International Journal of Innovative Research in Computer Science and Technology (IJIRCST)
ISSN (Online): 2347-5552, Volume-12, Issue-4, July 2024
https://doi.org/10.55524/ijircst.2024.12.4.15
Article ID IJIRD-1347, Pages 92-98
www.ijircst.org

# Exploring Anomaly Detection and Risk Assessment in Financial Markets Using Deep Neural Networks

**Bingxing Wang[1], Yuxin Dong [2], Jianhua Yao[3], Honglin Qin[4], and Jiajing Wang[5]**

[1]Illinois Institute of Technology, Chicago, USA
[2] Wake Forest University, Winston-Salem, USA
[3]Trine University, Phoenix, USA
[4]Stevens Institute of Technology, Hoboken, USA
[5]Columbia University, New York, USA

Correspondence should be addressed to Bingxing Wang; wanguo0528@gmail.com

Received 25 June 2024; Revised 12 July 2024; Accepted 26 July 2024

Copyright © 2024 Made Bingxing Wang et al. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**ABSTRACT**—In this paper, deep learning technology, along with a Gated Recurrent Unit (GRU) combined with an attention mechanism, is used to enhance the recognition ability and risk assessment accuracy of abnormal trading behavior in financial markets. The GRU effectively solves the problem of gradient vanishing in traditional recurrent neural networks through its unique gated structure, allowing the model to learn more stable and effective feature representations in long sequence data. On this basis, the contextual attention (CA) module in the attention mechanism is introduced, enabling the model to automatically learn and assign different weights to various parts of the input sequence. Combined with bidirectional GRU and the attention mechanism, the model can not only capture temporal dependencies in the sequence but also highlight the key features that affect market anomalies, thus improving the model's ability to understand complex market dynamics.

**KEYWORDS**—Deep Learning; GRU, Attention Mechanism, CA Module; Anomaly Detection

## I. INTRODUCATION

By merging sophisticated machine learning advancements, this manuscript integrates a Gated Recurrent Unit (GRU) augmented with a consideration-focused mechanism to establish a potent and precise market oversight framework. This endeavor aims to proficiently detect unusual trading activities and meticulously assess underlying risks within intricate financial ecosystems.

Prevailing techniques for monitoring financial markets frequently encounter boundaries set by manually crafted regulations and a lack of intricate pattern discernment, rendering them ill-equipped to grapple with the multifaceted nature of temporal financial datasets—marked by extensive dimensions, non-linear traits, and ever-changing dynamics. Conversely, deep learning methodologies, specifically recurrent neural networks (RNNs), have demonstrated immense promise in parsing sequential information. Nonetheless, vanilla RNNs confront the obstacle of vanishing gradients when dealing with long data sequences, thereby curtailing their learning prowess. To surmount this impediment, our work adopts the GRU, which ingeniously manipulates 'reset' and 'update' gates to determine what data to forget or retain, thereby mastering long-term dependencies and excelling in deciphering financial time series.

Capitalizing on this sturdy groundwork, we proceed to incorporate a consideration-centric approach, an ingenious scheme empowering the model to spontaneously recalibrate its focus based on varying segments' contributions to the final forecast. The essence of attention diverges from uniform input valuation; instead, it allocates emphasis in proportion to each input's relevance to the immediate task. Applied to anomaly detection and risk assessment in financial markets, this empowers the model to instinctively filter out market cues vital for anomaly detection and risk assessment, sidestepping extraneous disturbances. This flexible accentuation of features markedly bolsters the model's accuracy and resilience.

Employing a GRU with a bi-directional traversal, coupled with an emphasis module, fosters an exhaustive comprehension of chronological data. The bi-directional facet empowers the model to reflect not only on past market maneuvers but also to proactively incorporate impending data, providing a comprehensive view for anomaly recognition. Embedding the emphasis mechanism equips our field of view with precision optics, assuring the model zeros in on the market's most revelatory movements. Such a dual-pronged strategy vastly augments the model's grasp of intricate market rhythms and responsiveness, enabling it to detect market fluctuations promptly, meticulously gauge affiliated risks, and relay invaluable alerts to financial stewards and supervisors. It not merely stretches the theoretical horizons of deep learning usage in finance but also equips practitioners with a formidable implement, poised to occupy a pivotal position in future surveillance frameworks. By pinpointing anomalies with precision and adeptly appraising risks, the model facilitates timely regulatory interventions, safeguarding investors against market upheavals. Furthermore, it paves fresh avenues for financial entities to refine risk management blueprints and enhance decision-making caliber. As technology evolves, confidence grows that deep learning-powered market oversight infrastructures will progressively bolster financial stability and foster economic prosperity.

