$

The weight is updated by subtracting a learning rate n multiplied by the partial derivative of the error ε_t for the weight W_{hk}. The learning rate is a crucial part of the training process of neural networks. In this section, the most common optimizers that are used in machine and deep learning will be presented (Goodfellow, Bengio, & Courville, 2016).

Stochastic Gradient Descent (SGD)

The stochastic gradient descent algorithm avoids the need for redundant calculations as it updates the parameters for each iteration in a dataset. For this reason, convening is faster and can also be used for online learning. However, in SGD, frequent parameter updates have a hefty price dispersion, resulting in fluctuations in the loss function. This phenomenon can lead the algorithm to converge to other and potentially better local minima, in contrast to the gradient descent algorithm that always converges to the same minimum.

In this thesis, the defining mean square error function is given as

$$\mathcal{Q}(w) = \frac{1}{\mathcal{T}}\sum_{i=1}^{\mathcal{T}}(y_i - \hat{y}_i)^2 = \frac{1}{\mathcal{T}}\sum_{i=1}^{\mathcal{T}}\mathcal{Q}_i(w), \tag{3.38}$$

which calculates the average squared difference between the actual outputs y_i and the predicted outputs $\hat{y}i$ for a set of $\mathcal{T}$ training data examples. The function can be represented as the sum of individual losses $\mathcal{Q}i(w)$ divided by $\mathcal{T}$.

The gradient descent update rule for adjusting a model's parameter vector w, calculated as,

$$w_{t+1} = w_t - n_t \nabla \mathcal{Q}(w_t). \tag{3.39}$$

The parameter vector is updated by subtracting a learning rate n_t multiplied by the gradient $\nabla \mathcal{Q}(w_t)$ of the loss function $\mathcal{Q}$ concerning the parameters at time t.

The full gradient of the loss is calculated as

$$\nabla \mathcal{Q}(w) = \frac{1}{\mathcal{T}} \sum_{i=1}^{\mathcal{T}} \nabla \mathcal{Q}_i(w), \tag{3.40}$$

where the full gradient of the loss function $\mathcal{Q}(w)$, which is the average of the gradients $\nabla \mathcal{Q}i(w)$ of individual losses $\mathcal{Q}i(w)$ for each training example.

The stochastic gradient descent is given as,

$$\nabla \mathcal{Q}(w) = \frac{1}{|\mathbf{b}|} \sum_{v \in |\mathbf{b}|} \nabla \mathcal{Q}_i(w), \tag{3.41}$$

where the gradient $\nabla \mathcal{Q}(w)$ is calculated as the average of the gradients $\nabla \mathcal{Q}_i(w)$ of a randomly selected batch of training examples, denoted by $\mathbf{b}$.

The update rule for stochastic gradient descent is denoted as,

$$w_{t+1} = w_t - n_t \frac{1}{|\mathbf{b}|} \sum_{v \in |\mathbf{b}|} \nabla \mathcal{Q}_i(w), \tag{3.42}$$

where the parameter vector w is updated by subtracting a learning rate n_t multiplied by the average of the gradients $\nabla \mathcal{Q}_i(w)$ over a batch of training examples.

Momentum

The Gradient Descent algorithm can not deal with surfaces with sharp curves in a common dimension near local minima. The algorithm oscillates between the slopes in these cases, making little progress towards the local optima. Momentum is solving this problem, as it is a method that accelerates convergence in one direction and, at the same time, reduces oscillation, calculated as,

$$z_t = \gamma z_{t-1} + n \nabla \mathcal{Q}(w), \tag{3.43}$$

$$w = w - z_t \tag{3.44}$$

. The variable z_t represents the update vector that accumulates the previous update multiplied by a momentum term γ, and the current gradient $\nabla \mathcal{Q}(w)$ multiplied by the learning rate n. The parameter vector w is then updated by subtracting the update vector z_t.

Adaptive Gradient Algorithm (Adagrad)

Adagrad algorithm is a gradient-based optimisation algorithm that alters the learning rate for each parameter, and its use is suitable for sparse data. Adagrad algorithm has been found to improve the robustness of the Gradient Descent algorithm. Let $\mathcal{G}t, i$ represent the partial derivative (gradient) of the objective function $\mathcal{Q}i(w)$ concerning the parameter w at time step t. This derivative represents the rate of change of the objective function concerning the parameter calculated as

$$\mathcal{G}_{t,i} = \nabla \mathcal{Q}_i(w). \tag{3.45}$$

The update of each parameter w_i every time t is given as,

$$w_{t+1,i} = w_{t,i} - n\mathcal{G}_{t,i}, \tag{3.46}$$

The parameter is updated by subtracting the learning rate n multiplied by the gradient $\mathcal{G}_{t,i}$. The Adagrad algorithm introduces an adaptive learning rate adjustment calculated as

$$w_{t+1} = w_{t,i} - \frac{n}{\sqrt{S_{t,ii} + E}} \mathcal{G}_{t,i}. \tag{3.47}$$

The learning rate n for each parameter w_i at time step t is modified based on the previous gradients calculated for w_i. The denominator $\sqrt{S_{t,ii} + E}$ is a scaling factor that normalises the learning rate. S_t is a diagonal matrix of size $D \times D$ where each diagonal element $S_{t,ii}$ represents the sum of the squares of the gradients w_i up to time step t. The term E is a small constant (smoothing term) added to avoid division by zero. The above equation can be reduced to a product of arrays, calculated as

$$w_{t+1} = w_t - \frac{n}{\sqrt{S_t + E}} \odot \mathcal{G}_t. \tag{3.48}$$

Moreover, it represents the update of the parameter vector w as a whole rather than updating individual parameters separately. The $\odot$ symbol denotes element-wise multiplication. Here, the learning rate adjustment is applied to the entire parameter vector by element-wise multiplying it with the normalised gradient vector $\mathcal{G}_t$.

Adadelta

The Adadelta algorithm is an elongation of the Adagrad that fixes the monotonous reduction in the learning rate. Instead of accumulating the summary of all the previous squares of the gradients, the Adadelta algorithm limits the summary to a window of length w. For a window w, the algorithm retrospectively calculates the summary of the gradients as an average of all previous gradients. Let a time step t the average of the gradients is $\mathcal{E}$, depending by a fraction γ on the average of the previous ones, and by the same degree it has in the specific t, calculated as,

$$\mathcal{E}[\mathcal{G}^2]_t = \gamma\mathcal{E}[\mathcal{G}^2]_{t-1} + (1-\gamma)\mathcal{G}_t^2, \tag{3.49}$$

and for the update vector Δw_t calculated as,

$$\Delta w_t = -n \odot \mathcal{G}_t, \tag{3.50}$$

$$w_{t+1} = w_t + \Delta w_t. \tag{3.51}$$

By placing the updated vector Δwt is calculated by multiplying the learning rate n with the gradient $\mathcal{G}t$ divided by the square root of the average squared gradients plus a small constant E, we have, respectively,

$$\Delta w_t = -\frac{n}{\sqrt{S_t + E}} \odot \mathcal{G}_t. \tag{3.52}$$

The parameter vector w is updated by adding the update vector Δwt denoted as,

$$\Delta w_t = -\frac{n}{\sqrt{\mathcal{E}[\mathcal{G}^2]_t + E}}\mathcal{G}_t. \tag{3.53}$$

The variable $\mathcal{E}[\mathcal{G}^2]t$ represents the exponentially decaying average of the squared gradients.

RMSProp

RMSProp is similar to Adadelta and is the same as the first update vector of the Adadelta algorithm. By setting γ to a similar value as the momentum term, around 0.9, we have

$$\mathcal{E}[\mathcal{G}^2]_t = 0.9\mathcal{E}[\mathcal{G}^2]_{t-1} + 0.1\mathcal{G}_t^2, \tag{3.54}$$

where, $\mathcal{E}[\mathcal{G}^2]t$ represents the exponentially weighted moving average (EWMA) of the squared gradients $\mathcal{G}^2t$. The EWMA is calculated by taking a weighted sum of the previous average $\mathcal{E}[\mathcal{G}^2]t-1$ (weighted by 0.9) and the squared gradient $\mathcal{G}^2t$ (weighted by 0.1). This average estimates the second moment (variance) of the gradients. The update rule for the parameter vector w at time step t is calculated as,

$$w_{t+1} = w_t - \frac{n}{\sqrt{\mathcal{E}[\mathcal{G}^2]_t + E}}\mathcal{G}_t. \tag{3.55}$$

The parameter is updated by subtracting the learning rate n multiplied by the gradient $\mathcal{G}t$ divided by the square root of the average squared gradients $\mathcal{E}[\mathcal{G}^2]t$ plus a small constant E.

Adaptive Moment Estimation (Adam)

Adam similarly RMSprop, and Adadelta store the average of previous square gradients. The average of the gradients and the average squares of the gradients derived from the previous iterations are given by the equations 3.56, and 3.57:

$$z_t = \beta_1 z_{t-1} + (1 - \beta_1)\mathcal{G}_t, \tag{3.56}$$

where z_t represents the exponentially decaying average (first moment) of the gradients $\mathcal{G}t$. It is similar to the momentum term and is calculated by taking a weighted sum of the previous average z_{t-1} (weighted by β_1) and the current gradient $\mathcal{G}t$ (weighted by $1 - \beta 1$). This average captures the tendency or direction of the gradients.

$$q_t = \beta_2 q_{t-1} + (1 - \beta_2)\mathcal{G}_t^2, \tag{3.57}$$

where q_t represents the exponentially decaying average (second moment) of the squared gradients $\mathcal{G}_t^2$. It is calculated similarly to z_t but with the squared gradients instead. This average provides an estimate of the variance of the gradients.

The bias-corrected estimate $\widehat{z}t$ of the first moment by dividing zt by the bias-correction factor $1 - \beta_1^t$. The bias-correction factor adjusts the estimate to account for the initialisation bias in the first few time steps, calculated as,

$$\widehat{z}_t = \frac{z_t}{1 - \beta_1^t}, \tag{3.58}$$

The bias-corrected estimate $\widehat{q}t$ of the second moment by dividing qt by the bias-correction factor $1 - \beta_2^t$. Similar to equation 3.58, this correction factor accounts for the initialisation bias in the first few time steps, calculated as,

$$\widehat{q}_t = \frac{q_t}{1 - \beta_2^t}. \tag{3.59}$$

The parameters β_1 and β_2 control the decay rates for the first and second moments, respectively. They are typically set to values close to 1, such as 0.9 and 0.999, to emphasise recent and squared gradients more.

The update rule for the parameter vector w at time step $t.$, is given as,

$$w_{t+1} = w_t - n\frac{\widehat{z}_t}{\sqrt{\widehat{q}_t} + E}. \tag{3.60}$$

The parameter is updated by subtracting the learning rate n multiplied by the bias-corrected average gradient $\widehat{z}t$ divided by the square root of the bias-corrected average squared gradient $\widehat{q}t$ plus a small constant E. The division by the square root of $\widehat{q}_t$ acts as an adaptive learning rate based on the computation of the variance of the gradients.

This chapter has provided a comprehensive exploration of the computational properties of neural networks by examining their fundamental computational properties, including their architecture. This foundation allowed us to grasp the core principles that enable neural networks to process and transform input data and learn from examples. Furthermore, the computational properties of Monte Carlo dropout and orthogonal weight initialisation methods were examined. These techniques have shown promise in enhancing the performance and reliability of neural network models in various domains. By incorporating these methods, we could improve prediction accuracy, quantify uncertainty, and optimise the proposed modelling process. Additionally, different optimisation algorithms are utilised in training neural networks. In the next chapter, the proposed model is introduced.

4. Proposed Novel Bio-inspired Model Architecture For Forex Market Predictions

This study proposes a novel bio-inspired Modular Convolutional orthogonal Recurrent MCD–ANN (MCoRNNMCD–ANN), aiming to encounter the limitations of the current monolithic architectures presented in the literature. The proposed modular network incorporates a new CNN architecture to address catastrophic forgetting, overfitting, vanishing and exploding gradient problems, and underspecification (Alzubaidi et al., 2021). Therefore, a proposed new CNN architecture incorporates a modular topology inspired by Tzilivaki et al. (2019), formulating a convolutional, orthogonal recurrent MCD replacing the pooling layers, followed by dense layers flattening their outputs. Compared with a typical CNN time series composed of convolutional, pooling, flattened, and dense layers, the proposed new CNN could enhance the robustness and forecasting performance of the Forex market (Aryal et al., 2019).

Consequently, in the proposed MCoRNNMCD–ANN, the modules selected from baselines (Tables 2.2 and 2.3) displayed better results in partitioning the input domain in anticipating EUR/GBP price movements. Hence, two separate and parallel features extraction convolutional with orthogonal kernel initialisation applied in simple RNN and a GRU coupled with MCD networks, receiving the closing prices and sentiment scores capture long-term dependencies in the EUR/GBP exchange rate hourly, replacing the pooling layers were considered. The replacement occurs to avoid the downsampling of feature sequences by losing valuable information since the pooling layers capture only the essential features in the data and ignore the less important ones, which can be vital (Liu, Ji, & Wang, 2020). The dense layers are also placed before the flattening operation in both modules in the proposed novel CNN architecture. This adaptation transpires because the dense layer's preliminary purpose is to increase the model's capacity to learn more complex patterns from the RNN's output. The flatten operation is then applied to reshape the result of the dense layer for each module into a one-dimensional tensor to prepare it for the combined outputs by integrating them into a final concatenation layer. Ultimately, the concatenated features passed in the final decision module consist of a three-layer feed-forward ANN that yields the anticipated hourly closing price of EUR/GBP. Finally, the acquired knowledge from the proposed MCoRNNMCD-ANN from predicting price fluctuations of the EUR/GBP is partially transferred. The partially transferred learning strived to foster a modular ANN coupled with MCD to possibly achieve better outcomes, receiving less data for predicting the EUR/USD exchange rate. Figure 4.1 displays the proposed MCoRNN-MCD–ANN model.

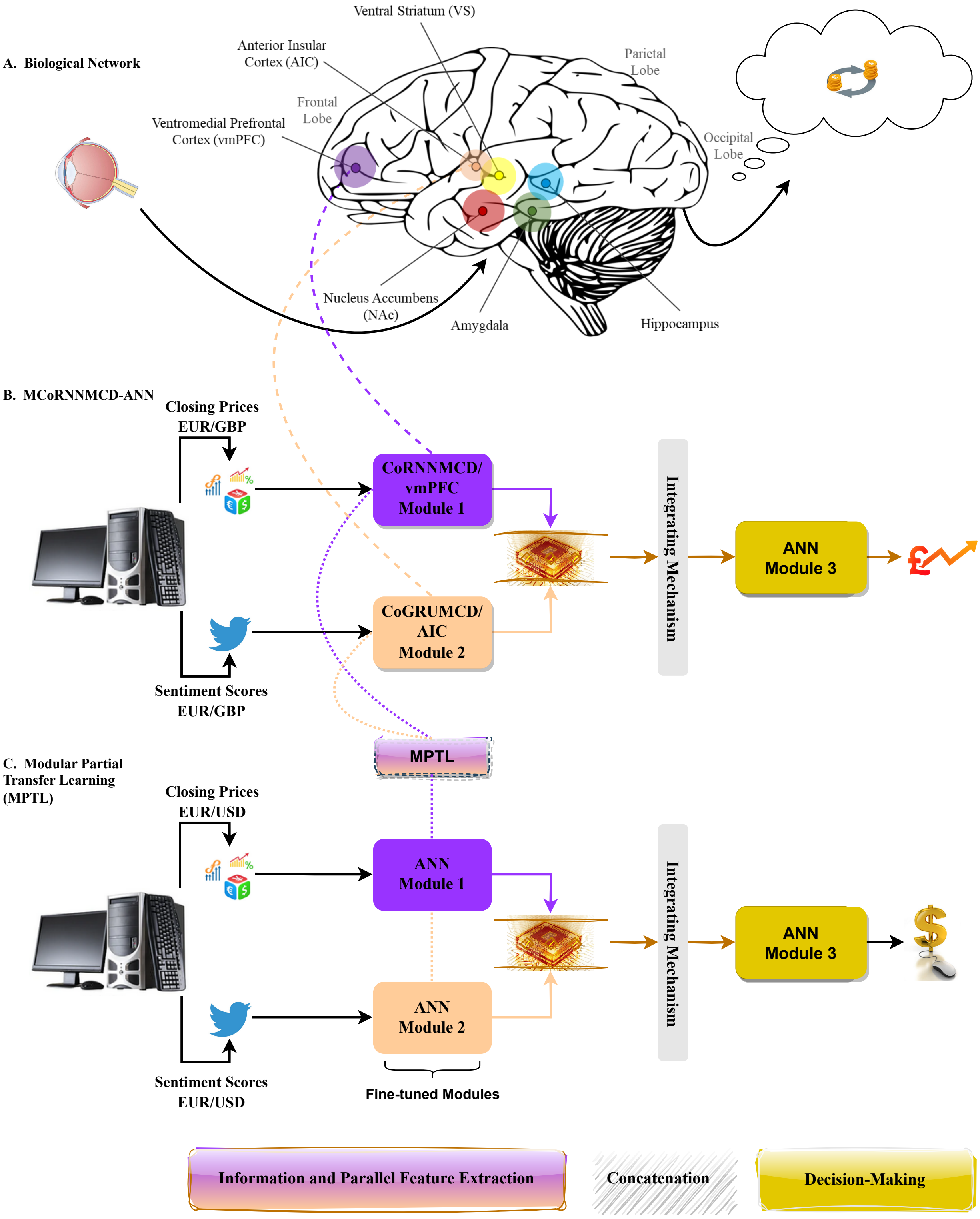


Figure 4.1: Proposed MCoRNNMCD-ANN. Part A) visualises the brain areas that impact investors' decision-making. Part B) illustrates the representation of the brain areas from the proposed MCoRNNMCD-ANN model. Finally, in part C), the generated knowledge of MCoRNNMCD-ANN is partially transferred to a new task, aiming to fine-tune a modular ANN.

4.1 Module 1: Convolutional orthogonal RNN-MCD (CoRNNMCD)

Let us consider time series data representing the hourly closing prices of the EUR/GBP currency pair. The input data can be characterized as a matrix $x_c \in \mathbb{R}^{dc \times l}$, where dc is the number of channels (in this case, one for a single currency pair) and l is the time series length for the one-hour time frame applied in this study. As already referred, the hourly rate is one of the best intraday time frames for price anticipation (Almasri & Arslan, 2018). However, in other cases, the l variable can take whatever value, such as one day, week, or month, in a time series prognosis.

In the initial CNN, the convolution operation can be mathematically represented as:

$$y[i] = f((x_c * w)[i] + b) \tag{4.1}$$

Here, x_c represents the input data, w denotes the filters or kernels, b is the bias term, and f is the activation function. The dot product operation $(x_c * w)[i]$ is performed between the filter w and the portion of the input data x_c the filter is currently "looking at". The activation function f is then applied element-wise to the result of the dot product, adding non-linearity to the output.

Moving on to Module 1, a 1D convolutional layer is applied to the input data. The convolution operation is performed using a set of filters or kernels, denoted as $w \in \mathbb{R}^{dc \times r}$, where r is the size of the filter. Mathematically, the convolution between a 1D filter w of size r and an input signal of length l can be defined as:

$$(x_c * w)[i] = \sum_{j=1}^{r} x_c[i+j]w[j] \tag{4.2}$$

Here, $*$ denotes the convolution operation, and i ranges from 1 to $(l - r + 1)$ to ensure the filter fits entirely within the input signal. The variable j ranges from 1 to r and represents the position within the filter and the corresponding elements in length l input signal.

After the convolution, the activation function is applied element-wise to each element of the convolution result, adding non-linearity. Next, the convolution operation generates a new feature representation, denoted as $W \in \mathbb{R}^{(l-r+1) \times mc}$, where mc is the number of filters. The output feature map c of the 1D convolutional layer is defined as the input to the RNN, which directly replaces the max pooling layer. The feature map c is represented as a matrix W, where each row corresponds to a window vector $w_n = [x_n, x_{n+1}, \ldots, x_{n+r-1}]$.

To feed the window W into an RNN, the hidden state is computed as $h_t \in \mathbb{R}^{mh}$ where mh represents the dimension of the hidden state in the recurrent network at each time step t. The hidden state h_t in the equation of the simple RNN is calculated as:

$$h_t = \phi(W_x h x_{(c)_t} + W_{hh} h_{t-1} + b_h) \tag{4.3}$$

Here, $x_{(c)_t} \in \mathbb{R}^{mc}$ represents the input at time step t, $W_{xh} \in \mathbb{R}^{mc \times mh}$ and $W_{hh} \in \mathbb{R}^{mh \times mh}$ are weight matrices, h_{t-1} is the previous hidden state, and $b_h \in \mathbb{R}^{mh}$ is a bias term. The non-linear activation function ϕ, such as the Rectified Linear Units (ReLU), is applied element-wise to each hidden state h_t (Nair & Hinton, 2010).

Replacing the max pooling layer with an RNN allows capturing sequential dependencies in the time series data. In addition, the RNN considers the temporal information and improves the model's performance in predicting future values. After the RNN layer, a dense layer can be added to generate the network's output. The dense layer takes the hidden state h_t as input and applies the following equation:

$$y_{(c)} = f(W_y h_t + b) \tag{4.4}$$

Here, W_y is the weight matrix, b is the bias term, and f is the activation function, such as softmax, which converts the output into a likelihood distribution over the possible classes. Finally, a flattened layer takes the output of the dense layer as input, computed as:

$$F_c[i] = \text{flatten}(y_{(c)}[i]) \tag{4.5}$$

It is worth mentioning that the dense layer after the RNN can allow the model to learn complex relationships and mappings between the input and the desired output while flattening the outputs of the dense can simplify the data structure by collapsing the dimensions, making it compatible with following layers that expect one-dimensional inputs.

The backpropagation technique (BPTT) is utilized to train an RNN. However, RNNs require help to learn long-term dependencies during the BPTT training process since the gradients employed to update the weights increase exponentially, a procedure is known as the vanishing or exploding gradient problem.

In this thesis, orthogonal initialisation is considered one of the proper mechanisms to address the vanishing gradient issue in the RNNs (Golmohammadi et al., 2017). Therefore the kernel weights W will be transformed into O (section 3.4). Furthermore, the parametric rectified linear unit (PReLU) activation function is utilised instead of the tahn activation function since it is considered one of the keys to deep networks' recent success in time series analysis (Dong, Wang, & Guo, 2018). Furthermore, PReLU is denoted as:

$$f(x_i) = \begin{cases} x_i, & \text{if } x_i > 0 \\ a_i x_i, & \text{if } x_i \leq 0 \end{cases} \tag{4.6}$$

where $f(x_i)$ represents the output of the PReLU activation function for the input x_i, a_i a learnable parameter associated with the i-th unit of the PReLU. It controls the slope of the negative part of the function. When x_i is less than or equal to zero, the output becomes $a_i \cdot x_i$, where a_i is a non-negative constant. When x_i is greater than zero, the output remains x_i. In other words, the PReLU function allows for negative values in the output by introducing the learnable parameter a_i. If a_i is set to 0, the function reduces to the regular ReLU activation.

