SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Enhancing Risk Mitigation In Business Investments Using Deep Reinforcement Learning

Rakesh Kumar[1*], Dr.K. Sreedevi[2], V. Hemamalini[3], Dr.S. Rama Sree[4], Dr.Y. Shantharam[5], CH.V. Aruna[6]

[1*]Department of Computer Engineering & Applications, GLA University, Mathura, Uttar Pradesh, India.
E-mail: rakesh.kumar@gla.ac.in, https://orcid.org/0000-0002-3095-1005

[2]Assistant Professor, Department of Commerce, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, Chennai, Tamil Nadu, India. E-mail: sreedevicom@maher.ac.in, https://orcid.org/0009-0007-1245-6282

[3]Assistant Professor, Department of Electronics and Communication Engineering, New Prince Shri Bhavani College of Engineering and Technology, Chennai, Tamil Nadu, India. E-mail: hemamalini@newprinceshribhavani.com, https://orcid.org/0000-0002-8441-9321

[4]Professor, Department of Computer Science and Engineering, Aditya University, Surampalem, Andhra Pradesh, India.
E-mail: ramasree_s@adityauniversity.in, https://orcid.org/0000-0002-8771-6006

[5]Associate Professor, Civil Engineering, Mahendra Engineering College, Namakkal, Tamil, Nadu, India.
E-mail: shantharamy@mahendra.info

[6]Department of FED, Ramachandra College of Engineering, Eluru, Andhra Pradesh, India. E-mail: chittibommaaruna@gmail.com, https://orcid.org/0000-0003-3366-0585

*Corresponding author: Email: rakesh.kumar@gla.ac.in

Abstract

Deep reinforcement learning (DRL) has emerged as a game-changing technology in financial decision-making, offering new opportunities for enterprise risk management and investment management. Common risk assessment approaches such as static credit scoring, regression-based models, and traditional time series models like Long Short-Term Memory (LSTM) suffer from serious drawbacks in their ability to capture the dynamic, non-linear, and volatile nature of today's financial markets. This paper introduces a new framework named Adaptive Deep Reinforcement Learning Risk Mitigation (ADRL-RM), which combines Proximal Policy Optimization (PPO), Deep Q-Network (DQN), and multi-source data fusion—encompassing financial indicators (34 features), macroeconomic signals, market indices, and BERT-extracted news sentiment—to provide real-time, context-aware risk assessment for business investment portfolios. The architecture utilizes a Bidirectional LSTM (BiLSTM) encoder and Temporal Pattern Attention (TPA) for feature extraction; state representations are passed to an actor-critic module that jointly optimizes the risk-aversion parameters ($\lambda \in [0,1]$) as well as rebalancing horizons ($h \in [1,20]$ days). Empirical results on U.S. sector ETFs (n=12) and Dow Jones Industrial Average (DJIA) components (n=28) from 2003 to 2023 show that ADRL-RM outperforms baselines: PPO Mean-Variance achieves an annualized return of 15.7%, a Sharpe ratio of 0.887 (vs. 0.286 tangency; $\Delta SR=0.601$, $p<0.001$HAC-adjusted), and a max drawdown of 30.7% (a 25% reduction). The accuracy of enterprise risk classification is 81.4%, recall (high risk) is 87.6%, and RMSE is 139.29 ($R^2=0.91$). Ablation results include additive improvements, such as BiLSTM-TPA (+4.3% accuracy), sentiment fusion (+2.1% recall), and PPO (+3.1% accuracy compared to supervised models). The framework offers a management-friendly tool, converting AI signals into interpretable risk ratings (High/Medium/Low) and dynamic allocations, effectively supporting institutional investors.

Keywords Deep Reinforcement Learning, Proximal Policy Optimization, Portfolio Rebalancing, Bidirectional LSTM, Dynamic Risk Assessment, Multi-Source Data Fusion, Business Investment Management.