### *A. Research status of outlier detection*

The statistically grounded frameworks for anomaly detection confront obstacles, chiefly due to the obscure nature of real-world data distributions, rendering them elusive to perfect alignment with any pre-established mathematical schema. This, in turn, impacts the trustworthiness of inferred anomaly reports. Further exacerbating this challenge, the ascension in data dimensionality renders it a near-impossible feat to accurately delineate the data's dispersion profile.

Knorr demarcated distance-oriented anomaly recognition methodologies, positing that data points exhibiting solitude, characterized by a scarcity of proximate points, should be flagged as anomalies [1]. This premise was subsequently expanded upon by Ramaswamy and colleagues in 2009, who pinpointed the top n data instances as anomalies through a summation of their distances to their k closest neighbors [2]. Despite the Intuitive appeal and facile comprehension of distance-centric techniques, they exhibit sensitivity to parameter tuning and incur substantial computational overhead in scenarios involving high-dimensional datasets.

Clustering methodologies maintain a close kinship with the aforementioned density and distance-centric frameworks, isolating data points distant from cluster nuclei or inhabiting sparse territories as anomalies. The DBSCAN protocol exemplifies this application. Both the efficacy of clustering routines and the precision of anomaly identification hinge heavily on the precision of distance metrics, potentially restricting their pragmatic utility.

Moreover, extensive research has ventured into harnessing avant-garde technologies – neural networks, Support Vector Machines (SVM), and Bayesian networks – to model and forecast sequential data patterns[3]. Anomalies are then discerned by contrasting the discrepancy between the models' predicted outputs and the empirical observations[4], thereby enriching the anomaly detection landscape with a multiplicity of strategic approaches.

### *B. Research status of anomaly detection in financial markets*

Anomaly detection and risk assessment in financial markets have seen significant advancements through the application of deep learning technologies. Erfani et al. [5] explored the utilization of a linear one-class SVM in combination with deep learning for high-dimensional anomaly detection, demonstrating the potential of integrating traditional machine learning methods with deep neural networks. Chuah et al. [6] introduced the GMean model, which combines a semi-supervised GRU and K-means clustering for predicting transcription factor binding sites, showcasing the flexibility of GRUs in handling temporal data. In the realm of financial risk prediction, various studies have applied machine learning techniques to forecast loan defaults and evaluate credit risks. Yu and Zhu [7] developed a data-driven approach using machine learning algorithms to predict default risk in peer-to-peer lending, highlighting the application of machine learning in financial risk prediction. Emekter et al. [8] evaluated credit risk and loan performance using various predictive models, while Aksakalli et al. [9] employed Random Forests for risk assessment in social lending, illustrating the efficacy of ensemble methods in financial risk evaluation.

Byanjankar et al. [10] implemented neural networks to predict credit risk in peer-to-peer lending, underscoring the superior performance of deep learning models over traditional methods. Cahuantzi et al. [11] conducted a comparative analysis of LSTM and GRU networks for learning symbolic sequences, revealing insights into the strengths and weaknesses of each architecture, which are pertinent for selecting the appropriate model for specific tasks. Tu et al. [12] utilized a GRU-Informer model for real-time prediction of rate of penetration (ROP) in drilling operations, highlighting the potential of GRUs in real-time data processing. Liu et al. [13] presented a novel approach in few-shot learning for product description calibration, emphasizing the role of deep learning in handling sparse data scenarios, a common challenge in financial market analysis. The integration of GRUs with attention mechanisms, as well as the application of various deep learning models, has significantly enhanced the ability to detect anomalies and assess risks in financial markets. These advancements provide practical tools for financial market oversight and risk management, extending the theoretical boundaries of deep learning applications and offering substantial practical value for financial entities and regulators.

## II. THEORETICAL BASIS

Several studies have explored the use of recurrent neural networks (RNNs) and their variants to address the challenges posed by financial time series data. Xu et al. [14] investigated financial risk behavior prediction using deep learning and big data, highlighting the efficacy of advanced neural networks in capturing complex patterns within financial datasets. Their work underscores the potential of deep learning models, particularly Gated Recurrent Units (GRUs), in enhancing the accuracy of risk assessments.

In the realm of high-frequency trading, Sun et al. [15] demonstrated the integration of Long Short-Term Memory (LSTM) networks with Extreme Value Theory (EVT) to manage financial risk. This approach effectively handles the volatility and extremities inherent in high-frequency trading data, providing a robust framework for risk management.