Finally, to potentially enhance the performance of the orthogonal kernel initialized RNN receiving the outputs of the 1D-convolution closing price for EUR/GBP as inputs, the MCD is coupled to the oRNN layer (CoRNNMCD) in its ability to quantify model uncertainty, facilitating more informed decision making in Forex forecast (Gal & Ghahramani, 2015). The hidden state h_t in equation 4.3 is updated and computed as:

$$h_t = \text{PReLU}((Ox_h x_{(c)_t} + W_{hh} h_{t-1} + b_h) \odot \text{MCD}) \tag{4.7}$$

The output of the CoRNNMCD is fed to the dense layer and computed as:

$$y_{(c)_t} = \text{linear}(W_{hy} h_t + b_y) \tag{4.8}$$

where h_t is the hidden state at time t, x_t is the input at time t, W_{hh}, W_{hy} are weight matrices, b_h, and b_y are the bias vectors, O is an orthogonal matrix used to initialize the input weights, $\odot$ represents an element-wise multiplication, and MCD is the Monte Carlo Dropout.

Finally, a flattened layer receives as an input the output of the dense layer indicated as:

$$F_c[t] = \text{flatten}(y_{(c)}[t]) \tag{4.9}$$

Figure 4.2 illustrates the CoRNNMCD. CoRNNMCD pseudocode can be described as follows:

Algorithm 5 CoRNNMCD

1: **procedure** CoRNNMCD
2: initialize h_0
3: $O \leftarrow$ InitializeOrthogonalMatrix()
4: **for** each time step t **do**
5: $h_t = \text{PReLU}((Ox_h x_{(c)_t} + W_{hh}h_{t-1} + b_h) \odot \text{MCD})$
6: $y_{(c)_t} = \text{linear}(W_{hy}h_t + b_y)$
7: $F_c[t] = \text{flatten}(y_{(c)}[t])$
8: **end for**
9: **end procedure**

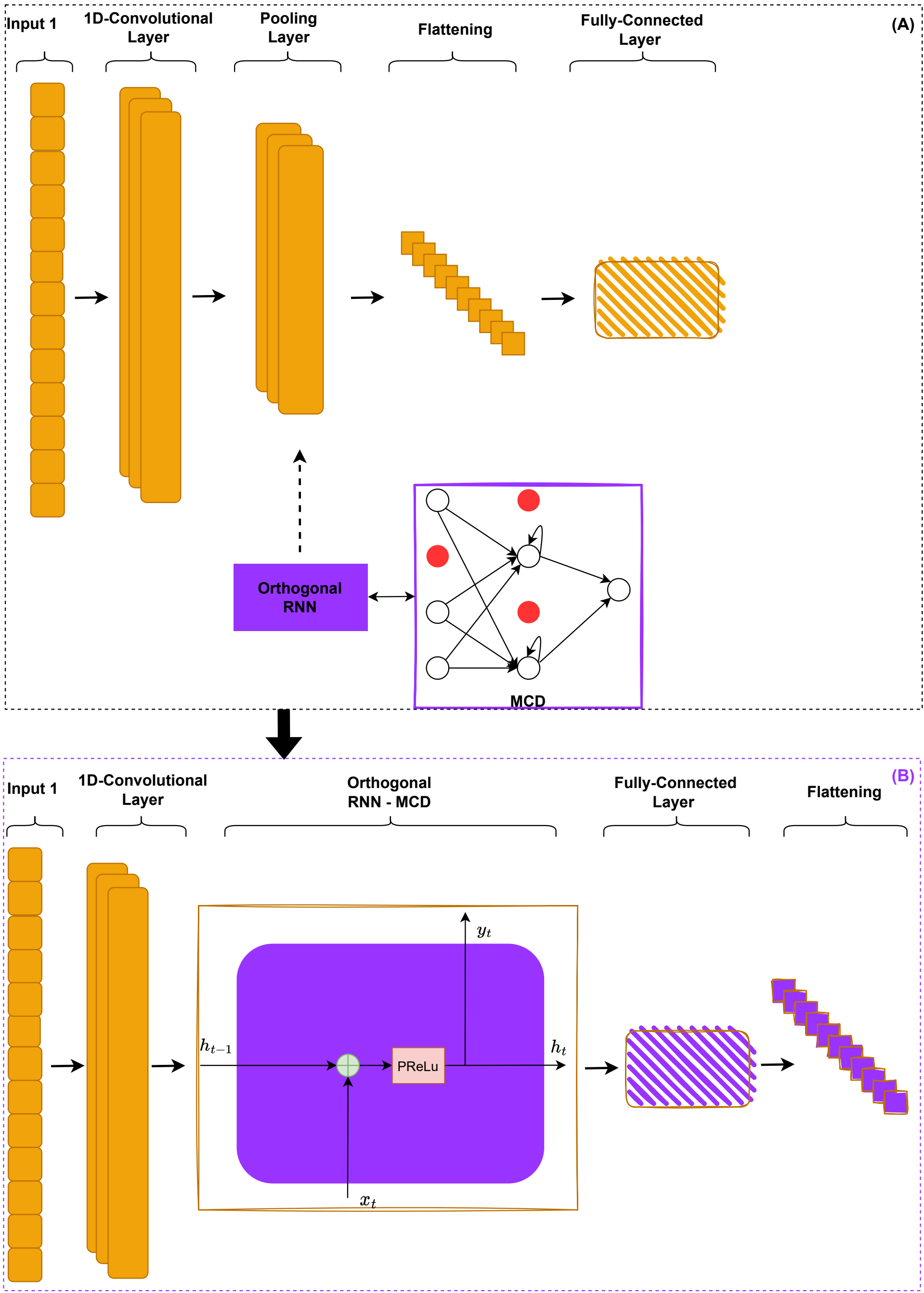


Figure 4.2: (A) Initial CNN Architecture (B) Orthogonal RNN-MCD Architecture. Replacement of Max-Pooling Layer by the Orthogonal RNN-MCD.

4.2 Module 2: Convolutional orthogonal GRU-MCD (CoGRUMCD)

Module 2 uses a 1D convolutional layer for sentiment analysis on a time-series task. Nevertheless, first, let us summarize the key components and equations: Given the input data $x_s \in \mathbb{R}^{d_s \times l}$, where d_s is the number of channels (1 in this matter), and l is the hourly length of the time series utilized in this study. x_s represents the input data at each time step t. The window w_n is formed by selecting r consecutive sentiment scores starting from the n-th timestamp, expressed as $w_n = [x_n, x_{n+1}, \ldots, x_{n+r-1}]$. The 1D convolutional layer processes the window w_n to extract convolutional features. The output of the convolutional layer, denoted as $s \in \mathbb{R}^{(l-r+1)\times m_s}$, consists of m_s feature maps. The parameter m_s determines the number of feature maps representing the filters used in the convolutional layer. The convolutional features in s are new window representations, capturing different patterns or representations in the input time series. The output feature maps in s are then fed into a GRU computed as:

$$r_t = \sigma(W_r[h_{t-1}, x_{(s)_t}] + b_r) \tag{4.10}$$

$$z_t = \sigma(W_z[h_{t-1}, x_{(s)_t}] + b_z) \tag{4.11}$$

$$\tilde{h}_t = \tanh(W_h[r_t \odot h_{t-1}, x_{(s)_t}] + b_h) \tag{4.12}$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \tag{4.13}$$

where h_t is the hidden state at time t, $x_{(s)_t}$ is the intake at time t, r_t and z_t are the reset and update gates, respectively. $\tilde{h}_t$ is the candidate hidden state, W_r, W_z, W_h are weight matrices, b_r, b_z, and b_h are the bias vectors, and $\odot$ represents an element-wise multiplication.

The 1D-convolutional orthogonal kernel initialized GRU coupled with MCD (CoGRUMCD) updates the GRU equations 4.10, 4.11, 4.12, 4.13 to incorporate the convolutional features and learn temporal dependencies in the sentiment scores as follows:

$$r_t = \sigma(O_r x_{(s)_t} + W_r h_{t-1} + b_r) \tag{4.14}$$

$$z_t = \sigma(O_z x_{(s)_t} + W_z h_{t-1} + b_z) \tag{4.15}$$

$$\tilde{h}_t = \text{PReLU}(O_h x_{(s)_t} + W_h(r_t \odot h_{t-1}) + b_h) \odot \text{MCD} \tag{4.16}$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \tag{4.17}$$

where r_t is the reset gate at time step t, z_t is the update gate at time step t, $\tilde{h}_t$ is the candidate value for the new hidden state at time step t, h_t is the new hidden state at time step t, O_r, O_z, O_h are orthogonal matrices used to initialize the input weights, W_r, W_z, W_h are weight matrices, b_r, b_z, and b_h are the bias terms, and $\odot$ represents an element-wise multiplication.
The output of the CoGRUMCD is fed to the dense layer denoted as

$$y_{(s)_t} = \text{linear}(W_{hy} h_t + b_y) \tag{4.18}$$

Finally, a flattened layer receives as an input the output of the dense layer computed as,

$$F_s[t] = \text{flatten}(y_{(s)}[t]) \tag{4.19}$$

Figure **??** illustrates the CoGRUMCD. CoGRUMCD pseudocode can be given as follows:

Algorithm 6 CoGRUMCD

```
procedure CoGRUMCD
    initialize h_0
    O_r ← InitializeOrthogonalMatrix()
    O_z ← InitializeOrthogonalMatrix()
    O_h ← InitializeOrthogonalMatrix()
    for each time step t do
        r_t = σ(O_r x_(s)_t + W_r h_{t-1} + b_r)
        z_t = σ(O_z x_(s)_t + W_z h_{t-1} + b_z)
        h̃_t = PReLU(O_h x_(s)_t + W_h(r_t ⊙ h_{t-1}) + b_h) ⊙ MCD
        h_t = (1 − z_t) ⊙ h_{t-1} + z_t ⊙ h̃_t
        y_(s)_t = linear(W_hy h_t + b_y)
        F_s[t] = flatten(y_(s)[t])
    end for
end procedure
```

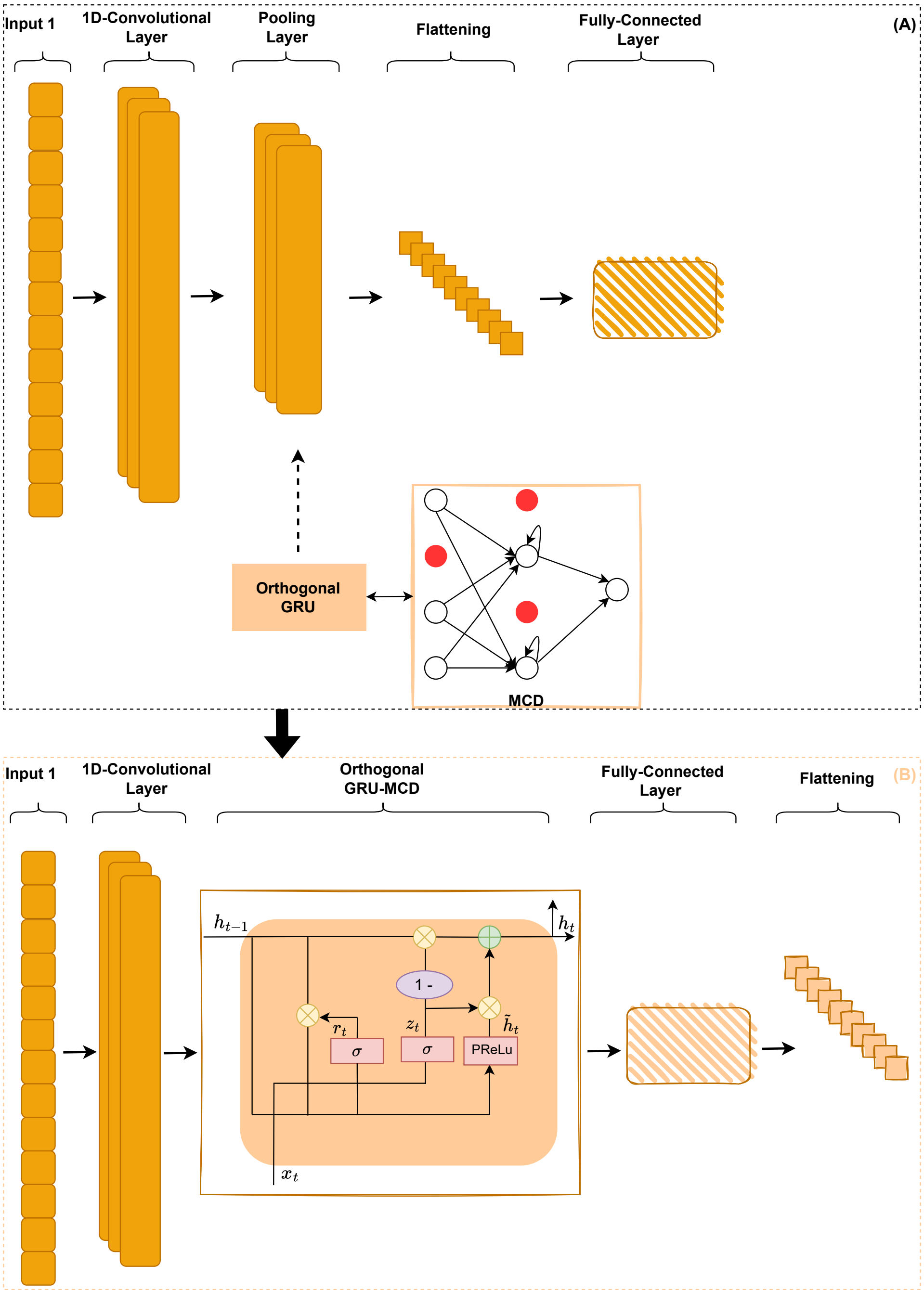


Figure 4.3: (A) Initial CNN Architecture (B) Orthogonal GRU-MCD Architecture. Replacement of Max-Pooling Layer by the Orthogonal GRU-MCD.

4.3 Parallel Feature Extraction and Concatenation

The parallel features extraction operation converges the two modules' tasks. The two modules are continuous, with hourly time frames. Let $M1$ be the first module (CoRNNMCD) with input feature vector x_c and output vector y_c flattened as $F_c[t]$. Let $M2$ be the second module (CoGRUMCD) with input feature vector x_s and output vector y_s flattened as $F_s[t]$.

The parallel processing operation can be represented as:

$$y_{1,2} = [M1(x_c), M2(x_s)] \tag{4.20}$$

The outputs from the parallel processing operating system, module 1 (M1) and module 2 (M2), that receive the closing price and the sentiment scores are merged in the concatenation layer and used as an integrating mechanism. The conjunct outputs are connected to the final module of the proposed MCoRN-NMCD–ANN model, aiming to yield the anticipated closing price for the EUR/GBP rate.

The information that merged in the concatenation layer is calculated as:

$$\mathcal{C} = M1 \cup M2 \tag{4.21}$$

4.4 Module 3: Decision Making

The final part of the proposed MCoRNNMCD–ANN model takes place to make the decision consisting of a three-layer feed-forward ANN. The first layer of the ANN receives the merged information and can be denoted as:

$$Dense_{d1} = \text{ReLU}(W_{Dense_{d1}}\mathcal{C} + b_{Dense_{d1}}). \tag{4.22}$$

In the second dense layer, a proposed altered version of the Swish activation function, namely HSwish$_{alt}$, is applied. The main difference between the Swish activation function and the HSwish$_{alt}$ function is that it utilized the hard sigmoid instead of the sigmoid in Swish (Courbariaux, Bengio, & David, 2015; Ramachandran, Zoph, & Le, 2017). Moreover, HSwish$_{alt}$ utilized a different scaling factor, as calculated below:

$$\text{Swish}(x, \beta = 1) = x \cdot \text{sigmoid}(\beta x) \tag{4.23}$$

The proposed HSwish$_{alt}$ function utilized $\beta = 0.5$ computed as:

$$\text{HSwish}_{alt}(x, \beta = 0.5) = x \cdot \text{hardsigmoid}(0.5x) \tag{4.24}$$

The rationale behind HSwish$_{alt}$ is to mitigate the issue of exaggerated responses to minor fluctuations in the input. In financial markets like Forex, where prices can exhibit high volatility and noisy fluctuations, prediction models need robustness and stability. Using HSwish$_{alt}$ with $\beta = 0.5$, the model can potentially introduce a dampening effect on the negative inputs, resulting in smoother and more controlled responses. This dampening effect could also be beneficial in scenarios where the model has to avoid harsh reactions to minor input fluctuations.

Hence, the second dense layer receives the output from the first dense layer, estimated as:

$$Dense_{d2} = \text{HSwish}_{alt}(W_{Dense_{d2}} Dense_{d1} + b_{Dense_{d2}}). \tag{4.25}$$

The final output layer is denoted as:

$$Dense_{d3} = \text{linear}(W_{Dense_{d3}} Dense_{d2} + b_{Dense_{d3}}). \tag{4.26}$$

ANN pseudocode can be given as follows:

Algorithm 7 ANN

procedure ANN($\mathcal{C}, W1, b1, W2, b2, W3, b3$)
  Dense$_{d1}$_output $\leftarrow$ ReLU($\mathcal{C} \cdot W1 + b1$)
  Dense$_{d2}$_output $\leftarrow$ HSwish$_{alt}$(Dense$_{d1} \cdot W2 + b2$)
  Dense$_{d3}$ $\leftarrow$ linear(Dense$_{d2} \cdot W3 + b3$)
  return output
end procedure

The final computational form of the proposed MCoRNNMCD-ANN model is presented below:

Algorithm 8 Proposed MCoRNNMCD-ANN

procedure MCORNNMCD-ANN(c, s)
  $cconv, sconv \leftarrow C1(c), C2(s)$
  $ORNN \leftarrow InitializeOrthogonalRNN()$
  $OGRU \leftarrow InitializeOrthogonalGRU()$
  for each timestep t **do**
    $crnn[t] \leftarrow ORNN(cconv[t])$
    $sgru[t] \leftarrow OGRU(sconv[t])$
    $crnn\_drop, sgru\_drop \leftarrow MonteCarloDropout(crnn, sgru)$
  end for
  $yrnn \leftarrow Dense(Flatten(crnn\_drop))$
  $ygru \leftarrow Dense(Flatten(sgru\_drop))$
  $yconcat \leftarrow Concatenate(yrnn, ygru)$
  $ANN \leftarrow InitializeANN(3)$
  $\hat{y} \leftarrow ANN(yconcat)$
  return $\hat{y}$
end procedure

4.5 Modular Partial Transferring Learning

A Modular Partial Transfer Learning (MPTL) method has been employed in this thesis to enhance the performance of a modular ANN coupled with MCD, denoted as MANNMCD, consisting of two modules in a similar task. This enhancement is accomplished by leveraging previously acquired proposed MCoRNNMCD-ANN knowledge from predicting price fluctuations of the EUR/GBP currency pair. This approach's primary goal is to enable MANNMCD to achieve better outcomes with less data for predicting the EUR/USD exchange rate.

From the proposed MCoRNNMCD-ANN model, only the $CoRNNMCD$ (module 1) and the $CoGRUMCD$ (module 2) will be transferred to two separate ANN modules consisting of 3-layers coupled with MCD, referred to as $ANNMCD1$ and $ANNMCD2$, for fine-tuning.

The equations for the $CoRNNMCD$ and $CoGRUMCD$ models are presented below:

$$h_t^{\text{CoRNNMCD}} = f_{\text{CoRNNMCD}}(h_{t-1}^{\text{CoRNNMCD}}, x_t; \theta_{\text{CoRNNMCD}})$$

$$h_t^{\text{CoGRUMCD}} = f_{\text{CoGRUMCD}}(h_{t-1}^{\text{CoGRUMCD}}, x_t; \theta_{\text{CoGRUMCD}})$$

Here, h_t^{CoRNNMCD} and h_t^{CoGRUMCD} represent the hidden states of the $CoRNNMCD$ and $CoGRUMCD$ models at time t, respectively. x_t is the input at time t, and θ_{CoRNNMCD} and θ_{CoGRUMCD} are the parameters of the respective models.

The output of the $CoRNNMCD$ and $CoGRUMCD$ models is utilised for training the $ANNMCD1$ and $ANNMCD2$ models using the Adam optimiser. The following equations describe the fine-tuning process:

$$\theta_{\text{ANNMCD1}} = \text{Adam}(L_{\text{target}}(\theta_{\text{ANNMCD1}}))$$

$$\theta_{\text{ANNMCD2}} = \text{Adam}(L_{\text{target}}(\theta_{\text{ANNMCD2}}))$$

Here, $L_{\text{target}}(\theta_{\text{ANNMCD1}})$ and $L_{\text{target}}(\theta_{\text{ANNMCD2}})$ represent the loss functions specific to the target task, and θ_{ANNMCD1} and θ_{ANNMCD2} are the parameters of $ANNMCD1$ and $ANNMCD2$, respectively.

The outputs of $ANNMCD1$ and $ANNMCD2$ are combined to obtain a combined output (Combined_Output) consisting of the inputs of the final decision module consisting of a 4-layer ANN. Refining the final module through training from scratch on target data could offer a distinct advantage by enabling the model to adjust intricately to the unique characteristics of the target domain. This approach may allow the neural network to develop tailored decision boundaries and capture specific patterns prevalent in the target data. By doing so, the model could become finely tuned to the intricacies of the task at hand, potentially enhancing its performance and generalisation capabilities within the specific context of the target domain. The Combined_Output is obtained by concatenating the outputs of $ANNMCD1$ and $ANNMCD2$. Subsequently, the combined output is processed by ANN layers using the following equations:

$$\text{Dense}_{d1} = \text{ReLU}(W1 \cdot \text{Combined\_Output} + b1)$$

$$\text{Dense}_{d2} = \text{HSwish}_{\text{alt}}(W2 \cdot \text{Dense}_{d1} + b2)$$

Finally, the output of the decision module is obtained by applying a linear transformation to Dense_{d2}:

$$y_t = \text{linear}(W3 \cdot \text{Dense}_{d2} + b3)$$

In the above equations, $W1$, $W2$, and $W3$ denote the weight matrices for the connections between layers in the decision module, and $b1$, $b2$, and $b3$ represent the bias vectors for the respective layers.

The Dense_{d1} and Dense_{d2} layers allow for additional non-linear transformations and feature extraction from the combined output of $ANNMCD1$ and $ANNMCD2$. Additionally, these layers provide the flexibility to capture complex relationships and patterns in the data, ultimately leading to the generation of the final output y_t.

It is crucial to note that the distinct activation functions, the number of hidden layers, and the architecture of the decision module can vary depending on the precise conditions of the task and the design choices made by the researchers.