The optimization of natural language processing (NLP) models through multimodal deep learning has also garnered attention. Sun et al. [16] focused on improving NLP models by incorporating multimodal data, which can be pivotal in financial market analysis where textual data from news and reports significantly influence market dynamics. Similarly, Wang et al. [17] advanced a multimodal deep learning architecture for image-text matching, a technique that can be adapted for analyzing financial reports and associated visual data, thereby enriching the data inputs for financial models.

Further contributions have been made in the area of emotional analysis using large language models, as explored by Yang et al. [18]. Their work, although primarily focused on emotional analysis, provides valuable insights into the application of large language models in understanding sentiment and its impact on financial markets.

### *A. GRU*

The configuration of the GRU unit is depicted in Figure 1. Our advanced model unites a long short-term memory network's admission control and a disregard control, integrating these dual functionalities concurrently.

It encompasses a dual-state system: one for retaining information, referred to as memory, and another for holding data, termed storage.

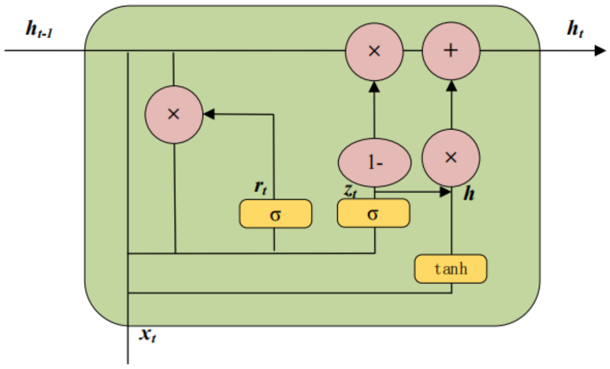



Figure 1: GRU Model structure

The educational procedure was bifurcated into a pair of phases. Initially, a framework was fabricated with the aim to transmute sequences of disparate lengths into sequences of a standardized extent for inscription. Sequentially, a subsequent framework was devised, tasked with the inverse operation - reverting this standardized-length sequence back to its original, variable-length configuration.

The task of the inscription is executed by employing a Recurrent Neural Network (RNN), accountable for scrutinizing each elemental component of the input sequence X in a consecutive manner. Every instance a novel constituent within the sequence is examined, the RNN readjusts its concealed state. Throughout the entire ingress procedure, the RNN amasses these concealed states, culminating with the occurrence of the sequence's termination cue, and consolidates them into a vector denoted as c.

The phase of decryption engages another RNN model, meticulously tailored for prognosticating and generating the succeeding output constituent x, guided by the procured concealed state c. Reflecting the abridged explanation of RNNs, the concealed state of the decryption network at the ongoing temporal step t is contingent upon the state $h_{\langle t-1\rangle}$ from the antecedent step, fused with the intelligence garnered from c. This latter is computed by employing a distinct algorithm, which incrementally materializes the output sequence through a stepwise generation process.

$$h_{\langle t\rangle} = \sigma\left(w_{wh}h_{\langle t-1\rangle} + w_{wy}y_{t-1} + w_{wc}c + b\right) \quad (1)$$

Likewise, the resultant at temporal instant t can be attained:

$$y_t = w'_{wh}h_{\langle t\rangle} + w'_{wy}y_{t-1} + w'_{wc}c + b \quad (2)$$

Should you require forecasting a categorization issue, it may be addressed via the utilization of an activation function's invocation, alongside dynamically adjusting recollection and discarding latent elements:

$$r_j = \sigma([W_r x]_j + [U_r h_{\langle t-1\rangle}]_j) \quad (3)$$

$W_r$ and $U_r$ symbolize the entirely crimson matrix undergoing acquisition, while the analogous refreshment gate $z_j$ is delineated as:

$$z_j = \sigma([W_z x]_j + [U_z h_{\langle t-1\rangle}]_j) \quad (4)$$

Under this scenario, the veiled condition tied to the ongoing j constituent is characterized as:

$$h_j^{\langle t\rangle} = z_j h_j^{\langle t-1\rangle} + (1 - z_j)\tilde{h}_j^{\langle t\rangle} \quad (5)$$

Among them:

$$\tilde{h}_j^{\langle t\rangle} = \phi([Wx]_j + [U(r \circ h_{\langle t-1\rangle})]_j) \quad (6)$$

Upon the instance where the restart gate approximates zero, this signifies that the concealed status is set to mostly discard its prior condition, concentrating exclusively on the instantaneous input data for rejuvenation. Conversely, the modification gate assumes a moderating part in this progression, dictating the degree to which intelligence from the preceding concealed status is assimilated into the emerging concealed state. Given that every concealed unit is endowed with autonomous restart and modification gating machinery, they possess the capacity to individually concentrate on acquiring and seizing data linkages across disparate chronological scopes. More concretely, concealed units inclined towards apprehending instantaneous or brief-period correlations are distinguished by recurrently elicitation restart gates; Inversely, units adept at seizing prolonged-period correlations customarily manifest active modification-gate conduct.

### *B. CA module*

The concentration-directing mechanism steers the network towards emphasizing the nucleus components of the assignment, and subtly curtails the intrusion level of ancillary and disruptive data. This approach accomplishes a meticulous sifting and refinement of data, augmenting not just the velocity and precision of task accomplishment on a large scale, but also efficaciously alleviating the quandary of data surfeit, thereby enabling the network to dedicate more focus on excavating the elemental attributes of the task. Consequently, the directionality and productivity ratio of the learning process are fortified.

More concretely, the integration of the conduit concentration mechanism is purposed to empower the network to self-regulate, bestowing differential weights in accordance with the significance of separate conduits, spotlighting those traits pivotal for entity recognition, and concurrently minimizing the sway of unrelated data. Through the application of meticulously computed weights to individual conduits of the feature representation, the model becomes capable of capitalizing fully upon the conduit features most crucial for identification, directly fueling an upsurge in detection precision.

Within the CA protocol, the employment of mean pooling activity along the breadth and altitude of the input feature depiction not only seizes the essence of conduit hierarchy but also ingeniously melds spatial data to ascertain that the model can attend to both conduit attributes and spatial characteristics concurrently, achieving a dual reinforcement of conduit and spatial concentration. The layout of this construct is illustrated in Figure 2.

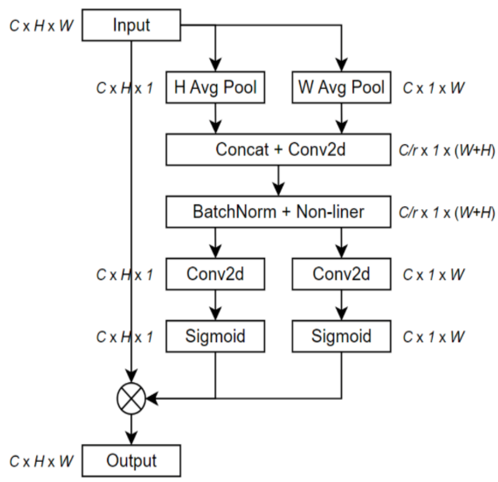



Figure 2: CA Structure Diagram

Observation of Figure 2 reveals that the CA submodule initiates by computing the mean of the incoming feature representation along separate axes of height and width:

$$z^h = \frac{1}{W} \sum_{0 \leq i < W} x(h, i) \quad (7)$$

$$z^w = \frac{1}{H} \sum_{0 \leq j < H} x(j, w) \quad (8)$$

In the formula, x represents the input feature map, which belongs to the space $\mathbb{R}C^{C\times H\times W}$. Where C, H and W represent the number, height and width of channels respectively. By performing an average pooling operation on x in the height dimension, we obtain the output $z^h$, which has the shape $\mathbb{R}^{C\times H\times 1}$ ; The same average pooling along the width dimension yields $z^w$ with the dimensions $\mathbb{R}^{C\times 1\times W}$. Following this, the duo of condensed outcomes are concatenated in a sequential fashion along the conduit axis. Thereupon, a dual-dimensional convolution stratum, batch normalization, and an energizing function are exerted onto the fused outcome to give rise to the ultimate yield, algebraically rendered as:

$$f = \delta\left(BN\left(Conv2d^{1\times 1}(z^h, z^w)\right)\right) \quad (9)$$

In the formula description, the operation (,) represents the concatenation along the channel dimension; The $onv2d^{1\times 1}$ convolution represents the operation using a two-dimensional convolution kernel of size 1×1; BN stands for batch normalized processing steps; $\delta$ indicates the type of nonlinear activation function to be applied; The middle feature graph $f$ has the dimension $\mathbb{R}^{C/r\times 1\times (H+W)}$, where $r$ is used to adjust the proportion of channel reduction as in the SE module. Subsequently, the intermediate result $f$ is split into two components $f^h$ and $f^w$, which undergoes two dimensional convolution operations respectively and are transformed by the Sigmoid function to produce two attention vectors $g^h$ and $g^w$. Ultimately, each of these dual concentration vectors is element-wise multiplied with the pristine input $x$ to procure the conclusive feature diagram output $y$, with the mathematical encapsulation thereof being outlined as follows:

$$f^h, f^w = Split(f) \quad (10)$$

$$g^h = \sigma\left(Conv2d^{1\times 1}(f^h)\right) \quad (11)$$

$$g^w = \sigma\left(Conv2d^{1\times 1}(f^w)\right) \quad (12)$$

$$y = g^h \cdot g^w \cdot x \quad (13)$$

Where $Split$ is the resolution tensor; $\sigma$ is the Sigmoid function. A further embodiment of the composite concentration schema uniting conduit and spatial aspects manifests in CA, distinguished by its ingenuity in deftly weaving spatial attributes into conduit characteristics, thereby accomplishing a streamlined fusion of data. CA is architected with an emphasis on minimalism and adaptability, empowering it to harmoniously blend into the canonical architectural modules of portable networks, enhancing model efficacy sans incurring excessive computational strain.

Empirical evidence illustrates that CA not merely excels in the realm of image categorization undertakings but also exhibits remarkable prowess in downstream deployments, including object recognition and semantic partitioning, underscoring its broad applicability and augmentative impact.

## III. GRU MODEL FUSED WITH CA MODULE

This manuscript unites the profound study paradigm rooted in the concentration-directing mechanism's CA module alongside GRU (Gated recurrent unit) to delve into its prospective deployment in identifying anomalies and gauging risks within the fiscal domain, with the architectural depiction of the model presented in Figure 3.

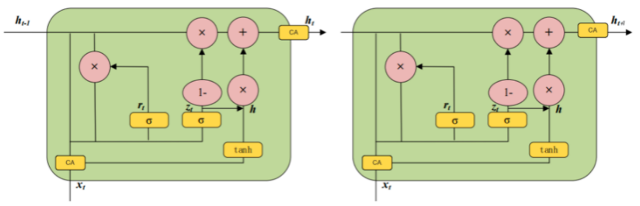


Figure 3: Deep Learning CA+GRU Model

The framework initiates by deploying GRU to metabolize sequential temporal information, thereby seizing the mutable facets of fiscal landscapes. GRU adeptly tackles the predicament of extended chronological dependencies through its idiosyncratic gating infrastructure (comprising the reset and refresh gates), assuring the model retains salient historical wisdom while recalibrating to incorporate novel inputs.

Supplementing GRU, the Context-Sensitive Focus (CA) submodule was incorporated to dynamically allocate significance coefficients to diverse segments of the ingress chronology. Via educating itself on the interconnectedness and primacy of the input dataset, the CA submodule empowers the model to concentrate on facets that are supremely pivotal for anomaly sleuthing and hazard appraisal, concurrently mitigating disruptive noise interferences, as illustrated in Figure 4.

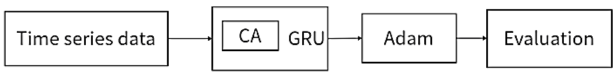



Figure 4: Flow chart of the experiment

The CA submodule equips the model with the faculty to adaptively zero in on information of utmost relevance to the undertaking, proficiently sift out extraneous disturbances, and amplify the model's discernment of market irregularities. By fusing GRU's sequential data handling prowess with CA's astute attention apportionment, the model attains heightened precision in encapsulating the intricate kinetics of fiscal ecosystems, thereby boosting the veracity of anomaly sleuthing and hazard appraisals.

Concurrently, the concentration-directing apparatus furnishes a conduit for instinctively grasping the model's rationale behind decision-making. Via meticulous data conditioning, superfluous computations are curtailed, enabling the model to function more expediently without sacrificing excellence, rendering it apt for voluminous datasets and real-time surveillance contexts.

## IV. EXPERIMENTAL ANALYSIS

### *A. Data preprocessing*

Financial time data contain structures that are trend, cyclical, and time-dependent. The differencing operation calculates the difference between the adjacent values of the series to eliminate the trend and periodicity and obtain a stationary time series. The dataset embraces the S&P 500 Index extracted from American equity market figures, encompassing parameters such as the initial price, terminal price, transaction volume, peak value, and nadir value, amongst others. Data is harvested at a frequency of one instance per trading session,'

### *B. Experimental Setup*

The chronological fiscal data corpus is divided into two partitions: a training set and a test set, wherein 70% of the total data is dedicated to training the model, and the remaining 30% is used for evaluating its performance. Due to measurement inconsistencies, the raw dataset is filled with signal disturbances and numerous deviant readings. To mitigate measurement discrepancies and reduce computational deviations caused by these outliers, it is imperative to homogenize the raw data. This process includes gap value rectification, uniformity adjustments, consolidation of various datasets, and chronology scaling, all aimed at ensuring the effectiveness of the model's training.