The proposed MCoRNNMCD-ANN framework comprehensively leverages partial transfer learning to potentially improve the $MANNMCD$ performance by incorporating knowledge from pre-trained $CoRNNMCD$ and $CoGRUMCD$ models. Furthermore, the fine-tuning process aims to allow the model to adapt to the target task while benefiting from the acquired knowledge, ultimately in an effort to improve the EUR/USD exchange rate predictions.

The pseudocode for the MPTL for MANNMCD-ANN is given as follows:

Algorithm 9 Modular Partial Transfer Learning for MANNMCD-ANN

procedure TRAINMCORNNMCD-ANN
  Initialize $CoRNNMCD$ and $CoGRUMCD$ models
  Pretrain $CoRNNMCD$ and $CoGRUMCD$ models on EUR/GBP data
  Initialize $ANN1$ and $ANN2$ models
  Fine-tuning Phase:
  for $epoch \leftarrow 1$ to N **do**
    Get outputs h_t^{CoRNNMCD} from $CoRNNMCD$ model
    Get outputs h_t^{CoGRUMCD} from $CoGRUMCD$ model
    Update $ANNMCD1$: $\theta_{\text{ANNMCD1}} \leftarrow \text{Adam}(\theta_{\text{ANNMCD1}}, L_{\text{target}}(\theta_{\text{ANNMCD1}}))$
    Update $ANNMCD2$: $\theta_{\text{ANNMCD2}} \leftarrow \text{Adam}(\theta_{\text{ANNMCD2}}, L_{\text{target}}(\theta_{\text{ANNMCD2}}))$
  end for
  Decision Module:
  Get output Combined_Output by concatenating $ANNMCD1$ and $ANNMCD2$ outputs
  Calculate $\text{Dense}_{d1} = \text{ReLU}(W1 \cdot \text{Combined\_Output} + b1)$
  Calculate $\text{Dense}_{d2} = \text{HSwish}_{\text{alt}}(W2 \cdot \text{Dense}_{d1} + b2)$
  Calculate final output: $y_t = \text{linear}(W3 \cdot \text{Dense}_{d2} + b3)$
  Return: y_t
end procedure

4.6 Discussion

This chapter introduced a pioneering bio-inspired framework, termed Modular Convolutional orthogonal Recurrent MCD–ANN (MCoRNNMCD–ANN), designed to overcome the limitations of prevailing monolithic architectures in forecasting hourly price movements in the Forex market. Comprising two modules, CoRNNMCD and CoGRUMCD, this novel architecture replaces pooling layers to retain comprehensive information crucial for accurate price prediction, starkly contrasting pooling methods that discard specific data. The new adaptative mechanism consisted of Monte Carlo dropout and orthogonal kernel initialisation, incorporating it into recurrent layers within a convolutional modular network. It aims to enhance forecasting performance and execution time to minimise loss errors compared to benchmarks outlined in Chapter 2.

Critical theoretical points included the strategic placement of dense layers after the layers of the RNNs, where dense layers introduce learnable parameters facilitating the model's ability to discern intricate relationships between information and desired output. Thus, flattening operations simplify data structures, aligning them with subsequent layers expecting one-dimensional inputs.

The chapter also introduced the MPTL, a technique leveraging knowledge from the proposed MCoRNNMCD-ANN predicting EUR/GBP price fluctuations to enhance a modular ANN coupled with MCD denoted as MANNMCDANN. The CoRNNMCD and CoGRUMCD modules are transferred to the initial two layers of the modular ANN, followed by fine-tuning, showcasing the model's adaptability across the anticipation of EUR/USD price fluctuations, receiving fewer data.

Crucially, the decision-making of the modular ANN does not utilise prior knowledge from the decision module of the proposed MCoRNNMCD–ANN, emphasising the significance of training the decision module from scratch on target data. This approach may align the model more closely with decision boundaries and patterns specific to the target domain, fostering improved generalisation.

The following chapter presents information on data collection and the parametrisation of the proposed MCoRNNMCD–ANN. It delineates a comprehensive design comparison between the proposed MCoRNNMCD–ANN model and benchmarks. Moreover, it compares 2 modular ANNs as presented above, with and without transfer learning receiving less data in a new task with an extensive comparison between these two models. Ultimately, results are discussed, providing a thorough understanding of the proposed MCoRNNMCD–ANN, state-of-the-art hybrid, ensemble, single, and MPTL models' efficacy in an endeavour to predict hourly closing prices of EUR/GBP and EUR/USD currency pairs.

5. Modelling and Forecasting

This chapter provides a comprehensive overview of the data collection process, the proposed modular neural network model setup, the comparative analysis, and the results obtained from the experiments. Firstly, the data collection is described, including the selection of relevant Forex market data sources and the preprocessing steps implemented to ensure the quality and consistency of the dataset. The chapter then proceeds to discuss the comparative analysis conducted, where the performance of the modular neural network model is compared against existing monolithic architectures and other relevant benchmark models. Next, the evaluation metrics used to assess prediction accuracy, such as mean absolute error and root mean square error, are explained, and the significance of the results is discussed. Finally, the chapter concludes by presenting the results obtained from the experiments, showcasing the nimbleness and capabilities of the proposed model in predicting Forex price fluctuations. The findings highlight the superiority of the proposed modular approach over traditional monolithic architectures.

5.1 Data Collection

The EUR/GBP, exchange rate data consist of the closing price values and sentiment information retrieved from Yahoo Finance API and Twitter Streaming API, respectively. The EUR/GBP, exchange rate prices were acquired from the Finance Yahoo API based on an hourly rate. Therefore, the predicted hourly intraday trading of the closing price EUR/GBP rate is the defined target from January 2018 to December 2019 for 12,436 hours. However, because the Forex prices incorporate missing values, Twitter's sentiment data utilises the same hourly timeframe to be aligned with the pricing data. Therefore, the temporal data was partitioned according to the time intervals hourly in which they were collected.

Furthermore, a feature-level fusion based on the same hourly timeframe is considered to achieve the data coalition from both APIs. Finally, after the data fusion, the EUR/GBP exchange rate closing prices and the sentiment scores are provided for each module of the MCoRNNMCD-ANN model. Similarly, the EUR/USD exchange rate closing prices and the sentiment data have been collected and followed the same process above to be retrieved from Yahoo

Finance API and Twitter Streaming API, respectively, from January 2020 to December 2020.

5.1.1 Forex Closing Prices

The Forex closing prices of the EUR/GBP rate generated from the Yahoo Finance API from January 2018 to December 2019 are 12,436 h. The data includes the open, high, low, and close values. Only the closing price is taken into consideration as considered the most helpful indicator to foresee Forex markets (C.-C. Chen et al., 2015). One hour is deemed most suitable for better anticipating financial markets because it is shorter than daily or yearly forecasting (Almasri & Arslan, 2018). Finally, it is worth mentioning that research on data requirements for predicting time series using ANNs revealed that utilising data of one to two years yields the highest accuracy (Walczak, 2001).

5.1.2 Sentiment Data

The Twitter Streaming API is utilised by tuning the appropriate parameters for the needs of this study. The language parameter indicates whether a user wants to receive tweets only in one or some specific languages in terms of the tweet's text. More specifically, the "language = en" parameter is specified because extracting tweets from an English text was considered more appropriate, as all existing dictionaries support the English language. Using Tweepy enables handling the profile of a user and the data collection by assessing specific keywords; this study uses hashtags such as search words = "#eurgbp", "#forexmarket", and "#forex", referring to EUR/GBP currency pairs. Each tweet is accompanied by its corresponding timestamp value during its collection from Twitter API. The timestamp values are parsed using the Pandas to extract date and time information, facilitating time-based analysis. Subsequently, the tweets are grouped into hourly intervals based on the hour of posting. Tweets with the same hour counted as one, which, in this case, aggregates the text into a single data point. This aggregation can be helpful for various types of analysis, including sentiment analysis using tools like the Valence Aware Dictionary for Sentiment Reasoning (VADER). Finally, 3,265,896 have been retrieved from January 2018 to December 2019 for 12,436 h. Following a similar process, the sentiment data for the EUR/USD has been retrieved from the Twitter Streaming API from January 2020 to December 2020, defending the hashtags of "#eurusd", "#forexmarket", and "#forex".

VADER, a rule-based sentiment analysis lexicon, is utilised to extract each sentiment score from the Twitter data (Hutto & Gilbert, 2014). VADER has yielded enormous results, considering the labelling of a tweet that outperforms even from a human factor rating. VADER delivers a compound ratio, giving the negative, positive, and neutral sentiment scores. For example, from the 3,265,896 tweets of the EUR/GBP exchange rate, VADER yielded the following results: 747,890 (22.9%) negative, 930,780 (28.5%) positive, and 1,587,226 neutral (48.6%) tweets, from January 2018 to December 2019 for 12,436 h of EUR/GBP rate.

Similarly, from the 1,864,926 tweets of the EUR/GBP exchange rate, VADER yielded the following results: 345,012 (18.5%) negative, 488,610 (26.2%) positive and 1,031,304 neutral (55.3%) tweets, from January 2020 to December 2020 for 6,190 hours of EUR/USD rate. Figure 5.1 shows how the tweets can be parsed in hourly intervals and how the concatenation of sentiment scores yielded from VADER with the Forex closing price can occur.

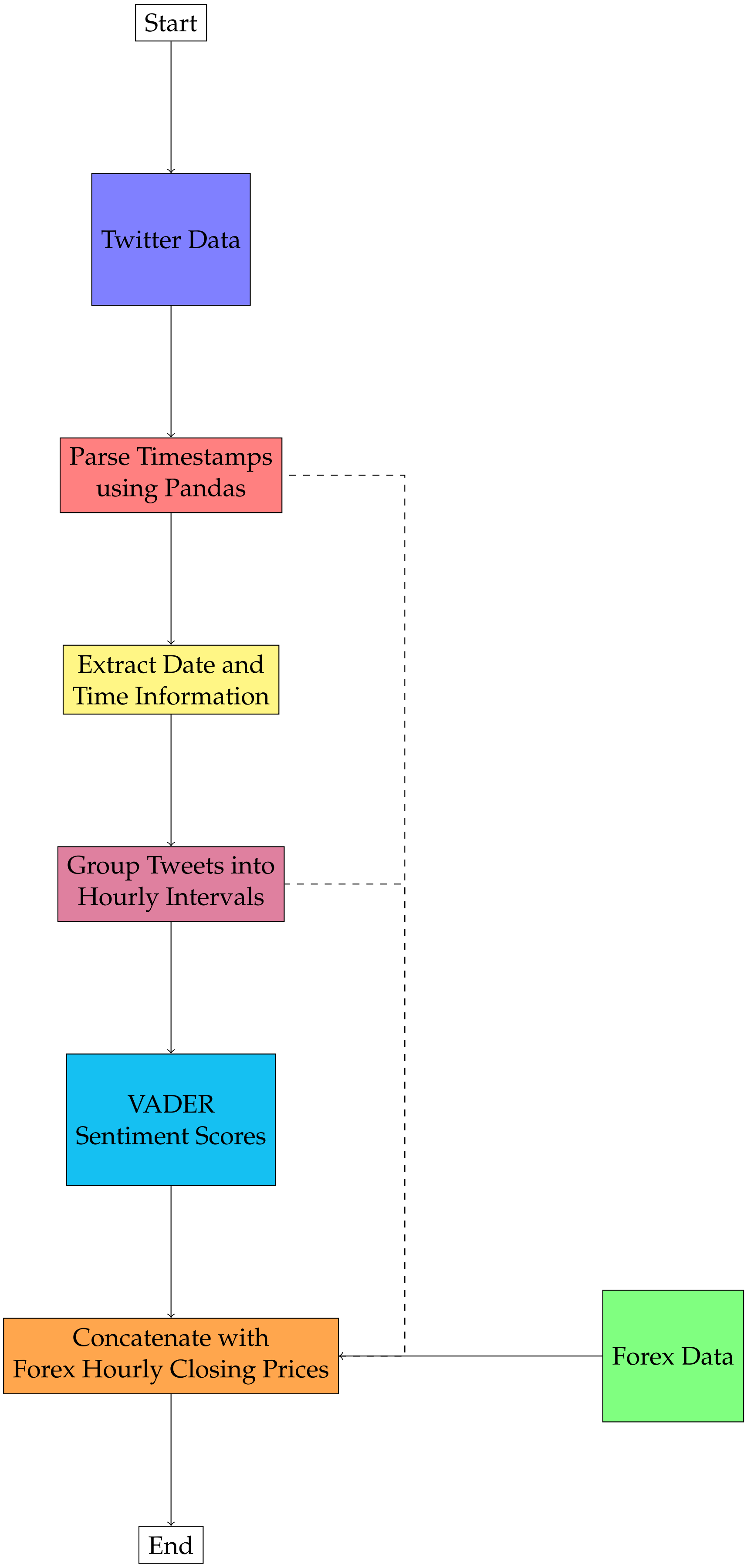


Figure 5.1: Time-Based Analysis Framework

5.2 Design and Implementation

In this thesis, to conduct the experiments, the proposed MCoRNNMCD–ANN model setting is as follows: First, each dataset has been acquired from the Yahoo Finance API and Twitter Streaming API, incorporating the hourly closing price and sentiment data, applying normalisation method, respectively. Second, the datasets are divided into training, validation, and testing sets, with the same portion of 60:20:20 used to improve the generalisability of the network. Third, the hyperparameters of the MCoRNNMCD–ANN model are encountered by employing the grid search method. Finally, the parameters below are considered to choose the most optimal for the proposed model receiving closing price and sentiment score evaluated by the MSE. The list of parameters is given below, and the best results are presented in Table 5.1 accordingly:

- Number of time steps (Lookback): 20, 30, 40, 50, 60
- Number filters per convolutional layer (Filters): 32, 64, 128, 256, 512
- Number of nodes per hidden layer (Nodes per HL): 25, 30, 50, 60, 100
- MCD rates: 10% to 50%
- Batch sizes (BS): 10, 20, 30, 60, 100

Table 5.1: Top parameters extracted from grid search in closing prices (CP) and sentiment (Sent) scores of EUR/GBP.

Model	Lookback	Filters	Nodes per HL	MCD Rate	BS	MSE
CoRNNMCD (CP)	60	128	50	0.1	20	**0.0010**
CoRNNMCD (CP)	60	30	30	0.1	20	0.0012
CoGRUMCD (Sent)	60	128	50	0.1	20	**0.2967**
CoGRUMCD (Sent)	20	64	30	0.1	20	0.3110

It is worth noting that the MSE used as an objective metric of evaluation in the grid search algorithm is evaluating the performance of its model and not its final predictions that have different calculations in the shake of hourly Forex forecasting closing price. Accordingly, grid search produces the optimal hyperparameters described:

- The lookback window uses a time step of 60. Furthermore, 128 filters are selected as the optimum numbers of the 1D convolutional layer in modules one and two, incorporating the ReLu activation function. Additionally, in the orthogonal kernel initialised RNN and GRU layers coupled with MCD with 0.1 rates, supplanting the max-pooling layer in the initial

CNN architecture, 50 neurons have been selected, utilising PReLU as the optimal activation function;

- The dense layers in modules one and two consist of 50 neurons integrating the ReLu activation function connected to the flattened layers. The decision-making ANN module consists of 3 layers receiving the merge features from modules one and two. The first dense layer also includes 50 neurons incorporating the ReLU activation function. Likewise, the second dense layer includes 50 neurons containing the HSwishalt. The output of the decision-making part, receiving one neuron selecting the linear activation function, as it is appropriate for regression tasks, yielding the predicted hourly closing price fluctuations of the EUR/GBP exchange rate;

- A batch size of 20 has been selected. The early stopping method is employed to identify the optimum number of epochs for training (Kingma, Salimans, & Welling, 2015). Early stopping has also been used in the baseline models to determine the optimum number of epochs for training. According to N. Srivastava, Hinton, Krizhevsky, Sutskever, and Salakhutdinov (2014), it is worth noting that early stopping is only sometimes utilised to combat overfitting. Laves, Ihler, Fast, Kahrs, and Ortmaier (2020) also indicated that the early stopping is not optimal for the squared error on training and testing data. The Adam optimiser with a learning rate of 0.0001 has been chosen as it proved effective for non-stationary objectives and problems with very noisy gradients, and the MSE as the loss function has been utilised during the proposed MCoRNNMCD–ANN for its training process. Each experiment of the proposed MCoRNNMCD–ANN against benchmarks has been repeated fifty times to be reliable;

- A computer with the following characteristics has been used to execute the experiments: Intel® Core™ i7-9750H (Hyper-Threading Technology), 16 GB RAM, 512 GB PCIe SSD, NVIDIA GeForce RTX 2070 8 GB. The Anaconda computational environment with Keras and TensorFlow in Python (version 3.6) programming language has been utilised to conduct the experiments.

After implementing the well-suited parameters in the proposed MCoRNNMCD–ANN model, its performance based on the MSE, MAE, and MSLE is provided in Table 5.2. Furthermore, the MCoRNNMCD–ANN outperformed all the baselines.

Table 5.2: MCoRNNMCD–ANN performance metrics in closing prices and sentiment scores of EUR/GBP.

Model	Metrics	Train	Valid	Test	R^2	Time Duration
MCoRNNMCD-ANN	MSE	6.5486×10^{-5}	6.1332×10^{-5}	5.7488×10^{-5}	0.99	2:35
	MAE	0.00534	0.00526	0.00518		
	MSLE	3.1117×10^{-5}	2.7342×10^{-5}	2.6051×10^{-5}		

5.2.1 Objective Evaluation Metrics

This thesis has applied three objective evaluation metrics, such as the MSE, MAE, and MSLE, to reveal which models performed better. The proposed MCoRNNMCD-ANN model was compared with the benchmarks and single models to predict EUR/GBP hourly price fluctuations, as presented in sections 5.2.2 and 5.2.3. Likewise, the modular partial transfer learning results are presented in section 5.2.4.

- The Mean Squared Error (MSE) is the summary of the square of the forecast error, which is the original output y minus the foreseen output $\hat{y}$, squared and divided by the total number of data points $\mathcal{T}$. It is calculated as:
$$\text{MSE} = \frac{1}{\mathcal{T}} \sum_{i=1}^{\mathcal{T}} (y - \hat{y})^2. \tag{5.1}$$

- The Mean Absolute Error (MAE) is the summary of the absolute difference between the original output y and the foreseen output $\hat{y}$, divided by the total number of data points $\mathcal{T}$. It is calculated as:
$$\text{MAE} = \frac{1}{\mathcal{T}} \sum_{i=1}^{\mathcal{T}} |y - \hat{y}|. \tag{5.2}$$

- The Mean Squared Logarithmic Error (MSLE) is the summary of the squared difference between the original values y_i and the foreseen values $\hat{y}_i$ after applying a logarithmic transformation. The sum is divided by the total number of data points $\mathcal{T}$. It is calculated as:
$$J(y, \hat{y}) = \frac{1}{\mathcal{T}} \sum_{i=0}^{\mathcal{T}} \left(\log(y_i + 1) - \log(\hat{y}_i + 1)\right)^2. \tag{5.3}$$

5.2.2 Benchmarks Models

The objective evaluation metrics of the proposed MCoRNNMCD–ANN shown in Table 5.2 revealed a decline in errors of the hybrid and ensemble models presented in Table 2.5. For instance, MCoRNNMCD–ANN decreased to 89.17%, 19.70%, 83.56%, 195.51% and 192.95% for the test MSE of the BiCuDNNLSTM, CNN–LSTM, LSTM–GRU, CLSTM and ensemble learning model. The test MAE of MCoRNNMCD–ANN decreased to 50.30%, 6.36%, 52.53%, 166.91% and 150.47% for the test MAE of the BiCuDNNLSTM, CNN–LSTM, LSTM–GRU, CLSTM and ensemble learning model. The test MSLE of MCoRNNMCD–ANN decreased to 91.20%, 20.77%, 85.28%, 195.59%, 192.99% for the test MSLE of the BiCuDNNLSTM, CNN–LSTM, LSTM–GRU, CLSTM and ensemble learning approach. The difference in time elapsed in minutes between the proposed MCoRNNMCD–ANN and the hybrid and ensemble benchmark models presented in Table 2.5 has also been considered regarding their execution time. As a result, the execution time of MCoRNNMCD–ANN decreased to 28 min for the execution time of the LSTM–GRU. The execution time of MCoRNNMCD–ANN was increased to 76, 54, 17 and 133 minutes for the execution time of BiCuDNNLSTM, CNN–LSTM, CLSTM, and ensemble model, respectively. Based on the outcomes, in most cases, the execution time of a model can be tremendously affected by the size of the window length and the complexity of the layers used in each model. It is worth mentioning that the BiCuDNNLSTM with the default parameters needs less execution time as it runs in a GPU using CUDA, which accelerates deep learning models. Finally, the LSTM–GRU takes more execution time than the proposed MCoRNNMCD–ANN, even though it utilises a default size window of 30. This effect can result from the more utilised neurons and complex architecture since it employs only LSTM and GRU models. MCoRNNMCD–ANN outperformed benchmarks.

To conduct a fairer comparison, modified versions of the hybrid benchmarks implemented the parameters from the proposed MCoRNNMCD–ANN model to investigate their performance as below:

- The modified parameters of BiCuDNNLSTM utilise a window length of 60 instead of the default 50-time steps, a convolution layer with a filter size of 128 instead of its default 64, a dropout layer with a rate of 0.1 instead of 0.2, the HSwishalt activation function in the dense layer after the flattening layer instead of the default ReLU, linear as the output activation function instead of ReLU, MSE as the loss function instead of MAE, a batch size of 20 instead of 64, and early stopping is applied instead of 32 epochs;

- The modified parameters of the CNN–LSTM neural network model are a window length of 60 instead of the default 50-time steps, a convolution layer with a filter size of 128 instead of its default 32 with a ReLU activation function instead of tanh, an LSTM layer with 50 hidden units instead of 64, and the activation function used in this layer is parametric ReLU instead of that, MSE as the loss function instead of MAE, a batch size of 20 instead of 64, and early stopping is applied instead of 100 epochs;

- The modified parameters of the LSTM–GRU neural network model are a window length of 60 instead of the default 30-time steps, LSTM and GRU layers with 50 hidden units instead of 100 with the activation function PReLU for both layers instead of a hyperbolic tangent, without the inner activations to be set as hard sigmoid functions, Adam optimiser trains the network with the learning of 0.0001 instead of the rate of 0.001, and early stopping is applied instead of 20 epochs;

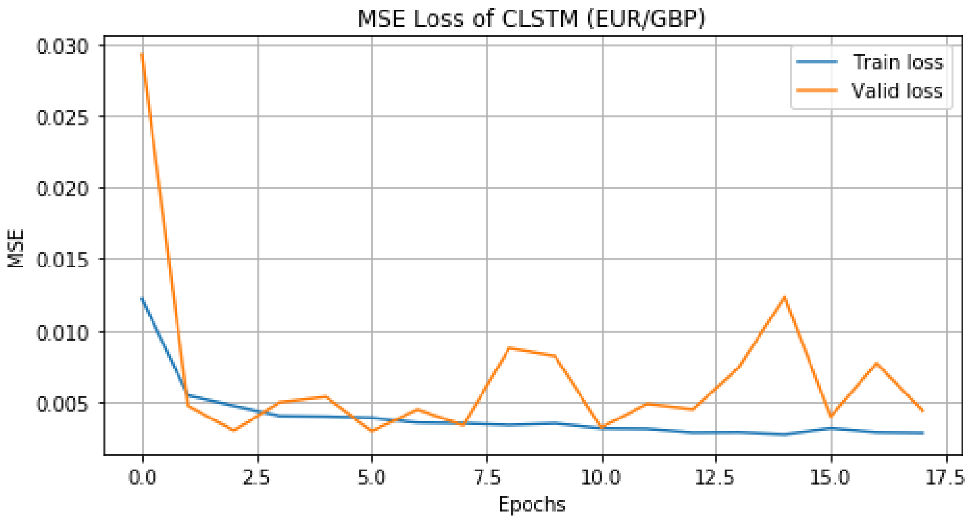
- The CLSTM model was adjusted with 128 filters in the 1D convolutional layer, 60-time steps instead of 15, and 50 neurons instead of 200, 100, and 150 neurons in the dense and LSTM layers. Moreover, LSTM employed MCD with PReLu instead of traditional dropout and ReLu activation function, applying early stopping instead of 100.