Employing the lattice quest methodology, the consignment magnitude is calibrated to 128 units, alongside a tally of 3 for the enigmatic stratum constituents. The Mean Squared Deviation Root (MSDR) is a popular measure for assessing the predictive efficacy of chronological financial sequences. It quantifies the disparity between the anticipated value and the empirical outcome, manifesting as a positive real number. Here, a lower score indicates heightened predictive precision. Its computation is as follows:

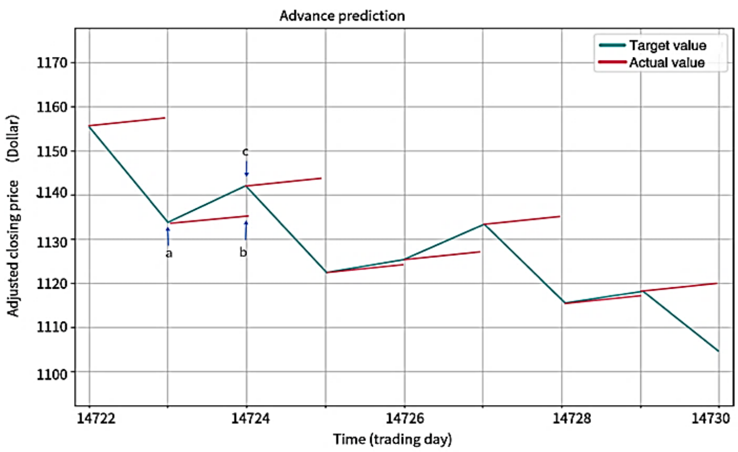
$$RMSE = \sqrt{\frac{1}{n}\Sigma_{i=1}^{n}(x_i - \hat{x}_i)^2} \tag{14}$$

Where, $n$ is the total number of observations, $x_i$ is the target value, and $\hat{x}_i$ is the actual value.

### *C. Experimental results*

The algorithm is trained using the training subset and then assessed on the test subset, with the benchmark being its ability to identify atypical trading behavior. Concurrently, the risk intensity is assessed based on the fluctuations forecasted by the algorithm. The algorithm's input data consists of previous observations of the adjusted terminal value, while its output, or labeling, represents the impending valuation of the adjusted terminal price.

The rectified concluding quotation of the S&P 500 equity index epitomizes a solitary-variable chronological progression forecasting challenge, entailing multi-stage anticipation through a profound learning architecture, as visually elucidated in Figure 5.




[18]

Figure 5: The CA+GRU model is used for one-step ahead prediction

Chronological progression is delineated along the horizontal expanse of the chart, demarcating the passage of individual trading epochs, whereas the perpendicular axis charts the deviation in the rectified terminal pricing. The azure line depicts the trajectory of the rectified closure, whereas the crimson line illustrates the succeeding (viz., the successive trading epoch) rectified conclusion predicated upon a mono-stage prognosis methodology. By way of illustration, the punctuate indicators $a$ and $c$ correlate respectively to the veridical rectified closure valuations on epochs 14723 and its immediate successor, epoch 14724. The distinctive juncture $b$ on the scarlet line embodies the forecasted rectified closure price for epoch 14723 inferred from the bourse data pertaining to epoch 14724. Put differently, the dot $b$ symbolizes a prospective stride, a conjecture of the succeeding day's terminal price hinged upon the antecedent day's intelligence. The congruent rationale and modus operandi are applicable for recursively forecasting and mapping the rectified closure prices for all ensuing epochs.

Accordingly, the MSDR indicator is invoked to gauge the predictive efficacy of the multi-phase forecasting prototype for the inaugural trading epoch, the subsequent trading epoch, and the tertiary trading epoch, deploying both the GRU prototype and the GRU profound learning prototype amalgamated with the CA submodule. The empirical consequences are tabulated in Table 1.