  Table 5.3 confirmed that the MCoRNNMCD–ANN outperforms the state-of-the-art hybrid benchmarks adjusting with the parameters of the MCoRNNMCD–ANN.

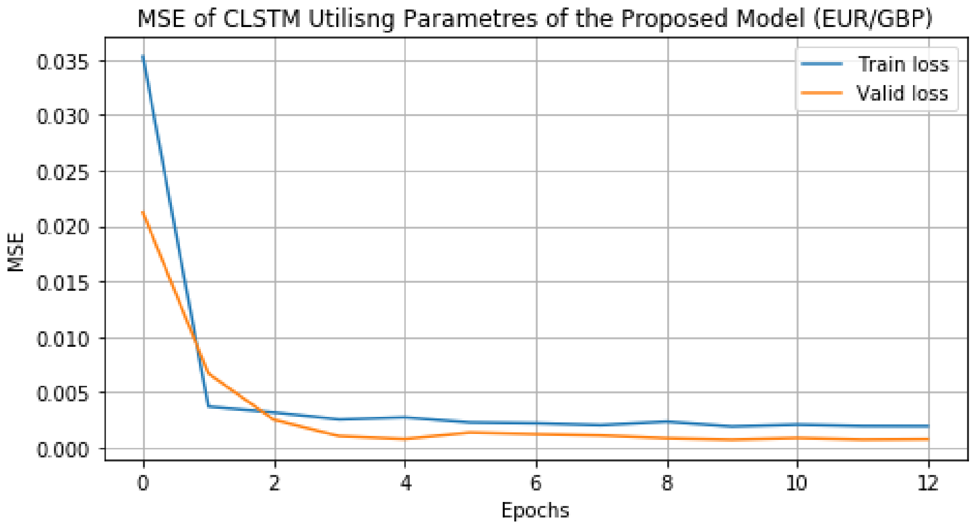
Table 5.3: MCoRNNMCD–ANN performance metrics against adjusted (adj.) hybrid benchmarks.

Model	Metrics	Train	Valid	Test	R^2	Time Duration
BiCuDNNLSTM adj.	MSE	0.00011	0.00010	0.00010	0.99	2:54
	MAE	0.00761	0.00754	0.00746		
	MSLE	5.1970×10^{-5}	4.7161×10^{-5}	4.6294×10^{-5}		
CNN-LSTM adj.	MSE	7.2831×10^{-5}	6.7592×10^{-5}	6.2947×10^{-5}	0.99	7:13
	MAE	0.00572	0.00563	0.00551		
	MSLE	3.4659×10^{-5}	3.0217×10^{-5}	2.8432×10^{-5}		
LSTM-GRU adj.	MSE	0.00011	0.00010	0.00010	0.99	20:43
	MAE	0.00739	0.00721	0.00721		
	MSLE	5.4301×10^{-5}	4.6677×10^{-5}	4.6560×10^{-5}		
CLSTM adj.	MSE	0.00145	0.00151	0.00134	0.93	8:21
	MAE	0.02352	0.02381	0.02353		
	MSLE	0.00061	0.00063	0.00059		
MCoRNNMCD-ANN	MSE	6.5486×10^{-5}	6.1332×10^{-5}	$\mathbf{5.7488 \times 10^{-5}}$	0.99	**2:35**
	MAE	0.00534	0.00526	**0.00518**		
	MSLE	3.1117×10^{-5}	2.7342×10^{-5}	$\mathbf{2.6051 \times 10^{-5}}$		

The objective evaluation metrics revealed that the test MSE of MCoRNN-MCD–ANN decreased to 53.98%, 9.10%, 53.98%, and 183.54% for the test MSE of the BiCuDNNLSTM, CNN–LSTM, LSTM–GRU, and CLSTM by adjusting their parameters with the parameters of the proposed MCoRNNMCD–ANN. Likewise, the test MAE of MCoRNNMCD–ANN decreased to 36.08%, 6.18%, 32.77%, and 127.83% for the test MAE of the adjusted BiCuDNNLSTM, CNN–LSTM, LSTM–GRU, and CLSTM. The test MSLE of MCoRNNMCD–ANN decreased to 55.96%, 8.74%, 56.49%, and 183.01% for the test MSLE of the BiCuDNNLSTM, CNN–LSTM, LSTM–GRU, and CLSTM, containing the parameters of the MCoRNNMCD–ANN. The difference in time elapsed in minutes between the MCoRNNMCD–ANN and the hybrid benchmark models adjusted with the parameters of the proposed model has shown that the execution time of MCoRNNMCD–ANN decreased to 19, 278, 1088, and 346 min for the execution time for the modified BiCuDNNLSTM, CNN– LSTM, LSTM–GRU, and CLSTM. Consequently, the execution time of hybrid benchmarks increased when the window length increased at 60-time steps incorporating the MCD when usable. That validated the previous assumption that the time steps play a remarkable role in the execution time of the models. Notably, the predictive error of the benchmarks adjusted with the proposed model parameters was reduced significantly and yielded better outcomes. Finally, the proposed MCoRNNMCD–ANN significantly outperformed the adjusted benchmarks. MCoRNNMCD–ANN was faster, validating the modular architecture and the innovative orthogonal kernel initialised RNN layers coupled with the MCD mechanism applied in the proposed model. All the models in Table 5.3 also presented a high R^2 value. Figure 5.2 illustrates an example of the MSE's tremendous improvement by utilising the parameters of the proposed MCoRNNMCD–ANN in the CLSTM and the best-performed hybrid benchmark MSE, namely CNN–LSTM adj., and the MSE of the MCoRNNMCD–ANN.

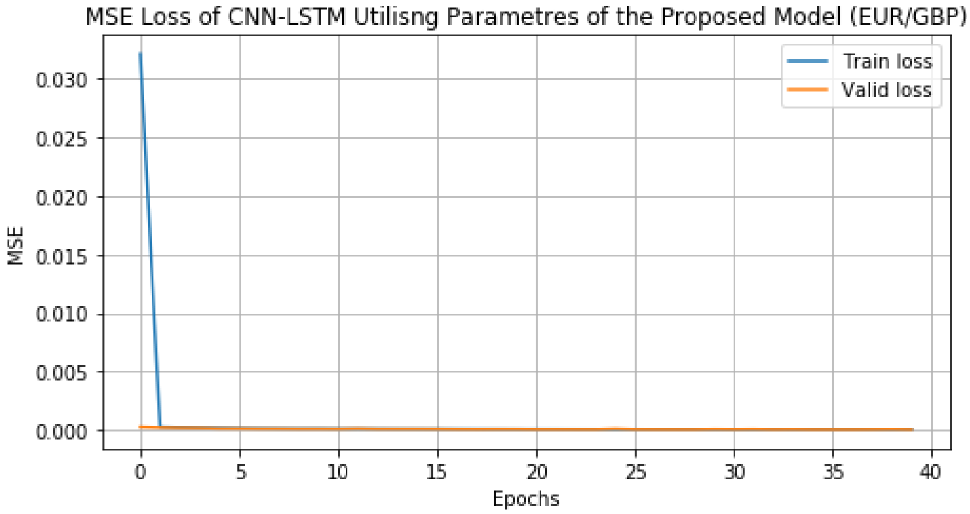


(a)

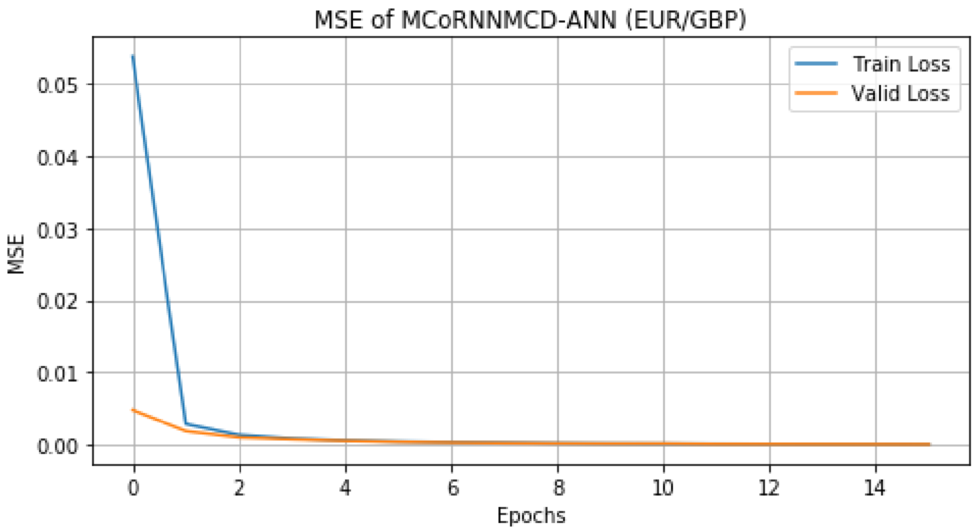


(b)

Figure 5.2: (cont.)



(c)



(d)

Figure 5.2: MSEs of CLSTMs, CNN–LSTM, and the MCoRNNMCD–ANN model: (a) CLSTM default; (b) CLSTM adjusted; (c) CNN–LSTM adjusted; and (d) proposed MCoRNNMCD–ANN.

5.2.3 Single Benchmark Models

In Table 2.6, the results of the single benchmark models have been shown. The objective evaluation metrics demonstrated that the test MSE of MCoRNNMCD–ANN decreased to 70.44%, 45.91%, and 196.90% for the test MSE of the 2D–CNN, GRU, and LSTM, respectively. The test MAE of MCoRNNMCD–ANN decreased to 38.50%, 16.96%, and 161.91% for the 2D–CNN, GRU, and LSTM test MAE, respectively. The test MSLE of MCoRNNMCD–ANN decreased to 77.81%, 54.77%, and 196.94% for the test MSLE of the 2D–CNN, GRU, and LSTM, respectively. The difference in time elapsed between the proposed MCoRNNMCD–ANN and the benchmark-single models in minutes has also been considered regarding their execution time. As a result, the execution time of MCoRNNMCD–ANN was increased to 104, 108, and 145 min for the execution time of 2D–CNN, GRU, and LSTM with default parameters. The execution time of the models can again be tremendously affected by the size of the window length and the complicatedness of each model. For instance, when the window length of the 2D–CNN, GRU, and LSTM single model incorporates a smaller time window length equal to 5-time steps, decreasing the execution time training. On the other hand, even though the LSTM is more complex than the 2D–CNN and GRU, it took significantly less time to be trained since it utilised fewer neurons (30) than GRU (50 neurons) and an Adam optimiser that can obtain a faster convergence rate leading to being faster against Adagrad for CNN (Kingma & Ba, 2017). However, LSTM has the highest predictive error. Finally, MCoRNNMCD–ANN presented a minor prediction error by significantly outperforming the single models.

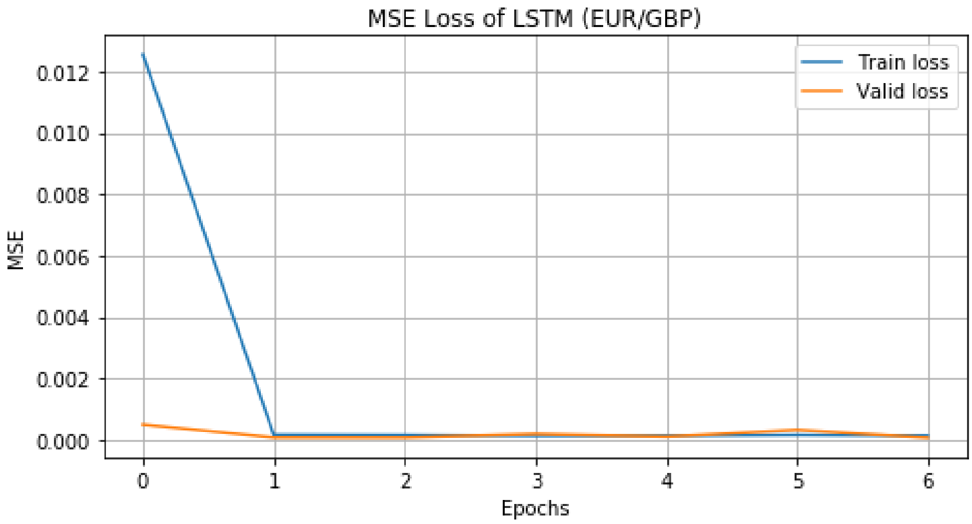
Similarly to the hybrid benchmark models, the single benchmarks (Pokhrel et al., 2022) will be adjusted with the proposed MCoRNNMCD–ANN parameters for a fairer comparison. The modified parameters of the CNN, LSTM, and GRU models are a window length of 60 instead of the default 5-time steps, a convolution layer with a filter size of 128 instead of its default 30 for CNN, an LSTM with 50 hidden units instead of 30 utilising the activation function of PReLU. Furthermore, a batch size of 20 is used. Finally, Adam is employed instead of Adagrad for GRU and CNN, while the Adam learning rate is set to 0.0001 instead of 0.1 in LSTM.

Table 5.4 proved that the error in the test performance of the MCoRNNMCD–ANN based on the evaluation metrics MSE, MAE, and MSLE was the smallest one in hourly EUR/GBP closing forecasting price outperforming the single benchmarks adjusting with the parameters of the proposed model.

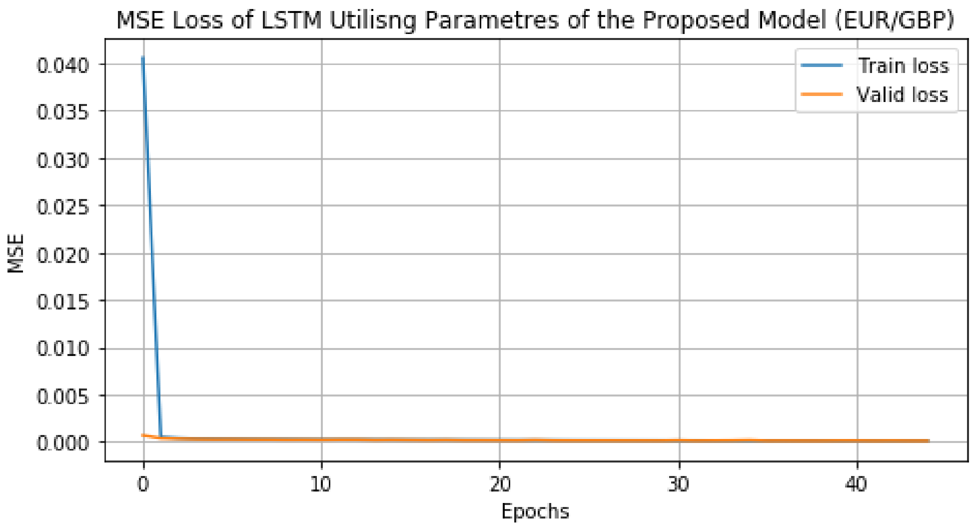
Table 5.4: MCoRNNMCD–ANN performance metrics against adjusted (adj.) single benchmarks.

Model	Metrics	Train	Valid	Test	R^2	Time Duration
2D-CNN adj.	MSE	0.00017	0.00016	0.00017	0.99	**0:39**
	MAE	0.00901	0.00897	0.00912		
	MSLE	8.3038×10^{-5}	7.3716×10^{-5}	7.6177×10^{-5}		
GRU adj.	MSE	7.3054×10^{-5}	6.6247×10^{-5}	6.2578×10^{-5}	0.99	4:54
	MAE	0.00558	0.00544	0.00536		
	MSLE	3.4814×10^{-5}	2.9724×10^{-5}	2.8312×10^{-5}		
LSTM adj.	MSE	8.7955×10^{-5}	8.0258×10^{-5}	7.6894×10^{-5}	0.99	8:43
	MAE	0.00635	0.00621	0.00612		
	MSLE	4.1941×10^{-5}	3.6171×10^{-5}	3.4807×10^{-5}		
MCoRNNMCD-ANN	MSE	6.5486×10^{-5}	6.1332×10^{-5}	$\mathbf{5.7488 \times 10^{-5}}$	0.99	2:35
	MAE	0.00534	0.00526	**0.00518**		
	MSLE	3.1117×10^{-5}	2.7342×10^{-5}	$\mathbf{2.6051 \times 10^{-5}}$		

More specifically, the test MSE of MCoRNNMCD–ANN decreased to 98.91%, 8.48%, and 28.88% for the test MSE of the 2D–CNN, GRU, and LSTM adjusted with the parameters of the proposed MCoRNNMCD–ANN. The test MAE of MCoRNNMCD–ANN decreased to 55.10%, 3.41%, and 16.63% for the test MAE of the 2D–CNN, GRU, and LSTM adjusted with the full parameters of the proposed MCoRNNMCD–ANN. The test MSLE of MCoRNNMCD–ANN decreased to 98.06%, 8.32%, and 28.77% for the test MSE of the 2D–CNN, GRU, and LSTM adjusted with the parameters of the proposed MCoRNNMCD–ANN. The difference in time elapsed between the proposed MCoRNNMCD–ANN and the benchmark-single models in minutes has also been considered regarding their execution time. As a result, the execution time of MCoRNNMCD–ANN decreased to 139 and 368 for the execution time of the GRU and LSTM modified with full parameters of the proposed MCoRNNMCD–ANN, respectively. The execution time of MCoRNNMCD–ANN was increased to 116 min for the time of 2D–CNN utilising the parameters of the proposed MCoRNNMCD–ANN. Based on the results, it has been observed that the 2D–CNN performs faster when adjusted with the proposed model parameters despite the timestep being increased to 60; this could result from the Adam optimiser that led to a faster training process instead of the Adagrad in CNN. However, the modified 2D–CNN performed worse than the default 2D–CNN. However, all benchmarks hybrid and single networks, and also those that used 1D convolutions, show significant improvement when MCoRNNMCD–ANN parameters are applied, displaying lower MSE and confirming the effectiveness of MCD and orthogonal kernel initialisation and 1D–CNNs for time-series tasks. For the adjusted LSTM and GRU with 60-time steps, execution time is increased due to retaining information from previous steps and slowing training. All the models have also presented a high R^2 value. Figure 5.3 shows substantial MSE improvement with proposed MCoRNNMCD–ANN parameters in LSTM and the best GRU–utilised MCoRNNMCD–ANN parameters and MSE of the proposed model's MSE.

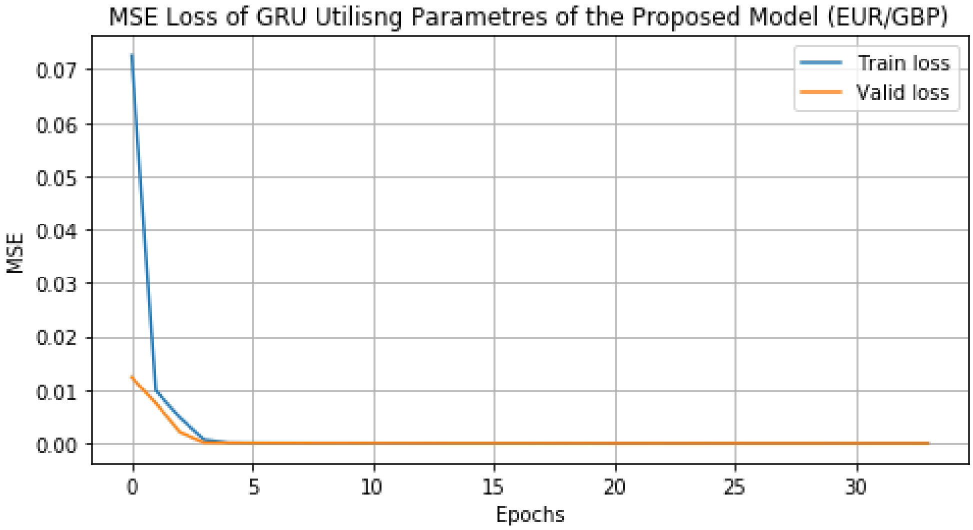


(a)

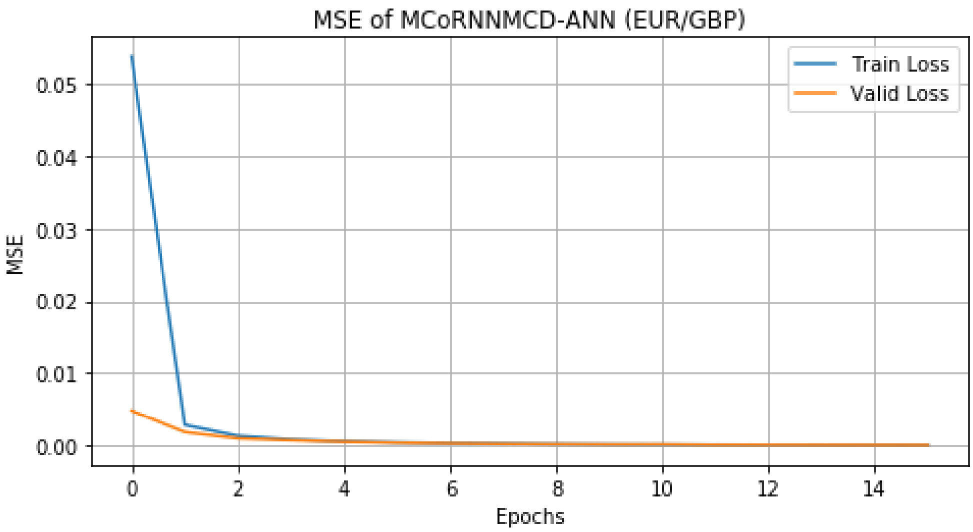


(b)

Figure 5.3: (cont.)



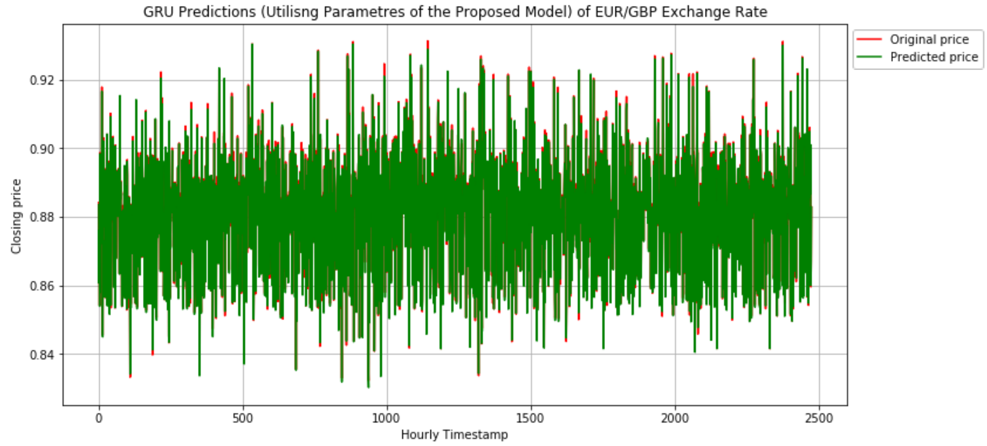
(c)



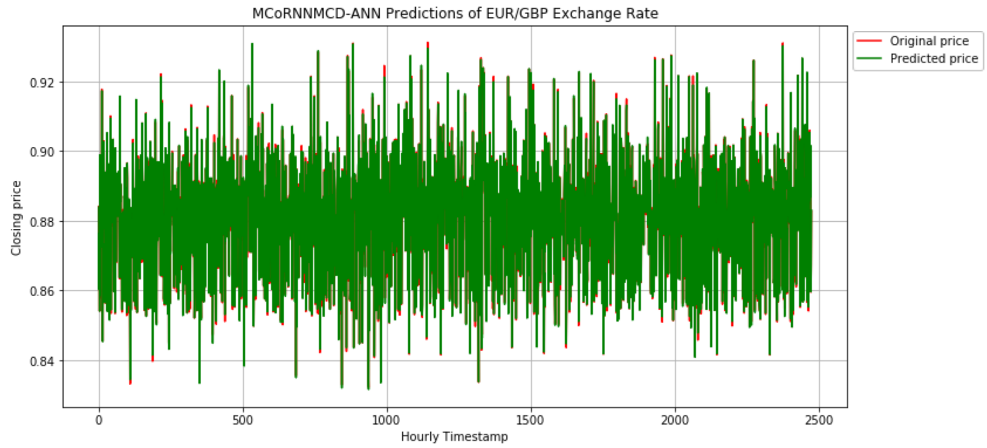
(d)

Figure 5.3: MSEs of LSTMs, GRU, and the MCoRNNMCD–ANN model: (a) LSTM default; (b) LSTM adjusted; (c) GRU adjusted; and (d) proposed MCoRNNMCD–ANN.

Concerning **RQ1** and **RQ2**, the outcomes of sections 5.2.2 and 5.2.3 further asset the concept of modularity alongside the adaptive mechanism consisting of MCD and orthogonality. Based on the experimental results, the proposed MCoRNNMCD–ANN outperformed all the hybrid, EM and single monolithic architectures. The GRU adjusted with the parameters of the proposed MCoRNNMCD-ANN was the second-best predictive model. Figure 5.4 displays the predictions in the price movement direction of the EUR/GBP rate, with the proposed MCoRNNMCD–ANN showing better performance than the adjusted GRU in the whole- and shorter-time frame. It is worth noting that the shorter period shows the first 50 hours of the EUR/GBP currency pair.

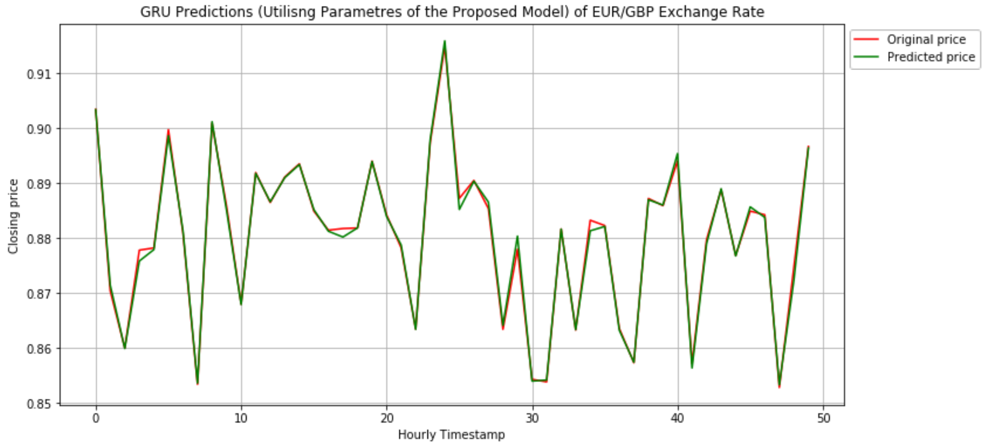


(a)

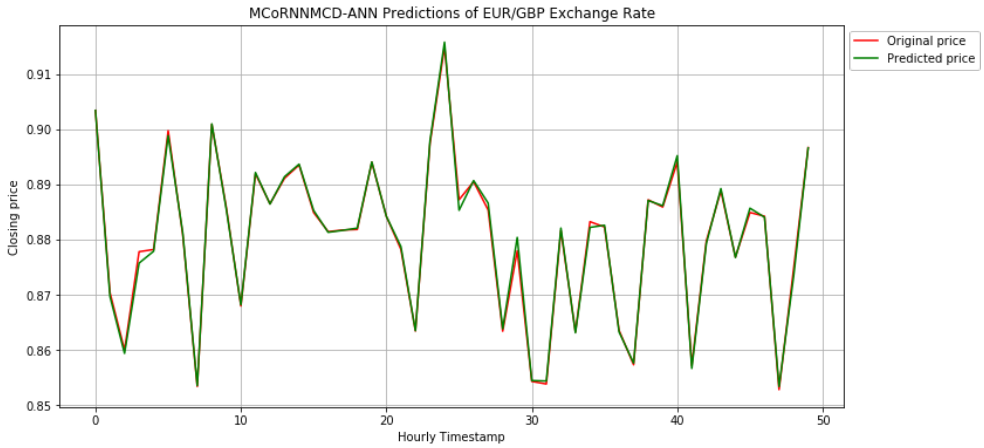


(b)

Figure 5.4: (cont.)



(c)



(d)

Figure 5.4: Predictions of GRU and the MCoRNNMCD–ANN model: (a) GRU adj. predictions whole time frame; (b) MCoRNNMCD–ANN predictions whole time frame;(c) GRU adj. predictions shorter time frame; and (d) MCoRNNMCD–ANN predictions shorter time frame.

5.2.4 Modular Partial Transfer Learning

The acquired knowledge from the proposed MCoRNNMCD-ANN model regarding the predicted hourly closing price of EUR/GBP utilised two years of data (2018-2019) has been partially transferred to fine-tuning the first two modules of modular ANN coupled with MCD in a new task. Consequently, the decision-making of the modular ANNs did not obtain prior knowledge of the decision module of the proposed MCoRNNMCD-ANN. Training the decision module of the modular ANNMCD from scratch on the target data can sustain the model better align with the decision boundaries and patterns specific to the target domain, improving generalisation. The new related task aimed to predict the hourly closing price of EUR/USD exchange rate fluctuations using only one year of data (2020). The pricing and sentiment data of 2020 were obtained and processed similarly, such as in the EUR/GBP task, using the same Yahoo Finance and Twitter Streaming APIs.

The modules of the MANNMCD consist of 1 hidden layer with 50 nodes incorporating the $HSwish_{alt}$ activation function coupled to an MCD layer with 0.1 rates and an output layer with one neuron utilising the linear activation function. The outcomes of each module are connected to the final part of an ANN without transferring the knowledge of the previous task of the final decision-making module, consisting of 2 hidden layers with 50 neurons applying the ReLu and $HSwish_{alt}$ activation functions, with the output incorporating one neuron with a linear activation function.

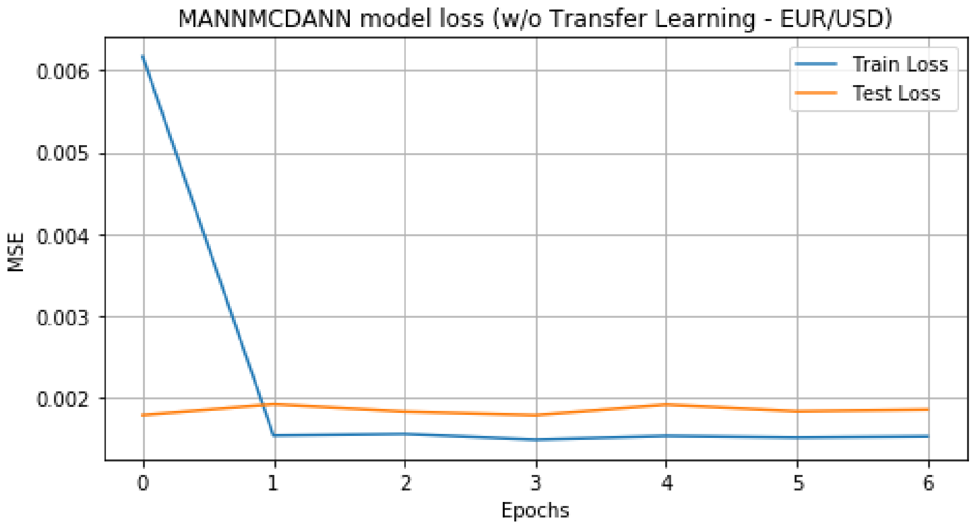
In Table 5.5, the evaluation metrics confirmed that the MANNMCDANN with PTL performs better than the MANNMCDANN without applying transfer learning (TL). Moreover, the test MSE of the MANN with TL decreased to 7.76% for the MANNMCDANN test MSE without using transfer learning. Similarly, the MAE of the MANNMCDANN with the TL model decreased to 10.2% for the MAE of the MANNMCD-ANN without TL. Finally, the test MSLE of the MANNMCDANN with TL decreased to 6.36% for the test MSLE of the MANNMCDANN without applying TL. As a result, the training time of the MANNMCDANN with TL is increased by 73 minutes. This increment in time can result from adapting the pre-trained model to the new task involving updating the modular model weights to fit the target data better. Depending on the extent of fine-tuning required, this adaptation process can be computationally expensive, leading to increased training time. However, reducing the loss error, which is this thesis's primary goal, has been achieved by improving the performance of the modular ANN that utilised the TL.

Concerning **RQ3**, the results of this section further support the notion that relevant tasks with limited data can enhance their generalisation performance with MPTL. Therefore, the modular ANNMCDANN applied transfer learning outperformed the model without being applied. Furthermore, it achieves significantly better results than a model that relies on the target domain alone for training.

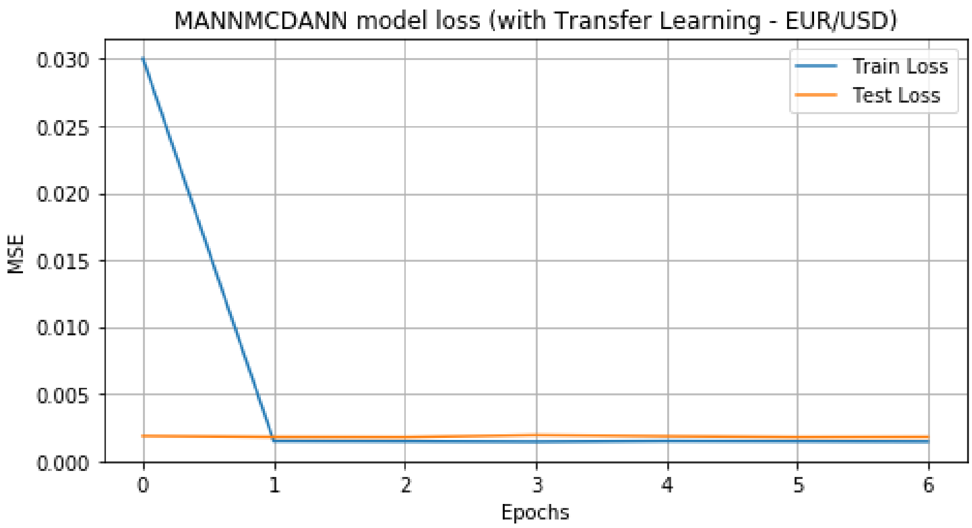
Table 5.5: MANNMCDANN model loss before and after applying partial TL.

Model	PTL	Metrics	Train	Valid	Test	Time Duration
MANNMCDANN	w/o Transfer Learning	MSE	0.00156	0.00189	0.00174	0:45
		MAE	0.01851	0.01894	0.01953	
		MSLE	0.00069	0.00083	0.00081	
MANNMCDANN	with Transfer Learning	MSE	0.00144	0.00176	**0.00161**	1:58
		MAE	0.01658	0.01711	**0.01763**	
		MSLE	0.00064	0.00078	**0.00076**	

Figure 5.5 displays the MANNMCDANN model loss before and after applying partial TL, showing a clear improvement in the model that utilised the PTL. Figure 5.6 shows the predicted closing price fluctuation with the MANNMCDANN of applying the partial transfer learning performs better than without involving the technique of partial TL in the MANNMCDANN.

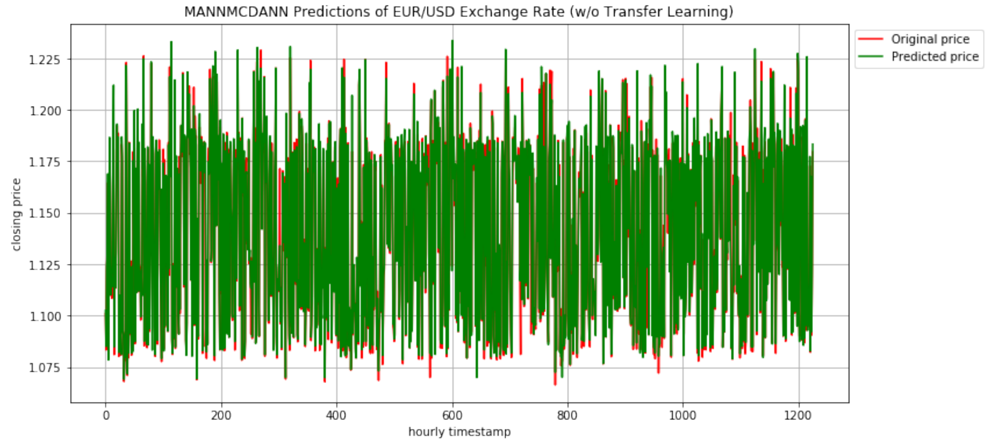


(a)

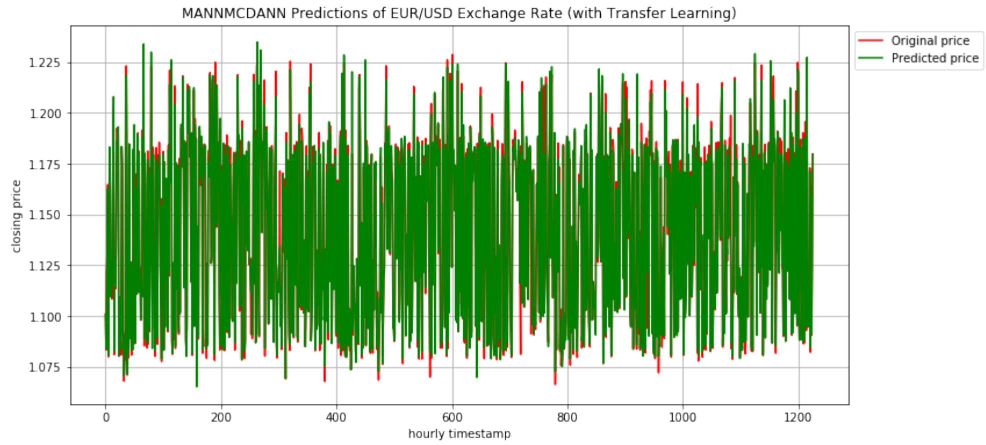


(b)

Figure 5.5: MSEs of the MANNMCDANN model: (a) without (w/o) Transfer Learning and (b) with Transfer Learning.



(a)



(b)

Figure 5.6: Predictions of the MANNMCDANN model: (a) without (w/o) Transfer Learning and (b) with Transfer Learning.

5.3 Discussion

This chapter compares the proposed MCoRNNMCD-ANN model against the benchmark models regarding the MSE, MAE, and MSLE tests in anticipating the Forex market price directions.

The results convey volumes about the exceptional performance leap achieved by MCoRNNMCD–ANN. A noticeable decrease in MSE, MAE, and MSLE across various benchmark models is evident. For example, the test MSE for BiCuD-NNLSTM, CNN–LSTM, LSTM–GRU, CLSTM, and the ensemble learning model witnessed reductions ranging from a notable 19.70% to an outstanding 195.51%, underscoring the superior predictive prowess of MCoRNNMCD–ANN. In addition, the objective evaluation metrics reveal a substantial performance enhancement achieved by MCoRNNMCD–ANN when its parameters are adjusted for hybrid models. For instance, the test MSE decreased impressively, showcasing the MCoRNNMCD–ANN model's adaptability and improved predictive accurateness across adjusted for hybrid models ranging from 9.10% to 183.54%. Also, the evaluation of single benchmark models, as indicated in the results, reveals a substantial betterment in the performance of MCoRNNMCD–ANN. For example, the objective evaluation metrics demonstrated a significant decrease in the MSE for the 2D–CNN, GRU, and LSTM ranging from 45.91% to 196.90%. Finally, accommodating MCoRNNMCD–ANN parameters for single benchmark models demonstrated significant improvement. For instance, for the test of MSE reductions across adjusted single benchmark models ranging from 8.48% to 98.91%

The exploration into time execution efficiency provided a comprehensive understanding of the adaptability the proposed MCoRNNMCD–ANN exhibited, especially compared to hybrid models. The hybrid with their default parameters, BiCuDNNLSTM, experienced a lower execution time of 76 minutes, while CNN–LSTM, CLSTM, and the ensemble model incurred 54, 17, and 133 minutes, respectively. For the specific case of the LSTM–GRU architecture, the execution time of MCoRNNMCD–ANN demonstrated a significant reduction, reaching a mere 28 minutes. These findings underscore the substantial impact of window length and layer complexities on execution time dynamics. Particularly noteworthy is the efficient performance of BiCuDNNLSTM with default parameters, designed for GPU acceleration using CUDA, resulting in lower execution times.

Conversely, the LSTM–GRU, despite utilising a default window size of 30, required more time due to heightened neuron utilisation and the intricate architecture associated with models relying solely on LSTM and GRU components. Deeper insights emerge from comparing MCoRNNMCD–ANN with hybrid benchmark models adjusted to the proposed MCoRNNMCD–ANN parameters. The proposed MCoRNNMCD–ANN reduction in execution times ranged from 19 to 1088 minutes for the modified BiCuDNNLSTM, CNN–LSTM, LSTM–GRU, and CLSTM configurations. This outcome underscores the critical role of window length, especially at 60-time steps, validating the assumption that time steps significantly influence model execution time. Hence, the state-of-the-art hybrids adjusted models that used the proposed MCoRNNMCD–ANN parameters proved that the MCoRNNMCD-ANN was significantly faster against them.

The review of time efficiency extends to comparing the proposed MCoRNNMCD–ANN and single benchmark models, shedding light on the nuanced intricacies of their execution times. The execution time of MCoRNNMCD–ANN witnessed increments to 104, 108, and 145 minutes for the default parameter configurations of 2D–CNN, GRU, and LSTM, respectively. This observed increase in execution time again underlines the sensitivity of single models to window length and complexity. Notably, when the window length of the 2D–CNN, GRU, and LSTM single models with their default parameters stands at a smaller time window of 5-time steps, a notable decrease in execution time during training is discernible. This result emphasises the impact of window length on training efficiency, offering a trade-off between temporal scope and computational demands. This thorough examination of time execution efficiency not only elucidates the nuanced influence of window length and layer complexities but also establishes MCoRNNMCD–ANN as a faster and more efficient alternative.

In the context of modified single models, parameter adjustments yielded substantial improvements in the models' efficiency. For instance, when proposed MCoRNNMCD–ANN parameters were applied to the GRU, its execution time notably decreased to 139 minutes, underscoring the MCoRNNMCD–ANN model's adaptability to optimisation. Similarly, when the MCoRNNMCD–ANN parameters counted to LSTM, its time execution was reduced to 368 minutes, emphasising the proposed model's parameters' efficiency in handling the complexities associated with long short-term memory networks. The execution time of 2D–CNN increased to 116 minutes with adjusted parameters of the proposed model, suggesting that, despite optimisation efforts, the 2D–CNN model prompts a nuanced consideration of the trade-off between model complexity and computational efficiency. While the 2D–CNN exhibited increased execution time with adjusted parameters, the overall efficiency gains for MCoRNN-MCD–ANN reaffirm its effectiveness in time-series tasks.

MCoRNNMCD–ANN's theoretical ingenuity shines through its fusion of modular convolutional orthogonal kernel-initialised RNN layers with the Monte Carlo dropout mechanism. This amalgamation sets the proposed MCoRNN-MCD–ANN apart as an innovative departure from conventional computational approaches. The proposed model consistently outperformed hybrid and single benchmarks, confirming its theoretical robustness in capturing intricate temporal patterns. Eventually, opportunities for exploration lie in adapting the proposed model architecture for different time step lengths and evaluating its performance across various financial time series. The remarkable reduction in predictive errors for the adjusted benchmarks further strengthens the case for the efficacy of the proposed model parameters. MCoRNNMCD–ANN consistently outperformed all benchmarks, highlighting the robustness of its modular architecture and innovative orthogonal convolutional RNN and GRU MCD mechanisms.

The knowledge derived from the initial development of the MCoRNNMCD-ANN model, primarily designed for predicting the hourly closing price of EUR/GBP, has been strategically leveraged in a new task focused on forecasting hourly closing price fluctuations for EUR/USD exchange rates using a one-year dataset. This transfer of knowledge is facilitated through the utilisation of the first two modules within the modular ANN-MCD architecture, wherein the prior insights from the CoRNNMCD and CoGRUMCD modules of the original MCoRNNMCD-ANN are adapted to the prediction framework for EUR/USD exchange rates.