Table 1:Model Performance Comparison

| Methods | 1 | 2 | 3 |
|---|---|---|---|
| GRU | 13.28 | 13.27 | 13.29 |
| CA+GRU | 9.76 | 9.78 | 9.74 |

This graphical representation illustrates the contrasting performances of a dual set of models in gauging fiscal market hazards. Evident from the tabular data, under the RMSE index scrutiny, the prognostic efficacy of the GRU prototype remains fairly consistent over the triad of trading epochs. Conversely, the GRU profound learning schema fused with the CA submodule witnesses a marked amelioration in its forecasting prowess.

Through juxtaposing the datasets from these twin frameworks, we discern that the CA+GRU schema registers a conspicuous upsurge in prediction precision, notably during the anticipation of the inaugural and tertiary trading sessions, with the discrepancy value declining approximately by a third with respect to the GRU prototype. This evidences that the CA submodule potentiates the GRU model's learning faculty significantly, thereby facilitating an enhancement in the exactness of protracted forecasts.

The empirical outcomes manifest that the GRU schema augmented with CA surpasses the foundational model in pinpointing atypical transactions and hazard appraisals, with special prominence in intricate and disturbance-laden data ecologies. This superiority is rooted in the attention dispersion chart provisioned by the concentration-directing apparatus, which intuitively showcases the traits that captivate the model's focus, thereby reinforcing the model's explicability. Concurrently, via an exploratory dissection of the computational velocity and memory footprint data emanating from the model's engagement with voluminous datasets, it is corroborated that this hybrid construct bolsters resource deployment efficacy and markedly augments the practicability within real-world implementations.

## V. CONCLUSION

This paper focuses on leveraging advanced deep-learning methodologies to enhance the efficacy and precision of financial market oversight, particularly in detecting unconventional trading activities and estimating impending risks. Our research devises and implements an innovative method that integrates an attention-augmented Gated Recurrent Unit (GRU) to provide more precise and meticulous surveillance of complex fluctuations within financial markets. Current surveillance methodologies largely rely on pre-established regulations. Their limitations lie in disregarding the complexity and dynamism inherent in financial datasets and their inadequacy in swiftly adapting to emerging anomalous patterns in the marketplace. Conversely, deep learning technologies, notably the Recurrent Neural Networks (RNNs) family, have shown immense promise in handling sequential data, although the vanishing gradient problem confronting conventional RNNs in prolonged sequences hampers their efficacy. Therefore, we opt for the GRU as the foundational architecture, which mitigates the long-term dependency challenge through a single gating mechanism, thereby enhancing the learning proficiency and forecasting accuracy of the model when dealing with financial time series data.

By merging the bidirectional GRU with the CA submodule, the model is empowered to both review historical market behavior and predict upcoming trends, fostering a comprehensive understanding of chronological sequences. The combination of the bidirectional architecture with the attention mechanism enables the swift identification of critical dynamics within complex market intelligence. This capability is pivotal for the timely detection of market fluctuations and the accurate assessment of associated risks.

Our model not only advances the theoretical boundaries of deep learning technology in financial market supervision but also underscores its vast practical potential. It serves as a powerful tool for financial entities and regulators, enabling the prompt detection of market turbulence and facilitating timely regulatory intervention to safeguard investors' interests. Additionally, it aids financial bodies in adopting more targeted risk management strategies, refining investment approaches, and enhancing decision-making capabilities. As technology advances and model refinements continue, the deep learning-powered financial market surveillance system will evolve, playing an increasingly crucial role in maintaining fiscal stability and promoting steady economic growth. The outcomes of this investigation not only drive innovations in financial market surveillance technology but also contribute intellectual and operational strength to the development of the global financial security framework, ushering in a new era of more efficient and intelligent financial market oversight.

## CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

## REFERENCES

[1] E. M. Knorr and R. T. Ng, "Algorithms for Mining Distance-Based Outliers in Large Data-Sets," in Proceedings of the 24th International Conference on Very Large Databases, New York, Aug. 1998, pp. 392-403. Available from: https://www.vldb.org/conf/1998/p392.pdf

[2] S. Ramaswamy, R. Rastogi, and K. Shim, "Efficient algorithms of mining outliers from large datasets," in Proceedings of the ACM SIGMOD International Conference on Management of Data, Dallas, USA, 2000, vol. 29, no. 2, pp. 427-438. Available from: https://doi.org/10.1145/342009.335437

[3] X. Yan, W. Wang, M. Xiao, Y. Li, and M. Gao, "Survival prediction across diverse cancer types using neural networks," in Proceedings of the 2024 7th International Conference on Machine Vision and Applications, Mar. 2024, pp. 134-138. Available from: https://doi.org/10.1145/3653946.3653966