Adopting a modular partial transfer learning approach presents adaptability and improves overall performance in capturing the underlying patterns within the new dataset, as underscored in Table 5.5. Maintaining the decision module as a fixed component during modular partial transfer learning while transferring the other two modules mitigates the risk of overfitting and facilitates improved generalisation to new samples. Refraining from fine-tuning the decision module is deliberate, aiming to preserve the general knowledge and patterns acquired during the pre-training stage. This strategic choice enhances the model's ability to generalise effectively to novel samples, showcasing the value of retaining foundational knowledge while adapting to the nuances of a distinct prediction task.

This chapter can also offer a practical implication. For instance, in Forex trading, even minor improvements in prediction accuracy can significantly impact profitability. Forex trading is often done with high leverage, meaning traders can control significant positions with relatively small amounts of capital. For example, consider a trader who makes 100 trades daily with an average profit of $10. Over a year, this trader would profit $260,000 (assuming 250 trading days per year). Suppose the trader can improve their average profit per trade by just 1%. In this case, their average profit per trade would increase to $10.10, and their annual profit would increase to $266,000. These gains may appear as minor returns growth, but they represent a 2.3% increase in profitability, which can be substantial. Furthermore, high trading frequencies and large trading volumes are expected in Forex trading, meaning that even minor improvements in prediction accuracy can be magnified. For example, suppose a trader has a winning percentage of 55% on their trades, meaning they win 55 out of every 100 trades. If they can increase their winning percentage to 56%, they can make more profitable trades and increase their overall profitability. It may seem slightly improved, but can significantly impact profitability throughout thousands of trades. Table B1 presents the predicted price movements by the MCoRNNMCD-ANN in Forex of the EUR/GBP. The thesis concludes in the Chapter below, presenting its limitations and future directions.

6. Conclusions and Future Research

This thesis has tackled the formidable challenges in Forex forecasting by introducing a novel computational approach that has combined economic theories, sentiments and recent neuroscience advancements through a modular neural network architecture. The proposed MCoRNNMCD-ANN framework showcased notable success in predicting price fluctuations across various currency pairs, surpassing existing state-of-the-art hybrid, ensemble and single models in the predictions of the hourly price fluctuations in EUR/GBP.

Adopting a modular architecture inspired by the modularity observed in human and animal brains has proven to be remarkably effective in capturing diverse patterns and features of data dynamics compared to monolithic architectures. Combining the new adaptative mechanism consisting of Monte Carlo dropout and orthogonal kernel initialisation into recurrent layers within a convolutional modular network, replacing the standard pooling layer has enhanced Forex prediction performance. Moreover, incorporating modular partial transfer learning from the proposed MCoRNNMCD-ANN has facilitated knowledge transfer between currency pairs. Consequently, the generalisation capabilities of the neural model utilised under data scarcity conditions have been augmented. The MANNMCDANN that applied this partial transfer learning technique has shown significantly better outcomes than the model that did not use it in the predictions of the hourly price fluctuations in EUR/USD.

Modular neural network architectures are poised to improve accuracy and robustness in Forex prediction models by potentially furnishing invaluable insights for traders navigating the foreign exchange market. Furthermore, integrating RCT and neuroscience through AI systems can elevate decision-making processes within Forex in an effort to foster more informed and successful trading strategies. Through ongoing research and exploration, this thesis endeavours to propel the development of more reliable and accurate methods for forecasting price movements in the dynamic and volatile Forex market.

In addition to the technical advancements made in this thesis, it is crucial to underscore the ethical considerations inherent in developing and applying predictive models in financial markets. Integrating AI, particularly in Forex forecasting, demands a thoughtful examination of potential ethical implications. As these models become increasingly influential in decision-making processes, the responsible and transparent use of such technology is crucial. Moreover, the responsible handling of sensitive financial data and protecting user privacy should be paramount. As AI systems rely on vast amounts of data, safeguarding the confidentiality and privacy of individuals and institutions involved in Forex trading is imperative. Striking a balance between data-driven insights and ethical data practices is vital to engender trust and maintain the integrity of the digital financial ecosystem.

6.1 Contributions

- First, a novel modular neural network architecture that drew inspiration from rational choice theory and cognitive neuroscience to model human decision-making in Forex price fluctuation predictions was presented. This approach was aimed to bring significant novelty to the field by incorporating modularity, rationality, and emotions. Modularity allows for decomposing the decision-making process into distinct modules, enabling a better understanding of how different factors, such as the sentiment of investors, contribute to price fluctuations. By integrating rationality and emotions, the proposed model has shown promising results in capturing psychological factors that could influence Forex trading decision-making. This approach pushes the boundaries of existing knowledge by combining interdisciplinary insights from economics neuroscience to be simulated from AI, providing new perspectives and potential advancements in predictive modelling in the Forex domain.

- The second approach introduced a new adaptation mechanism that combined Monte Carlo dropout and orthogonal kernel initialisation within a convolutional modular network. This mechanism was incorporated into the recurrent layer, replacing the traditional pooling mechanism from the CNN. The use of Monte Carlo dropout allowed adaptive width of the recurrent layers, enabling dynamic adjustments of capacity based on the complexities of Forex data. Furthermore, the proposed MCoRNNMCD-ANN model reduced redundancy and enhanced the representation of features in the data by initialising orthogonal kernels. This novel adaptation mechanism addresses the limitations of traditional pooling layers that neglect important features during their operation, offering a more flexible and efficient approach to handling Forex data complexities. Also, because the Monte Carlo dropout introduces a stochastic element that allows different subsets of units to be active or inactive during training, this stochastic behaviour enables the model to explore different feature combinations and prevents over-reliance on specific features. As a result, the model can become more adaptable and less prone to discarding essential features. Integrating Monte Carlo dropout and orthogonal kernel, initialisation introduces a novel aspect to the field of neural networks, tailored explicitly for modular designs, and offers new approaches for possibly improving prediction performance, uncertainty quantification, and reliability of Forex forecast modelling.

- The third approach proposed a novel method for addressing the challenge of data scarcity in Forex prediction by employing modular partial transfer learning. This approach utilised information from a previous task, such as the acquired knowledge (EUR/GBP) of the proposed MCoRNNMCD-ANN that transferred to fine-tune the first two modules of a modular ANN coupled with MCD in a new task with less data (EUR/USD). Accordingly, the modular ANNs' decision-making did not receive prior knowledge from the decision module of the proposed MCoRNNMCD-ANN. Furthermore, training the decision module of the MANNMCDANN (EUR/USD) from scratch on the target data can sustain the model better align with the decision boundaries and patterns specific to the target domain, improving generalisation performance. Partial transfer learning allows extracting relevant features and patterns from other data while avoiding the pitfalls of complete transfer learning, where irrelevant information may negatively impact performance. This novel approach acknowledges the scarcity of Forex data. Furthermore, it leverages the potential of transfer learning to overcome this challenge, offering a unique perspective on handling data limitations in predictive modelling.

In summary, all three approaches emphasise novelty in their respective areas. The first approach introduced a novel modular neural network architecture combining rational choice theory and cognitive neuroscience insights. The second approach presented a novel adaptation mechanism incorporating Monte Carlo dropout and orthogonal kernel initialisation in a recurrent layer within a convolutional modular network that replaced the pooling layer. Finally, the third approach proposes a novel strategy utilising partial transfer learning to address data scarcity in Forex prediction. These approaches provide new insights and potential advancements in predictive modelling in the Forex domain.

Regarding the research questions, the results presented in Chapter 5 support the hypothesis that:

- **RQ1**: The investigation into whether bio-inspired modular neural network architectures outperform monolithic ANN architectures in predicting price fluctuations in exchange rates has yielded positive results. The critical analysis and comparative evaluation have demonstrated that the proposed MCoRNNMCD-ANN significantly outperforms the state-of-the-art models. Furthermore, the modular architecture has shown superior prediction accuracy, successfully capturing complex market dynamics and outperforming monolithic architectures in decreasing prediction error. These findings highlighted the effectiveness of modularity in enhancing Forex price prediction models and its ability to capture and model the intricate dynamics of the Forex market.

- **RQ2**: Exploring the potential benefits of the new adaptative mechanism consisting of Monte Carlo dropout and orthogonal kernel initialisation into recurrent layers within a convolutional modular network, replacing the standard pooling layer, has yielded promising results. The practical evaluation has demonstrated the positive impact of these techniques on predicting Forex price fluctuations. By incorporating Monte Carlo dropout, the prediction error has been significantly reduced, and uncertainty quantification has been improved. Additionally, applying orthogonal weight initialisation methods has enhanced the optimisation process, leading to improved performance and robustness of the neural network models. These conclusions highlight the potential benefits of incorporating these techniques in enhancing the accuracy, reliability, and optimisation process of Forex price prediction models.

- **RQ3**: The investigation into how modular neural network architectures can leverage knowledge gained from the proposed MCoRNNMCD-ANN to enhance the performance and generalisation capabilities of ANNs in Forex predictions, particularly in data scarcity scenarios, has provided valuable insights. By incorporating partial transfer learning into modular architecture, the study successfully demonstrated the ability of modular architectures to leverage knowledge acquired from one currency pair and apply it to a different relevant task. This approach has effectively enhanced prediction models' generalisation capabilities and reliability in data-scarce scenarios. These findings address the challenge of limited data availability in Forex predictions and emphasise the potential of modular architectures in improving performance and generalisation capabilities.

In summary, the answers to the research questions confirm the effectiveness of bio-inspired modular neural network architectures in outperforming state-of-the-art and monolithic models, emphasise the benefits of incorporating Monte Carlo dropout and orthogonal weight initialisation methods, and highlight the ability of modular architectures to leverage knowledge and enhance generalisation capabilities in data scarcity scenarios. These findings contribute to advancing Forex price prediction models, providing practical implications for traders and investors in making informed decisions in the dynamic and complex Forex market.

6.2 Limitations and Future Research

There are some limitations in this thesis, which can be addressed in further studies:

- Firstly, it should be acknowledged that this research only focused on two specific Forex pairs (EUR/GBP and EUR/USD) as examples for forecasting price fluctuations. While these pairs are commonly traded and widely studied, they may not fully represent the entire Forex market. Therefore, the performance and effectiveness of the proposed MCoRNNMCD-ANN model may differ when applied to other Forex pairs. Thus, further studies should consider exploring more currency pairs to validate and generalise the findings.
- Secondly, while modular networks offer increased adaptability and modifiability compared to monolithic architectures, they may also introduce additional complexity in implementation and maintenance. The design and implementation of modular architectures require careful consideration of module interconnections, module design choices, and module coordination. Future research should focus on developing efficient tools and methodologies to facilitate the practical implementation and management of modular network architectures.
- Lastly, it is essential to note that while the proposed MCoRNNMCD-ANN model may improve the prediction error of Forex forecasting, it does not guarantee profits or eliminate all risks associated with trading in the Forex market. Therefore, trading in the Forex market always involves inherent risks, and it should be approached with caution and proper risk management strategies.

Future Research

Delving deeper into the incorporation of explainable AI (XAI) principles unveils a critical avenue for future research, particularly within the proposed modular network architecture context. As the sophistication and complexity of AI models continue to advance, the imperative for transparency and interpretability becomes increasingly pronounced, particularly in sensitive domains such as mental health. In these domains, where the consequences of AI-driven decisions significantly impact individuals, establishing a clear understanding of how the model arrives at its predictions is paramount. The integration of XAI aligns seamlessly with overarching ethical considerations and emerging regulatory requirements in deploying AI systems. Transparent models could satisfy the demand for accountability and contribute to responsible and trustworthy AI practices.

Moreover, the proposed modular network architecture could be applied to mental health prognoses. By leveraging the advantages of modularity, it may be possible to identify and predict patterns of mental health conditions, enabling early intervention and treatment. Specifically, the modular architecture can be employed to analyse and integrate various data sources, such as electronic health records, social media posts, and physiological data, to build a comprehensive profile of an individual's mental health. By leveraging the flexibility and adaptability of modular architectures, future studies can explore the application of these models in the mental health domain to possibly enhance prediction accuracy and enable personalised interventions.

The modular network approach may also have a broader applicability beyond mental health predictions. It could be utilised in other fields where integrating multiple data sources is necessary, such as environmental monitoring, cybersecurity, or energy management. The modular network approach offers a flexible and adaptable framework for integrating diverse data types, providing new insights and opportunities for prediction and optimisation in various domains.

In summary, future research should aim to address the limitations of this study by considering a more comprehensive range of Forex pairs, exploring the general applicability of hybrid activation functions, facilitating the implementation and maintenance of modular architectures, and investigating the potential of modular networks in mental health predictions and other relevant domains. The incorporation of XAI, therefore, transcends a mere technical enhancement; it becomes a cornerstone in the ethical and responsible use of AI in critical domains. By addressing these areas, researchers can further enhance prediction models' accuracy, reliability, and applicability and open new technological advancements.

References

Aasi, B., Imtiaz, S. A., Qadeer, H. A., Singarajah, M., & Kashef, R. (2021). Stock Price Prediction Using a Multivariate Multistep LSTM: A Sentiment and Public Engagement Analysis Model. In *2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)* (pp. 1–8). Toronto, ON, Canada: IEEE. doi: 10.1109/IEMTRONICS52119.2021.9422526

Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Mohamed, N. A., & Arshad, H. (2018). State-of-the-art in artificial neural network applications: A survey. *Heliyon*, *4*(11), e00938. doi: 10.1016/j.heliyon.2018.e00938

Acunzo, D. J., Low, D. M., & Fairhall, S. L. (2022). Deep neural networks reveal topic-level representations of sentences in medial prefrontal cortex, lateral anterior temporal lobe, precuneus, and angular gyrus. *NeuroImage*, *251*, 119005. doi: 10.1016/j.neuroimage.2022.119005

Adolphs, R. (1999). Social cognition and the human brain. *Trends in Cognitive Sciences*, *3*(12), 469–479. doi: 10.1016/S1364-6613(99)01399-6

Aggarwal, A., Gupta, I., Garg, N., & Goel, A. (2019). Deep Learning Approach to Determine the Impact of Socio Economic Factors on Bitcoin Price Prediction. In *2019 Twelfth International Conference on Contemporary Computing (IC3)* (pp. 1–5). Noida, India: IEEE. doi: 10.1109/IC3.2019.8844928

Ahmad, F. (2020). Personality traits as predictor of cognitive biases: moderating role of risk-attitude. *Qualitative Research in Financial Markets*, *12*(4), 465–484. doi: 10.1108/QRFM-10-2019-0123

Ajoku, K. K., Nwokonkwo, O. C., John-Otumu, A. M., & Oleji, C. P. (2021). A Model for Stock Market Value Forecasting using Ensemble Artificial Neural Network. *Journal of Advances in Computing, Communications and Information Technology*, *2*, 1–13. doi: 10.37121/jaccit.v2.162

Alamia, A., Gauducheau, V., Paisios, D., & VanRullen, R. (2020). Comparing feedforward and recurrent neural network architectures with human behavior in artificial grammar learning. *Scientific Reports*, *10*(1), 22172. doi: 10.1038/s41598-020-79127-y

Alghamdi, D., Alotaibi, F., & Rajgopal, J. (2021). A Novel Hybrid Deep Learning Model For Stock Price Forecasting. In *2021 International Joint Conference on Neural Networks (IJCNN)* (pp. 1–8). Shenzhen, China: IEEE. doi:

10.1109/IJCNN52387.2021.9533553

Ali, M., Sarwar, A., Sharma, V., & Suri, J. (2019). Artificial neural network based screening of cervical cancer using a hierarchical modular neural network architecture (HMNNA) and novel benchmark uterine cervix cancer database. *Neural Computing and Applications*, *31*(7), 2979–2993. doi: 10.1007/s00521-017-3246-7

Allman, J. M., Tetreault, N. A., Hakeem, A. Y., Manaye, K. F., Semendeferi, K., Erwin, J. M., ... Hof, P. R. (2010). The von Economo neurons in frontoinsular and anterior cingulate cortex in great apes and humans. *Brain Structure and Function*, *214*(5-6), 495–517. doi: 10.1007/s00429-010-0254-0

Almasri, E., & Arslan, E. (2018). Predicting Cryptocurrencies Prices with Neural Networks. In *2018 6th International Conference on Control Engineering & Information Technology (CEIT)* (pp. 1–5). Istanbul, Turkey: IEEE. doi: 10.1109/CEIT.2018.8751939

Alonso-Monsalve, S., Suárez-Cetrulo, A. L., Cervantes, A., & Quintana, D. (2020). Convolution on neural networks for high-frequency trend prediction of cryptocurrency exchange rates using technical indicators. *Expert Systems with Applications*, *149*, 113250. doi: 10.1016/j.eswa.2020.113250

Al-Sulaiman, T. (2022). Predicting reactions to anomalies in stock movements using a feed-forward deep learning network. *International Journal of Information Management Data Insights*, *2*(1), 100071. doi: 10.1016/j.jjimei.2022.100071

Althelaya, K. A., El-Alfy, E.-S. M., & Mohammed, S. (2018). Evaluation of bidirectional LSTM for short-and long-term stock market prediction. In *2018 9th International Conference on Information and Communication Systems (ICICS)* (pp. 151–156). Irbid: IEEE. doi: 10.1109/IACS.2018.8355458

Alvarez-Diez, S., Baixauli-Soler, J. S., & Belda-Ruiz, M. (2019). Co-movements between the British pound, the euro and the Japanese yen: the Brexit impact. *Journal of Economic Studies*, *46*(2), 467–481. doi: 10.1108/JES-01-2018-0007

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., ... Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, *8*(1), 53. doi: 10.1186/s40537-021-00444-8

Amer, M., & Maul, T. (2019). A review of modularization techniques in artificial neural networks. *Artificial Intelligence Review*, *52*(1), 527–561. doi: 10.1007/s10462-019-09706-7

Arjovsky, M., Shah, A., & Bengio, Y. (2015). Unitary Evolution Recurrent Neural Networks.

doi: 10.48550/ARXIV.1511.06464

Arnott, D., & Gao, S. (2019). Behavioral economics for decision support systems researchers. *Decision Support Systems*, *122*, 113063. doi: 10.1016/j.dss.2019.05.003

Aryal, S., Nadarajah, D., Kasthurirathna, D., Rupasinghe, L., & Jayawardena, C. (2019). Comparative analysis of the application of Deep Learning techniques for Forex Rate prediction. In *2019 International Conference on Advancements in Computing (ICAC)* (pp. 329–333). Malabe, Sri Lanka: IEEE. doi: 10.1109/ICAC49085.2019.9103428

Awunyo-Vitor, D. (2018). Theoretical and Conceptual Framework of Access to Financial Services by Farmers in Emerging Economies: Implication for Empirical Analysis. *Acta Universitatis Sapientiae, Economics and Business*, *6*(1), 43–59. doi: 10.1515/auseb-2018-0003

Baccianella, S., Esuli, A., Sebastiani, F., et al. (2010). Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In *Lrec* (Vol. 10, pp. 2200–2204).

Baek, Y., & Kim, H. Y. (2018). ModAugNet: A new forecasting framework for stock market index value with an overfitting prevention LSTM module and a prediction LSTM module. *Expert Systems with Applications*, *113*, 457–480. doi: 10.1016/j.eswa.2018.07.019

Bag, S., Gupta, S., Kumar, A., & Sivarajah, U. (2021). An integrated artificial intelligence framework for knowledge creation and B2B marketing rational decision making for improving firm performance. *Industrial Marketing Management*, *92*, 178–189. doi: 10.1016/j.indmarman.2020.12.001

Baniqued, P. L., Gallen, C. L., Kranz, M. B., Kramer, A. F., & D'Esposito, M. (2019). Brain network modularity predicts cognitive training-related gains in young adults. *Neuropsychologia*, *131*, 205–215. doi: 10.1016/j.neuropsychologia.2019.05.021

Barcellos, S. H., & Zamarro, G. (2021). Unbanked status and use of alternative financial services among minority populations. *Journal of Pension Economics and Finance*, *20*(4), 468–481. doi: 10.1017/S1474747219000052

Bathla, G. (2020). Stock Price prediction using LSTM and SVR. In *2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC)* (pp. 211–214). IEEE. doi: 10.1109/PDGC50313.2020.9315800

Bein, O., Reggev, N., & Maril, A. (2020). Prior knowledge promotes hippocampal separation but cortical assimilation in the left inferior frontal gyrus. *Nature Communications*, *11*(1), 4590. doi: 10.1038/s41467-020-18364-1

Bein, O., Trzewik, M., & Maril, A. (2019). The role of prior knowledge in incremental associative learning: An empirical and computational approach. *Journal of Memory and Language*, *107*, 1–24. doi: 10.1016/j.jml.2019.03.006

Berradi, Z., Lazaar, M., Mahboub, O., & Omara, H. (2020). A Comprehensive Review of Artificial Intelligence Techniques in Financial Market. In *2020 6th IEEE Congress on Information Science and Technology (CiSt)* (pp. 367–371). IEEE. doi: 10.1109/CiSt49399.2021.9357175

Bertolero, M. A., Yeo, B. T. T., Bassett, D. S., & D'Esposito, M. (2018). A mechanistic model of connector hubs, modularity and cognition. *Nature Human Behaviour*, *2*(10), 765–777. doi: 10.1038/s41562-018-0420-6

Bianchi, F., Gómez-Cram, R., Kind, T., & Kung, H. (2023). Threats to central bank independence: High-frequency identification with twitter. *Journal of Monetary Economics*, *135*, 37–54. doi: 10.1016/j.jmoneco.2023.01.001

Bormpotsis, C., Sedky, M., & Patel, A. (2023). Predicting Forex Currency Fluctuations Using a Novel Bio-Inspired Modular Neural Network. *Big Data and Cognitive Computing*, *7*(3), 152. doi: 10.3390/bdcc7030152

(Bud) Craig, A. D. (2009). How do you feel — now? The anterior insula and human awareness. *Nature Reviews Neuroscience*, *10*(1), 59–70. doi: 10.1038/nrn2555

Buskens, V. (2015). Rational Choice Theory in Sociology. In *International Encyclopedia of the Social & Behavioral Sciences* (pp. 901–906). Elsevier. doi: 10.1016/B978-0-08-097086-8.32177-8

Butti, C., Santos, M., Uppal, N., & Hof, P. R. (2013). Von Economo neurons: Clinical and evolutionary perspectives. *Cortex*, *49*(1), 312–326. doi: 10.1016/j.cortex.2011.10.004

Cao, Y., Jia, M., Ding, P., & Ding, Y. (2021). Transfer learning for remaining useful life prediction of multi-conditions bearings based on bidirectional-GRU network. *Measurement*, *178*, 109287. doi: 10.1016/j.measurement.2021.109287

Carbo-Valverde, S., Lacomba-Arias, J. A., Lagos-García, F. M., Rodriguez-Fernandez, F., & Verdejo-Román, J. (2020). Brain substrates explain differences in the adoption and degree of financial digitalization. *Scientific Reports*, *10*(1), 17512. doi: 10.1038/s41598-020-74554-3

Caucheteux, C., & King, J.-R. (2022). Brains and algorithms partially converge in natural language processing. *Communications Biology*, *5*(1), 134. doi: 10.1038/s42003-022-03036-1

Cauda, F., Torta, D. M. E., Sacco, K., D'Agata, F., Geda, E., Duca, S., … Vercelli, A. (2013). Functional anatomy of cortical areas characterized by Von Economo neurons. *Brain Structure and Function*, *218*(1), 1–20. doi: 10.1007/s00429-012-0382-9

Cavalli, S., & Amoretti, M. (2021). CNN-based multivariate data analysis for bitcoin trend prediction. *Applied Soft Computing*, *101*, 107065. doi: 10.1016/j.asoc.2020.107065

Cecchi, R., Vinckier, F., Hammer, J., Marusic, P., Nica, A., Rheims, S., ... Bastin, J. (2022). Intracerebral mechanisms explaining the impact of incidental feedback on mood state and risky choice. *eLife*, *11*, e72440. doi: 10.7554/eLife.72440

Cen, Z., & Wang, J. (2019). Crude oil price prediction model with long short term memory deep learning based on prior knowledge data transfer. *Energy*, *169*, 160–171. doi: 10.1016/j.energy.2018.12.016

Chaddock-Heyman, L., Weng, T. B., Loui, P., Kienzler, C., Weisshappel, R., Drollette, E. S., ... Kramer, A. F. (2021). Brain network modularity predicts changes in cortical thickness in children involved in a physical activity intervention. *Psychophysiology*, *58*(10), e13890. doi: 10.1111/psyp.13890

Chatterjee, A., Gupta, U., Chinnakotla, M. K., Srikanth, R., Galley, M., & Agrawal, P. (2019). Understanding Emotions in Text Using Deep Learning and Big Data. *Computers in Human Behavior*, *93*, 309–317. doi: 10.1016/j.chb.2018.12.029

Chen, C.-C., Kuo, C., Kuo, S.-Y., & Chou, Y.-H. (2015). Dynamic Normalization BPN for Stock Price Forecasting. In *2015 IEEE International Conference on Systems, Man, and Cybernetics* (pp. 2855–2860). Kowloon Tong, Hong Kong: IEEE. doi: 10.1109/SMC.2015.497

Chen, K., Chen, K., Wang, Q., He, Z., Hu, J., & He, J. (2019). Short-Term Load Forecasting With Deep Residual Networks. *IEEE Transactions on Smart Grid*, *10*(4), 3943–3952. doi: 10.1109/TSG.2018.2844307

Chen, M., Li, H., Wang, J., Yuan, W., Altaye, M., Parikh, N. A., & He, L. (2020). Early Prediction of Cognitive Deficit in Very Preterm Infants Using Brain Structural Connectome With Transfer Learning Enhanced Deep Convolutional Neural Networks. *Frontiers in Neuroscience*, *14*, 858. doi: 10.3389/fnins.2020.00858

Chen, X., Necus, J., Peraza, L. R., Mehraram, R., Wang, Y., O'Brien, J. T., ... Taylor, J.-P. (2021). The functional brain favours segregated modular connectivity at old age unless affected by neurodegeneration. *Communications Biology*, *4*(1), 973. doi: 10.1038/s42003-021-02497-0

Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. (Publisher: arXiv Version Number: 3) doi: 10.48550/ARXIV.1406.1078

Christopoulos, G. I., Tobler, P. N., Bossaerts, P., Dolan, R. J., & Schultz, W. (2009). Neural Correlates of Value, Risk, and Risk Aversion Contributing to Decision Making under Risk. *The Journal of Neuroscience*, *29*(40), 12574–12583. doi: 10.1523/JNEUROSCI.2614-09.2009

Christov-Moore, L., Reggente, N., Douglas, P. K., Feusner, J. D., & Iacoboni, M. (2020). Predicting Empathy From Resting State Brain Connectivity: A Multivariate Approach. *Frontiers in Integrative Neuroscience*, *14*, 3. doi: 10.3389/fnint.2020.00003

Colander, D., Holt, R., & Rosser Jr, B. (2004). The changing face of mainstream economics. *Review of Political Economy*, *16*(4), 485–499.

Courbariaux, M., Bengio, Y., & David, J.-P. (2015). BinaryConnect: Training Deep Neural Networks with binary weights during propagations. (Publisher: arXiv Version Number: 3) doi: 10.48550/ARXIV.1511.00363

Dai, H.-J., & Jonnagaddala, J. (2018). Assessing the severity of positive valence symptoms in initial psychiatric evaluation records: Should we use convolutional neural networks? *PLOS ONE*, *13*(10), e0204493. doi: 10.1371/journal.pone.0204493

De Martino, B., Kumaran, D., Seymour, B., & Dolan, R. J. (2006). Frames, Biases, and Rational Decision-Making in the Human Brain. *Science*, *313*(5787), 684–687. doi: 10.1126/science.1128356

Dennison, J. B., Sazhin, D., & Smith, D. V. (2022). Decision neuroscience and neuroeconomics: Recent progress and ongoing challenges. *WIREs Cognitive Science*, *13*(3), e1589. doi: 10.1002/wcs.1589

Ding, Y., Ding, P., Zhao, X., Cao, Y., & Jia, M. (2022). Transfer Learning for Remaining Useful Life Prediction Across Operating Conditions Based on Multisource Domain Adaptation. *IEEE/ASME Transactions on Mechatronics*, *27*(5), 4143–4152. doi: 10.1109/TMECH.2022.3147534

Dong, Y., Wang, J., & Guo, Z. (2018). Research and application of local perceptron neural network in highway rectifier for time series forecasting. *Applied Soft Computing*, *64*, 656–673. doi: 10.1016/j.asoc.2017.12.022

Duan, H., & Wang, X. (2016). Echo State Networks With Orthogonal Pigeon-Inspired Optimization for Image Restoration. *IEEE Transactions on Neural Networks and Learning Systems*, *27*(11), 2413–2425. doi: 10.1109/TNNLS.2015.2479117

Duthie, R., & Budzynska, K. (2018). A Deep Modular RNN Approach for Ethos Mining. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence* (pp. 4041–4047). Stockholm, Sweden: International Joint Conferences on Artificial Intelligence Organization. doi: 10.24963/ijcai.2018/562

Duz Tan, S., & Tas, O. (2021). Social Media Sentiment in International Stock Returns and Trading Activity. *Journal of Behavioral Finance*, *22*(2), 221–234. doi: 10.1080/15427560.2020.1772261

Dürr, O., Murina, E., Siegismund, D., Tolkachev, V., Steigele, S., & Sick, B. (2018). Know When You Don't Know: A Robust Deep Learning Approach in the Presence of Unknown Phenotypes. *ASSAY and Drug Development Technologies*, *16*(6), 343–349. doi: 10.1089/adt.2018.859

Eichenbaum, H. (2004). Hippocampus. *Neuron*, *44*(1), 109–120. doi: 10.1016/j.neuron.2004.08.028

Ellefsen, K. O., Mouret, J.-B., & Clune, J. (2015). Neural Modularity Helps Organisms Evolve to Learn New Skills without Forgetting Old Skills. *PLOS Computational Biology*, *11*(4), e1004128. doi: 10.1371/journal.pcbi.1004128

Ezen-Can, A. (2020). A Comparison of LSTM and BERT for Small Corpus. (Publisher: arXiv Version Number: 1) doi: 10.48550/ARXIV.2009.05451

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., & Sunde, U. (2018). Global Evidence on Economic Preferences*. *The Quarterly Journal of Economics*, *133*(4), 1645–1692. doi: 10.1093/qje/qjy013

Fan, J., Zheng, P., & Li, S. (2022). Vision-based holistic scene understanding towards proactive human–robot collaboration. *Robotics and Computer-Integrated Manufacturing*, *75*, 102304. doi: 10.1016/j.rcim.2021.102304

Farsi, B., Amayri, M., Bouguila, N., & Eicker, U. (2021). On Short-Term Load Forecasting Using Machine Learning Techniques and a Novel Parallel Deep LSTM-CNN Approach. *IEEE Access*, *9*, 31191–31212. doi: 10.1109/ACCESS.2021.3060290

Fermin, A. S. R., Friston, K., & Yamawaki, S. (2022). An insula hierarchical network architecture for active interoceptive inference. *Royal Society Open Science*, *9*(6), 220226. doi: 10.1098/rsos.220226

Flesch, T., Juechems, K., Dumbalska, T., Saxe, A., & Summerfield, C. (2022). Orthogonal representations for robust context-dependent task performance in brains and neural networks. *Neuron*, *110*(7), 1258–1270.e11. doi: 10.1016/j.neuron.2022.01.005

Frydman, C., & Camerer, C. F. (2016). The Psychology and Neuroscience of Financial Decision Making. *Trends in Cognitive Sciences*, *20*(9), 661–675. doi: 10.1016/j.tics.2016.07.003

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, *36*(4), 193–202. doi: 10.1007/BF00344251

Gal, Y., & Ghahramani, Z. (2015). Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. (Publisher: arXiv Version Number: 6) doi: 10.48550/ARXIV.1506.02142

Gal, Y., & Ghahramani, Z. (2016). Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning* (pp. 1050–1059).

Galeshchuk, S., & Mukherjee, S. (2017). Deep networks for predicting direction of change in foreign exchange rates. *Intelligent Systems in Accounting,* *Finance and Management*, *24*(4), 100–110. doi: 10.1002/isaf.1404

Golmohammadi, M., Ziyabari, S., Shah, V., Von Weltin, E., Campbell, C., Obeid, I., & Picone, J. (2017). Gated recurrent networks for seizure detection. In *2017 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)* (pp. 1–5). Philadelphia, PA: IEEE. doi: 10.1109/SPMB.2017.8257020

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. book in preparation for mit press. *URL¡ http://www. deeplearningbook. org*, *1*.

Gradojevic, N., Gencay, R., & Kukolj, D. (2009). Option Pricing With Modular Neural Networks. *IEEE Transactions on Neural Networks*, *20*(4), 626–637. doi: 10.1109/TNN.2008.2011130

Greaves, A., & Au, B. (2015). Using the bitcoin transaction graph to predict the price of bitcoin. *No data*, *8*, 416–443.

Grossmann, T. (2013). The role of medial prefrontal cortex in early social cognition. *Frontiers in Human Neuroscience*, *7*. doi: 10.3389/fnhum.2013.00340

Hayward, R. (2018). Foreign Exchange Speculation: An Event Study. *International Journal of Financial Studies*, *6*(1), 22. doi: 10.3390/ijfs6010022

He, L., Li, H., Wang, J., Chen, M., Gozdas, E., Dillman, J. R., & Parikh, N. A. (2020). A multi-task, multi-stage deep transfer learning model for early prediction of neurodevelopment in very preterm infants. *Scientific Reports*, *10*(1), 15072. doi: 10.1038/s41598-020-71914-x

Hernandez, A. E., Claussenius-Kalman, H. L., Ronderos, J., Castilla-Earls, A. P., Sun, L., Weiss, S. D., & Young, D. R. (2019). Neuroemergentism: A framework for studying cognition and the brain. *Journal of Neurolinguistics*, *49*, 214–223. doi: 10.1016/j.jneuroling.2017.12.010

Herrera, G. P., Constantino, M., Su, J.-J., & Naranpanawa, A. (2022). Renewable energy stocks forecast using Twitter investor sentiment and deep learning. *Energy Economics*, *114*, 106285. doi: 10.1016/j.eneco.2022.106285

Hilbert, M., & Darmon, D. (2020). How Complexity and Uncertainty Grew with Algorithmic Trading. *Entropy*, *22*(5), 499. doi: 10.3390/e22050499

Hiser, J., & Koenigs, M. (2018). The Multifaceted Role of the Ventromedial Prefrontal Cortex in Emotion, Decision Making, Social Cognition, and Psychopathology. *Biological Psychiatry*, *83*(8), 638–647. doi: 10.1016/j.biopsych.2017.10.030

Hochreiter, S. (1991). Untersuchungen zu dynamischen neuronalen netzen. *Diploma, Technische Universität München*, *91*(1), 31.

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. doi: 10.1162/neco.1997.9.8.1735

Hossain, M. A., Karim, R., Thulasiram, R., Bruce, N. D. B., & Wang, Y. (2018). Hybrid Deep Learning Model for Stock Price Prediction. In *2018 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. 1837–1844). Bangalore, India: IEEE. doi: 10.1109/SSCI.2018.8628641

Hutto, C., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*(1), 216–225. doi: 10.1609/icwsm.v8i1.14550

Ishfaq, M., Nazir, M. S., Qamar, M. A. J., & Usman, M. (2020). Cognitive Bias and the Extraversion Personality Shaping the Behavior of Investors. *Frontiers in Psychology*, *11*, 556506. doi: 10.3389/fpsyg.2020.556506

Islam, M. S., Hossain, E., Rahman, A., Hossain, M. S., & Andersson, K. (2020). A Review on Recent Advancements in FOREX Currency Prediction. *Algorithms*, *13*(8), 186. doi: 10.3390/a13080186

Jiang, X., Li, S.-P., Mai, Y., & Tian, T. (2022). Study of multinational currency co-movement and exchange rate stability base on network game. *Finance Research Letters*, *47*, 102601. doi: 10.1016/j.frl.2021.102601

Jing, N., Wu, Z., & Wang, H. (2021). A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction. *Expert Systems with Applications*, *178*, 115019. doi: 10.1016/j.eswa.2021.115019

Kanwal, A., Lau, M. F., Ng, S. P., Sim, K. Y., & Chandrasekaran, S. (2022). BiCuDNNLSTM-1dCNN — A hybrid deep learning-based predictive model for stock price prediction. *Expert Systems with Applications*, *202*, 117123. doi: 10.1016/j.eswa.2022.117123

Kelkar, A., & Medaglia, J. D. (2018). Evidence of Brain Modularity. In T. K. Shackelford & V. A. Weekes-Shackelford (Eds.), *Encyclopedia of Evolutionary Psychological Science* (pp. 1–10). Cham: Springer International Publishing. doi: 10.1007/978-3-319-16999-6_2422-1

Khedr, A. M., Arif, I., El-Bannany, M., Alhashmi, S. M., & Sreedharan, M. (2021). Cryptocurrency price prediction using traditional statistical and machine-learning techniques: A survey. *Intelligent Systems in Accounting, Finance and Management*, *28*(1), 3–34. doi: 10.1002/isaf.1488

Kietzmann, T. C., McClure, P., & Kriegeskorte, N. (2017). *Deep Neural Networks in Computational Neuroscience* (preprint). Neuroscience. doi: 10.1101/133504

Kingma, D. P., & Ba, J. (2017). *Adam: A Method for Stochastic Optimization.* arXiv. (arXiv:1412.6980 [cs])

Kingma, D. P., Salimans, T., & Welling, M. (2015). Variational dropout and the local reparameterization trick. *Advances in neural information processing systems*, *28*.

Kirisci, M., & Cagcag Yolcu, O. (2022). A New CNN-Based Model for Financial Time Series: TAIEX and FTSE Stocks Forecasting. *Neural Processing Letters*, *54*(4), 3357–3374. doi: 10.1007/s11063-022-10767-z

Kong, X.-Z., Mathias, S. R., Guadalupe, T., ENIGMA Laterality Working Group, Glahn, D. C., Franke, B., ... Orhan, F. (2018). Mapping cortical brain asymmetry in 17,141 healthy individuals worldwide via the ENIGMA Consortium. *Proceedings of the National Academy of Sciences*, *115*(22). doi: 10.1073/pnas.1718418115

Kraus, M., & Feuerriegel, S. (2017). Decision support from financial disclosures with deep neural networks and transfer learning. *Decision Support Systems*, *104*, 38–48. doi: 10.1016/j.dss.2017.10.001

Kriegeskorte, N. (2015). Deep Neural Networks: A New Framework for Modeling Biological Vision and Brain Information Processing. *Annual Review of Vision Science*, *1*(1), 417–446. doi: 10.1146/annurev-vision-082114-035447

Kriegeskorte, N., & Douglas, P. K. (2018). Cognitive computational neuroscience. *Nature Neuroscience*, *21*(9), 1148–1160. doi: 10.1038/s41593-018-0210-5

Kroker, T., Wyczesany, M., Rehbein, M. A., Roesmann, K., Wessing, I., & Junghöfer, M. (n.d., November). Noninvasive stimulation of the ventromedial prefrontal cortex modulates rationality of human decision-making. *Scientific Reports*, *12*(1), 20213. doi: 10.1038/s41598-022-24526-6

Kumar, S., De, K., & Roy, P. P. (2020). Movie Recommendation System Using Sentiment Analysis From Microblogging Data. *IEEE Transactions on Computational Social Systems*, *7*(4), 915–923. doi: 10.1109/TCSS.2020.2993585

Kumaran, D., Summerfield, J. J., Hassabis, D., & Maguire, E. A. (2009). Tracking the Emergence of Conceptual Knowledge during Human Decision Making. *Neuron*, *63*(6), 889–901. doi: 10.1016/j.neuron.2009.07.030

LaBar, K. S., & Cabeza, R. (2006). Cognitive neuroscience of emotional memory. *Nature Reviews Neuroscience*, *7*(1), 54–64. doi: 10.1038/nrn1825

Lamm, C., & Singer, T. (2010). The role of anterior insular cortex in social emotions. *Brain Structure and Function*, *214*(5-6), 579–591. doi: 10.1007/s00429-010-0251-3

Lanbouri, Z., & Achchab, S. (2020). Stock Market prediction on High frequency data using Long-Short Term Memory. *Procedia Computer Science*, *175*, 603–608. doi: 10.1016/j.procs.2020.07.087

Lau, J. K. L., Ozono, H., Kuratomi, K., Komiya, A., & Murayama, K. (2020). Shared striatal activity in decisions to satisfy curiosity and hunger at the risk of electric shocks. *Nature Human Behaviour*, *4*(5), 531–543. doi: 10.1038/s41562-020-0848-3

Laves, M.-H., Ihler, S., Fast, J. F., Kahrs, L. A., & Ortmaier, T. (2020). Well-calibrated regression uncertainty in medical imaging with deep learning. In *Medical imaging with deep learning* (pp. 393–412).

Lee, S. W., & Kim, H. Y. (2020). Stock market forecasting with super-high dimensional time-series data using ConvLSTM, trend sampling, and specialized data augmentation. *Expert Systems with Applications*, *161*, 113704. doi: 10.1016/j.eswa.2020.113704

Leong, J. K., Ho, T. C., Colich, N. L., Sisk, L., Knutson, B., & Gotlib, I. H. (2021). White-matter tract connecting anterior insula to nucleus accumbens predicts greater future motivation in adolescents. *Developmental Cognitive Neuroscience*, *47*, 100881. doi: 10.1016/j.dcn.2020.100881

Leong, J. K., Pestilli, F., Wu, C. C., Samanez-Larkin, G. R., & Knutson, B. (2016). White-Matter Tract Connecting Anterior Insula to Nucleus Accumbens Correlates with Reduced Preference for Positively Skewed Gambles. *Neuron*, *89*(1), 63–69. doi: 10.1016/j.neuron.2015.12.015

Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and Decision Making. *Annual Review of Psychology*, *66*(1), 799–823. doi: 10.1146/annurev-psych-010213-115043

Li, B., & Rangarajan, S. (2022). A conceptual study of transfer learning with linear models for data-driven property prediction. *Computers & Chemical Engineering*, *157*, 107599. doi: 10.1016/j.compchemeng.2021.107599

Li, H., Parikh, N. A., & He, L. (2018). A Novel Transfer Learning Approach to Enhance Deep Neural Network Classification of Brain Functional Connectomes. *Frontiers in Neuroscience*, *12*, 491. doi: 10.3389/fnins.2018.00491

Li, R., & Zhang, J. (2020). Review of computational neuroaesthetics: bridging the gap between neuroaesthetics and computer science. *Brain Informatics*, *7*(1), 16. doi: 10.1186/s40708-020-00118-w

Liu, S., Ji, H., & Wang, M. C. (2020). Nonpooling Convolutional Neural Network Forecasting for Seasonal Time Series With Trends. *IEEE Transactions on Neural Networks and Learning Systems*, *31*(8), 2879–2888. doi: 10.1109/TNNLS.2019.2934110

Livet, P. (2010). Rational choice, neuroeconomy and mixed emotions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1538), 259–269. doi: 10.1098/rstb.2009.0177

Loria, S., et al. (2018). textblob documentation. *Release 0.15*, *2*(8), 269.