[4] Z. Liu, X. Xia, H. Zhang, and Z. Xie, "Analyze the impact of the epidemic on New York taxis by machine learning algorithms and recommendations for optimal prediction algorithms," in Proceedings of the 2021 3rd International Conference on Robotics Systems and Automation Engineering, May 2021, pp. 46-52. Available from: https://doi.org/10.1145/3475851.3475861

[5] S. M. Erfani, S. Rajasegarar, S. Karunasekera, et al., "High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning," Pattern Recognition, vol. 58, pp. 121-134, 2016. Available from: https://doi.org/10.1016/j.patcog.2016.03.028

[6] C. W. Chuah, W. He, and D.-S. Huang, "GMean—a semi-supervised GRU and K-mean model for predicting the TF binding site," Scientific Reports, vol. 14, no. 1, p. 2539, 2024. Available from: https://doi.org/10.1038/s41598-024-52933-4

[7] J. Yu and Y. Zhu, "A Data-Driven Approach to Predict Default Risk of Loan for Online Peer-to-Peer (P2P) Lending," in Proceedings of the Fifth International Conference on Communication Systems & Network Technologies, IEEE, 2015, pp. 609-613. Available from: https://doi.org/10.1109/CSNT.2015.25

[8] R. Emekter, Y. Tu, B. Jirasakuldech, et al., "Evaluating credit risk and loan performance in online Peer-to-Peer (P2P) lending," Applied Economics, vol. 47, no. 1, pp. 54-70, 2015. Available from: https://doi.org/10.1080/00036846.2014.962222

[9] V. Aksakalli, Malekipirbazari, et al., "Risk Assessment in Social Lending via Random Forests," Expert Systems with Applications, vol. 42, no. 10, pp. 4621-4631, 2015. Available from: https://doi.org/10.1016/j.eswa.2015.02.001

[10] A. Byanjankar, M. Heikkilä, and J. Mezei, "Predicting Credit Risk in Peer-to-Peer Lending: A Neural Network Approach," in Proceedings of the IEEE Symposium on Computational Intelligence, IEEE, 2015, pp. 719-725. Available from: https://doi.org/10.1109/SSCI.2015.109

[11] R. Cahuantzi, X. Chen, and S. Güttel, "A comparison of LSTM and GRU networks for learning symbolic sequences," in Science and Information Conference, Cham: Springer Nature Switzerland, 2023. Available from: https://doi.org/10.1007/978-3-031-37963-5_53

[12] B. Tu, et al., "Real-time prediction of ROP based on GRU-Informer," Scientific Reports, vol. 14, no. 1, p. 2133, 2024. Available from: https://doi.org/10.1038/s41598-024-52261-7

[13] Z. Liu, M. Wu, B. Peng, Y. Liu, Q. Peng, and C. Zou, "Calibration Learning for Few-shot Novel Product Description," in Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, July 2023, pp. 1864-1868. Available from: https://doi.org/10.1145/3539618.3591959

[14] K. Xu, Y. Wu, Z. Li, R. Zhang, and Z. Feng, "Investigating Financial Risk Behavior Prediction Using Deep Learning and Big Data," International Journal of Innovative Research in Engineering and Management, vol. 11, no. 3, pp. 77-81, 2024. Available from: https://doi.org/10.55524/ijirem.2024.11.3.12

[15] M. Sun, Z. Feng, Z. Li, W. Gu, and X. Gu, "Enhancing Financial Risk Management through LSTM and Extreme Value Theory: A High-Frequency Trading Volume Approach," Journal of Computer Technology and Software, vol. 3, no. 3, 2024. Available from: https://doi.org/10.5281/zenodo.12669410

[16] D. Sun, Y. Liang, Y. Yang, Y. Ma, Q. Zhan, and E. Gao, "Research on Optimization of Natural Language Processing Model Based on Multimodal Deep Learning," arXiv preprint arXiv:2406.08838, 2024. Available from: https://doi.org/10.48550/arXiv.2406.08838

[17] J. Wang, H. Zhang, Y. Zhong, Y. Liang, R. Ji, and Y. Cang, "Advanced Multimodal Deep Learning Architecture for Image-Text Matching," arXiv preprint arXiv:2406.15306, 2024. Available from: https://doi.org/10.1109/ICETCI61221.2024.10594167

[18] H. Yang, Y. Zi, H. Qin, H. Zheng, and Y. Hu, "Advancing Emotional Analysis with Large Language Models," Journal of Computer Science and Software Applications, vol. 4, no. 3, pp. 8-15, 2024. Available from: https://doi.org/10.5281/zenodo.12204513