Loued-Khenissi, L., Pfeuffer, A., Einhäuser, W., & Preuschoff, K. (2020). Anterior insula reflects surprise in value-based decision-making and perception. *NeuroImage*, *210*, 116549. doi: 10.1016/j.neuroimage.2020.116549

Lu, W., Li, J., Li, Y., Sun, A., & Wang, J. (2020). A CNN-LSTM-Based Model to Forecast Stock Prices. *Complexity*, *2020*, 1–10. doi: 10.1155/2020/6622927

Luboeinski, J., & Tetzlaff, C. (2021). Memory consolidation and improvement by synaptic tagging and capture in recurrent neural networks. *Communications Biology*, *4*(1), 275. doi: 10.1038/s42003-021-01778-y

Lun, X., Yu, Z., Chen, T., Wang, F., & Hou, Y. (2020). A Simplified CNN Classification Method for MI-EEG via the Electrode Pairs Signals. *Frontiers in Human Neuroscience*, *14*, 338. doi: 10.3389/fnhum.2020.00338

Lydon-Staley, D. M., Ciric, R., Satterthwaite, T. D., & Bassett, D. S. (2019). Evaluation of confound regression strategies for the mitigation of micromovement artifact in studies of dynamic resting-state functional connectivity and multilayer network modularity. *Network Neuroscience*, *3*(2), 427–454. doi: 10.1162/netn_a_00071

M, H., E.A., G., Menon, V. K., & K.P., S. (2018). NSE Stock Market Prediction Using Deep-Learning Models. *Procedia Computer Science*, *132*, 1351–1362. doi: 10.1016/j.procs.2018.05.050

Ma, X., Zhang, J., Du, B., Ding, C., & Sun, L. (2019). Parallel Architecture of Convolutional Bi-Directional LSTM Neural Networks for Network-Wide Metro Ridership Prediction. *IEEE Transactions on Intelligent Transportation Systems*, *20*(6), 2278–2288. doi: 10.1109/TITS.2018.2867042

Mai, Y., Chen, H., Zou, J.-Z., & Li, S.-P. (2018). Currency co-movement and network correlation structure of foreign exchange market. *Physica A: Statistical Mechanics and its Applications*, *492*, 65–74. doi: 10.1016/j.physa.2017.09.068

Maneejuk, P., & Srichaikul, W. (2021). Forecasting foreign exchange markets: further evidence using machine learning models. *Soft Computing*, *25*(12), 7887–7898. doi: 10.1007/s00500-021-05830-1

McEwen, B. S., Bowles, N. P., Gray, J. D., Hill, M. N., Hunter, R. G., Karatsoreos, I. N., & Nasca, C. (2015). Mechanisms of stress in the brain. *Nature Neuroscience*, *18*(10), 1353–1363. doi: 10.1038/nn.4086

McNally, S., Roche, J., & Caton, S. (2018). Predicting the Price of Bitcoin Using Machine Learning. In *2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP)* (pp. 339–343). Cambridge: IEEE. doi: 10.1109/PDP2018.2018.00060

Merello, S., Ratto, A. P., Oneto, L., & Cambria, E. (2019). Ensemble Application of Transfer Learning and Sample Weighting for Stock Market Prediction. In *2019 International Joint Conference on Neural Networks (IJCNN)* (pp. 1–8). Budapest, Hungary: IEEE. doi: 10.1109/IJCNN.2019.8851938

Miok, K., Nguyen-Doan, D., Škrlj, B., Zaharie, D., & Robnik-Šikonja, M. (2019). Prediction Uncertainty Estimation for Hate Speech Classification. In C. Martín-Vide, M. Purver, & S. Pollak (Eds.), *Statistical Language and Speech Processing* (Vol. 11816, pp. 286–298). Cham: Springer International Publishing. (Series Title: Lecture Notes in Computer Science) doi: 10.1007/978-3-030-31372-2_24

Mohan, S., Mullapudi, S., Sammeta, S., Vijayvergia, P., & Anastasiu, D. C. (2019). Stock Price Prediction Using News Sentiment Analysis. In *2019 IEEE*

Fifth International Conference on Big Data Computing Service and Applications (BigDataService) (pp. 205–208). Newark, CA, USA: IEEE. doi: 10.1109/BigDataService.2019.00035

Mohr, P. N. C., Biele, G., & Heekeren, H. R. (2010). Neural Processing of Risk. *The Journal of Neuroscience*, *30*(19), 6613–6619. doi: 10.1523/JNEUROSCI.0003-10.2010

Morrone, J. A., Weber, J. K., Huynh, T., Luo, H., & Cornell, W. D. (2020). Combining Docking Pose Rank and Structure with Deep Learning Improves Protein–Ligand Binding Mode Prediction over a Baseline Docking Approach. *Journal of Chemical Information and Modeling*, *60*(9), 4170–4179. doi: 10.1021/acs.jcim.9b00927

Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Icml 2010* (pp. 807–814).

Nguyen, T.-T., & Yoon, S. (2019). A Novel Approach to Short-Term Stock Price Movement Prediction using Transfer Learning. *Applied Sciences*, *9*(22), 4745. doi: 10.3390/app9224745

Nielsen, F. A. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs.
(Publisher: arXiv Version Number: 1) doi: 10.48550/ARXIV.1103.2903

Nieuwenhuis, I. L., & Takashima, A. (2011). The role of the ventromedial prefrontal cortex in memory consolidation. *Behavioural Brain Research*, *218*(2), 325–334. doi: 10.1016/j.bbr.2010.12.009

Nikou, M., Mansourfar, G., & Bagherzadeh, J. (2019). Stock price prediction using DEEP learning algorithm and its comparison with machine learning algorithms. *Intelligent Systems in Accounting, Finance and Management*, *26*(4), 164–174. doi: 10.1002/isaf.1459

Nobata, C., Tetreault, J., Thomas, A., Mehdad, Y., & Chang, Y. (2016). Abusive Language Detection in Online User Content. In *Proceedings of the 25th International Conference on World Wide Web* (pp. 145–153). Montréal Québec Canada: International World Wide Web Conferences Steering Committee. doi: 10.1145/2872427.2883062

Ojo, S. O., Owolawi, P. A., Mphahlele, M., & Adisa, J. A. (2019). Stock Market Behaviour Prediction using Stacked LSTM Networks. In *2019 International Multidisciplinary Information Technology and Engineering Conference (IMITEC)* (pp. 1–5). Vanderbijlpark, South Africa: IEEE. doi: 10.1109/IMITEC45504.2019.9015840

Olsen, R. K., Moses, S. N., Riggs, L., & Ryan, J. D. (2012). The hippocampus supports multiple cognitive processes through relational binding and comparison. *Frontiers in Human Neuroscience*, *6*. doi: 10.3389/fnhum.2012.00146

Padoa-Schioppa, C. (2011). Neurobiology of Economic Choice: A Good-Based Model. *Annual Review of Neuroscience*, *34*(1), 333–359. doi: 10.1146/annurev-neuro-061010-113648

Park, Y. S., Konge, L., & Artino, A. R. (2020). The Positivism Paradigm of Research. *Academic Medicine*, *95*(5), 690–694. doi: 10.1097/ACM.0000000000003093

Peng, X.-r., Lei, X., Xu, P., & Yu, J. (2020). The Age-related Neural Strategy Alterations in Decision Making Under Risk. *Neuroscience*, *440*, 30–38. doi: 10.1016/j.neuroscience.2020.05.017

Phelps, E. A., & LeDoux, J. E. (2005). Contributions of the Amygdala to Emotion Processing: From Animal Models to Human Behavior. *Neuron*, *48*(2), 175–187. doi: 10.1016/j.neuron.2005.09.025

Pizzo, F., Roehri, N., Medina Villalon, S., Trébuchon, A., Chen, S., Lagarde, S., ... Bénar, C. G. (2019). Deep brain activities can be detected with magnetoencephalography. *Nature Communications*, *10*(1), 971. doi: 10.1038/s41467-019-08665-5

Pokhrel, N. R., Dahal, K. R., Rimal, R., Bhandari, H. N., Khatri, R. K., Rimal, B., & Hahn, W. E. (2022). Predicting NEPSE index price using deep learning models. *Machine Learning with Applications*, *9*, 100385. doi: 10.1016/j.mlwa.2022.100385

Price, J. L., & Drevets, W. C. (2010). Neurocircuitry of Mood Disorders. *Neuropsychopharmacology*, *35*(1), 192–216. doi: 10.1038/npp.2009.104

Prieto, A., Prieto, B., Ortigosa, E. M., Ros, E., Pelayo, F., Ortega, J., & Rojas, I. (2016). Neural networks: An overview of early research, current frameworks and new challenges. *Neurocomputing*, *214*, 242–268. doi: 10.1016/j.neucom.2016.06.014

Pujara, M. S., Wolf, R. C., Baskaya, M. K., & Koenigs, M. (2015, December). Ventromedial prefrontal cortex damage alters relative risk tolerance for prospective gains and losses. *Neuropsychologia*, *79*, 70–75. doi: 10.1016/j.neuropsychologia.2015.10.026

Qi, L., Khushi, M., & Poon, J. (2020). Event-Driven LSTM For Forex Price Prediction. In *2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)* (pp. 1–6). Gold Coast, Australia: IEEE. doi: 10.1109/CSDE50874.2020.9411540

Ramachandran, P., Zoph, B., & Le, Q. V. (2017). Searching for Activation Functions.
(Publisher: arXiv Version Number: 2) doi: 10.48550/ARXIV.1710.05941

Rangel, A., Camerer, C., & Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience*, *9*(7), 545–556. doi: 10.1038/nrn2357

Rangel, A., & Hare, T. (2010). Neural computations associated with goal-directed choice. *Current Opinion in Neurobiology*, *20*(2), 262–270. doi: 10.1016/j.conb.2010.03.001

Ray, R., Khandelwal, P., & Baranidharan, B. (2018). A Survey on Stock Market Prediction using Artificial Intelligence Techniques. In *2018 International Conference on Smart Systems and Inventive Technology (ICSSIT)* (pp. 594–598). IEEE. doi: 10.1109/ICSSIT.2018.8748680

Ren, J. (2021). Research on Financial Investment Decision Based on Artificial Intelligence Algorithm. *IEEE Sensors Journal*, *21*(22), 25190–25197. doi: 10.1109/JSEN.2021.3104038

Rilling, J. K., & Sanfey, A. G. (2011). The Neuroscience of Social Decision-Making. *Annual Review of Psychology*, *62*(1), 23–48. doi: 10.1146/annurev.psych.121208.131647

Roozendaal, B., McEwen, B. S., & Chattarji, S. (2009). Stress, memory and the amygdala. *Nature Reviews Neuroscience*, *10*(6), 423–433. doi: 10.1038/nrn2651

Ruissen, M. I., Overgaauw, S., & De Bruijn, E. R. A. (2018). Being right, but losing money: the role of striatum in joint decision making. *Scientific Reports*, *8*(1), 6711. doi: 10.1038/s41598-018-24617-3

Russin, J., O'Reilly, R. C., & Bengio, Y. (2020). Deep learning needs a prefrontal cortex. *Work Bridging AI Cogn Sci*, *107*(603-616), 1. URL `https://baicsworkshop.github.io/pdf/BAICS_10.pdf`.

Sadiq, S., Umer, M., Ullah, S., Mirjalili, S., Rupapara, V., & Nappi, M. (2021). Discrepancy detection between actual user reviews and numeric ratings of Google App store using deep learning. *Expert Systems with Applications*, *181*, 115111. doi: 10.1016/j.eswa.2021.115111

Sadr, M. A. M., Gante, J., Champagne, B., Falcao, G., & Sousa, L. (2022). Uncertainty Estimation via Monte Carlo Dropout in CNN-Based mmWave MIMO Localization. *IEEE Signal Processing Letters*, *29*, 269–273. doi: 10.1109/LSP.2021.3130504

Saunders, M. N. K., Lewis, P., & Thornhill, A. (2019). *Research methods for business students* (Eighth Edition ed.). New York: Pearson.

Saxe, A. M., McClelland, J. L., & Ganguli, S. (2013). Exact solutions to the nonlinear dynamics of learning in deep linear neural networks. (Publisher: arXiv Version Number: 3) doi: 10.48550/ARXIV.1312.6120

Seifollahi, S., & Shajari, M. (2019). Word sense disambiguation application in sentiment analysis of news headlines: an applied approach to FOREX market prediction. *Journal of Intelligent Information Systems*, *52*(1), 57–83. doi: 10.1007/s10844-018-0504-9

Sezer, O. B., Gudelek, M. U., & Ozbayoglu, A. M. (2020). Financial time series forecasting with deep learning : A systematic literature review: 2005–2019. *Applied Soft Computing*, *90*, 106181. doi: 10.1016/j.asoc.2020.106181

Shafiei, S. B., Lone, Z., Elsayed, A. S., Hussein, A. A., & Guru, K. A. (2020). Identifying mental health status using deep neural network trained by visual metrics. *Translational Psychiatry*, *10*(1), 430. doi: 10.1038/s41398-020-01117-5

Shahi, T. B., Shrestha, A., Neupane, A., & Guo, W. (2020). Stock Price Forecasting with Deep Learning: A Comparative Study. *Mathematics*, *8*(9), 1441. doi: 10.3390/math8091441

Sharwood Smith, M. (2019). The compatibility within a modular framework of emergent and dynamical processes in mind and brain. *Journal of Neurolinguistics*, *49*, 240–244. doi: 10.1016/j.jneuroling.2018.04.010

Shi, J., Wen, H., Zhang, Y., Han, K., & Liu, Z. (2018). Deep recurrent neural network reveals a hierarchy of process memory during dynamic natural vision. *Human Brain Mapping*, *39*(5), 2269–2282. doi: 10.1002/hbm.24006

Shiao, Y. C., Chakraborty, G., Chen, S. F., Hua Li, L., & Chen, R. C. (2019). Modeling and Prediction of Time-Series-A Case Study with Forex Data. In *2019 IEEE 10th International Conference on Awareness Science and Technology (iCAST)* (pp. 1–5). Morioka, Japan: IEEE. doi: 10.1109/ICAwST.2019.8923188

Sim, H. S., Kim, H. I., & Ahn, J. J. (2019). Is Deep Learning for Image Recognition Applicable to Stock Market Prediction? *Complexity*, *2019*, 1–10. doi: 10.1155/2019/4324878

Sirignano, J., & Cont, R. (2019). Universal features of price formation in financial markets: perspectives from deep learning. *Quantitative Finance*, *19*(9), 1449–1459. doi: 10.1080/14697688.2019.1622295

Sirimevan, N., Mamalgaha, I. U. H., Jayasekara, C., Mayuran, Y. S., & Jayawardena, C. (2019). Stock Market Prediction Using Machine Learning Techniques. In *2019 International Conference on Advancements in Computing (ICAC)* (pp. 192–197). Malabe, Sri Lanka: IEEE. doi: 10.1109/ICAC49085.2019.9103381

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, *104*, 333–339. doi: 10.1016/j.jbusres.2019.07.039

Sohangir, S., Petty, N., & Wang, D. (2018). Financial Sentiment Lexicon Analysis. In *2018 IEEE 12th International Conference on Semantic Computing (ICSC)* (pp. 286–289). Laguna Hills, CA, USA: IEEE. doi: 10.1109/ICSC.2018.00052

Souma, W., Vodenska, I., & Aoyama, H. (2019). Enhanced news sentiment analysis using deep learning methods. *Journal of Computational Social*

Science, *2*(1), 33–46. doi: 10.1007/s42001-019-00035-x

Sporns, O., & Betzel, R. F. (2016). Modular Brain Networks. *Annual Review of Psychology*, *67*(1), 613–640. doi: 10.1146/annurev-psych-122414-033634

Srivastava, M., Sharma, G. D., Srivastava, A. K., & Kumaran, S. S. (2020). What's in the brain for us: a systematic literature review of neuroeconomics and neurofinance. *Qualitative Research in Financial Markets*, *12*(4), 413–435. doi: 10.1108/QRFM-10-2019-0127

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, *15*(1), 1929–1958.

Sáenz-Royo, C., Chiclana, F., & Herrera-Viedma, E. (2022). Functional Representation of the Intentional Bounded Rationality of Decision-Makers: A Laboratory to Study the Decisions a Priori. *Mathematics*, *10*(5), 739. doi: 10.3390/math10050739

Tabari, N., Seyeditabari, A., Peddi, T., Hadzikadic, M., & Zadrozny, W. (2019). A Comparison of Neural Network Methods for Accurate Sentiment Analysis of Stock Market Tweets. In C. Alzate et al. (Eds.), *ECML PKDD 2018 Workshops* (Vol. 11054, pp. 51–65). Cham: Springer International Publishing. (Series Title: Lecture Notes in Computer Science) doi: 10.1007/978-3-030-13463-1_4

Tanaka, T., Yamamoto, T., & Haruno, M. (2017). Brain response patterns to economic inequity predict present and future depression indices. *Nature Human Behaviour*, *1*(10), 748–756. doi: 10.1038/s41562-017-0207-1

Tousignant, A., Lemaître, P., Precup, D., Arnold, D. L., & Arbel, T. (2019). Prediction of disease progression in multiple sclerosis patients using deep learning analysis of mri data. In *International conference on medical imaging with deep learning* (pp. 483–492).

Truscott, J., & Smith, M. S. (2004). Acquisition by processing: A modular perspective on language development. *Bilingualism: Language and Cognition*, *7*(1), 1–20. doi: 10.1017/S1366728904001178

Tsukiura, T., Shigemune, Y., Nouchi, R., Kambara, T., & Kawashima, R. (2013). Insular and hippocampal contributions to remembering people with an impression of bad personality. *Social Cognitive and Affective Neuroscience*, *8*(5), 515–522. doi: 10.1093/scan/nss025

Tzilivaki, A., Kastellakis, G., & Poirazi, P. (2019). Challenging the point neuron dogma: FS basket cells as 2-stage nonlinear integrators. *Nature Communications*, *10*(1), 3664. doi: 10.1038/s41467-019-11537-7

V., P., & Bhattacharyya, A. (2022). Human emotion recognition based on time–frequency analysis of multivariate EEG signal. *Knowledge-Based Systems*, *238*, 107867. doi: 10.1016/j.knosys.2021.107867

Van Kesteren, M. T. R., & Meeter, M. (2020). How to optimize knowledge construction in the brain. *npj Science of Learning*, *5*(1), 5. doi: 10.1038/s41539-020-0064-y

Walczak, S. (2001). An Empirical Analysis of Data Requirements for Financial Forecasting with Neural Networks. *Journal of Management Information Systems*, *17*(4), 203–222. doi: 10.1080/07421222.2001.11045659

Wang, C., Shen, D., & Li, Y. (2022). Aggregate Investor Attention and Bitcoin Return: The Long Short-term Memory Networks Perspective. *Finance Research Letters*, *49*, 103143. doi: 10.1016/j.frl.2022.103143

Wang, K.-J., & Zheng, C. Y. (2019). Toward a Wearable Affective Robot That Detects Human Emotions from Brain Signals by Using Deep Multi-Spectrogram Convolutional Neural Networks (Deep MS-CNN). In *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)* (pp. 1–8). New Delhi, India: IEEE. doi: 10.1109/RO-MAN46459.2019.8956382

Wang, Y., Dai, C., Shao, Y., Wang, C., & Zhou, Q. (2022). Changes in ventromedial prefrontal cortex functional connectivity are correlated with increased risk-taking after total sleep deprivation. *Behavioural Brain Research*, *418*, 113674. doi: 10.1016/j.bbr.2021.113674

Wang, Y., & Sun, Q.-Q. (2021). A long-range, recurrent neuronal network linking the emotion regions with the somatic motor cortex. *Cell Reports*, *36*(12), 109733. doi: 10.1016/j.celrep.2021.109733

Weissler, E. H., Naumann, T., Andersson, T., Ranganath, R., Elemento, O., Luo, Y., … Ghassemi, M. (2021). The role of machine learning in clinical research: transforming the future of evidence generation. *Trials*, *22*(1), 537. doi: 10.1186/s13063-021-05489-x

Willett, F. R., Avansino, D. T., Hochberg, L. R., Henderson, J. M., & Shenoy, K. V. (2021). High-performance brain-to-text communication via handwriting. *Nature*, *593*(7858), 249–254. doi: 10.1038/s41586-021-03506-2

Yang, J., & Counts, S. (2010). Predicting the Speed, Scale, and Range of Information Diffusion in Twitter. *Proceedings of the International AAAI Conference on Web and Social Media*, *4*(1), 355–358. doi: 10.1609/icwsm.v4i1.14039

Yang, S. (2021). A novel study on deep learning framework to predict and analyze the financial time series information. *Future Generation Computer Systems*, *125*, 812–819. doi: 10.1016/j.future.2021.07.017

Yao, H., Zhang, X., Zhou, X., & Liu, S. (2019). Parallel Structure Deep Neural Network Using CNN and RNN with an Attention Mechanism for Breast Cancer Histology Image Classification. *Cancers*, *11*(12), 1901. doi: 10.3390/cancers11121901

Yarushev, S. A., & Averkin, A. N. (2018). Time Series Analysis Based on Modular Architectures of Neural Networks. *Procedia Computer Science*, *123*, 562–567. doi: 10.1016/j.procs.2018.01.085

Yoo, S. B. M., & Hayden, B. Y. (2018). Economic Choice as an Untangling of Options into Actions. *Neuron*, *99*(3), 434–447. doi: 10.1016/j.neuron.2018.06.038

Yu, J., Mamerow, L., Lei, X., Fang, L., & Mata, R. (2016). Altered Value Coding in the Ventromedial Prefrontal Cortex in Healthy Older Adults. *Frontiers in Aging Neuroscience*, *8*. doi: 10.3389/fnagi.2016.00210

Zanc, R., Cioara, T., & Anghel, I. (2019). Forecasting Financial Markets using Deep Learning. In *2019 IEEE 15th International Conference on Intelligent Computer Communication and Processing (ICCP)* (pp. 459–466). Cluj-Napoca, Romania: IEEE. doi: 10.1109/ICCP48234.2019.8959715

Zey, M. A. (2015). Rational choice and organization theory. In *International Encyclopedia of the Social & Behavioral Sciences* (pp. 892–895). Elsevier.

Zhang, J., Yin, Z., Chen, P., & Nichele, S. (2020). Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review. *Information Fusion*, *59*, 103–126. doi: 10.1016/j.inffus.2020.01.011

Zhang, T., Schoene, A. M., Ji, S., & Ananiadou, S. (2022). Natural language processing applied to mental illness detection: a narrative review. *npj Digital Medicine*, *5*(1), 46. doi: 10.1038/s41746-022-00589-7

Zhang, Y., Chu, G., & Shen, D. (2021). The role of investor attention in predicting stock prices: The long short-term memory networks perspective. *Finance Research Letters*, *38*, 101484. doi: 10.1016/j.frl.2020.101484

Zhong, X., & Enke, D. (2019). Predicting the daily return direction of the stock market using hybrid machine learning algorithms. *Financial Innovation*, *5*(1), 24. doi: 10.1186/s40854-019-0138-0

Zhou, Q., Zhou, C., & Wang, X. (2022). Stock prediction based on bidirectional gated recurrent unit with convolutional neural network and feature selection. *PLOS ONE*, *17*(2), e0262501. doi: 10.1371/journal.pone.0262501

Zhu, L., & Laptev, N. (2017). Deep and Confident Prediction for Time Series at Uber. In *2017 IEEE International Conference on Data Mining Workshops (ICDMW)* (pp. 103–110). New Orleans, LA: IEEE. doi: 10.1109/ICDMW.2017.19

Appendices

Appendix

The MCoRNNMCD-ANN predictions aim to forecast price movements, and the outcomes are consistent with the actual movements. Table B3 compares the original price values with the corresponding predicted prices. Compared to the original prices, cell colouring is utilised for the classification and price fluctuations columns to emphasise whether the predicted prices indicate an upward or downward movement.

The MCoRNNMCD-ANN model effectively predicts price movements by aligning its predictions with the trends observed in the original prices. When the original prices indicate an upward movement, the model's predictions follow suit by suggesting higher prices, as indicated by the green-filled cells in the table. Similarly, when the original prices show a downward movement, the model's predictions often indicate lower prices, as denoted by the red-filled cells. This alignment between the original prices and the corresponding predicted prices showcases the model's ability to capture and replicate the trends present in the market, providing valuable insights into future price movements.

Table B1: MCoRNNMCD-ANN Predicted Price Fluctuations.

Original price	Predicted price	Classification	Price fluctuations
0.88414	0.883698	1	Up
0.86065	0.860085	0	Down
0.87611	0.876166	1	Up
0.85411	0.853859	0	Down
0.89866	0.899046	1	Up
0.89988	0.901466	1	Up
0.85953	0.859448	0	Down
0.86729	0.867039	1	Up
0.88284	0.882783	1	Up

More specifically:

- Original price: This column represents the actual or original price values.
- Predicted price: This column displays the predicted price values.
- Classification: This column indicates the classification of price movement based on the predicted values. The classification is represented by the colour-filled cells.
- Price fluctuations: This column describes the direction of price movement. Based on the classification, it shows whether the price is going up or down. The direction is also represented by the colour-filled cells.