1. Introduction

Over the last 20 years, there have been unparalleled changes in the global financial landscape; the market has become more complex, more correlated across various asset classes, and more volatile, while new asset classes

like structured derivatives and cryptocurrencies have come to the fore [15]. Historically, enterprise risk management, as a key component of sound business investment strategy, has been tied to static models that tacitly rely on linear relationships between financial variables and risk outcomes [22]. The classic portfolio framework of Markowitz (1952), known as the Mean-Variance framework, remains fundamental in portfolio theory, but it is based on fixed distributional assumptions which make it inappropriate for the dynamic regimes facing modern markets [5]. One of the paradigms shifts in the integration of AI, and especially reinforcement learning (RL), is that instead of trying to fit a static model to history, an RL agent continually interacts with the environment and updates the decision policy as it processes feedback to maximize cumulative reward [9].

A key challenge is the need to integrate RL and deep learning in a way that is suitable for high-dimensional, heterogeneous state space inputs such as raw price histories, macroeconomic indicators, and unstructured textual sentiment [18]. In the context of this research, the use of deep reinforcement learning (DRL) for business investment risk mitigation is timely and of practical importance [19]. Financial institutions are increasingly under pressure to implement risk models that are interpretable, auditable, and adaptive [12], utilizing advanced policy gradient methods like Proximal Policy Optimization (PPO) to ensure stable learning in complex environments [7]. Simultaneously, portfolio managers require tools that adaptively align competing objectives return maximization and risk control—across different market regimes [4].

In the current literature, enterprise risk (ER) prediction and portfolio risk mitigation face three fundamental shortcomings [16]. First, traditional supervised learning methods such as gradient boosting (XGBoost), LSTM, and convolutional neural networks (CNN) do not have the adaptive feedback loop required to adjust investment strategies to changing market conditions [3]. Second, most DRL applications in finance only consider trading signal generation or direct weight estimation, omitting the optimization of risk-aversion parameters and rebalancing frequency [2]. Third, existing risk prediction tools are largely based on single-source financial data, lacking the information density of multi-source signals, such as news sentiment and macroeconomic data, which have proven predictive power [21]. Addressing these gaps through adaptive, AI-driven insights is essential for transforming modern risk mitigation and enterprise decision-making [1].

This paper aims to fill the above-mentioned gaps and develop an integrated ADRL-RM framework that integrates PPO and DQN together with multi-source data fusion for enterprise-level investment risk mitigation as follows: Propose a BiLSTM-TPA feature extraction module that can learn the bi-directionality of time series and dynamic attention weighting from different types of financial time series. Develop an RL environment where agents can learn optimal risk-aversion levels and rebalancing horizons that are consistent with the efficient frontier. Test the framework with real-world sector ETF and DJIA data and validate the results with benchmark methods, showing better risk-adjusted performance. Develop a management-oriented user interface that will take policy learning and provide a manageable risk rating and investment advice.

The primary contributions and findings of the ADRL-RM framework are summarized as follows:

- A novel framework is presented that unifies risk classification, portfolio weighting, and rebalancing timing through the application of Deep Reinforcement Learning (DRL) to optimize enterprise-wide financial management.
- A BiLSTM-TPA encoder is utilized to synthesize diverse data streams—including financial ratios, market indices, macroeconomic indicators, and BERT-based news sentiment—into a unified, semantically rich state representation.
- Rigorous rolling out-of-sample back-tests demonstrate that PPO-guided Mean-Variance (MV) optimization achieves an annualized return of 15.7% and a Sharpe ratio of 0.887 on DJIA components, yielding significant alpha ($\alpha \approx 9–11\%$ p.a.; $p < 0.001$) relative to static benchmarks.
- Statistically rigorous ablation studies confirm the specific performance gains attributed to individual architectural elements, including the BiLSTM, TPA, sentiment fusion, and the reinforcement learning policy.
- An interactive scoring system is provided to translate continuous portfolio values into discrete risk categories (low, medium, and high), facilitating interpretable decision-making for enterprise stakeholders.

The rest of this paper is organized as follows. The literature review in Section 2 covers relevant literature on risk mitigation using DRL and on portfolio optimisation. Section 3 describes the proposed ADRL-RM methodology,

both mathematical formulations and algorithmic specifications. Section 4 discuss experimental setup, datasets, initialisation of the parameters, and the performance results along with the ablation studies. Findings are discussed thoroughly in Section 5. The section 6 provides suggestions for further research.

2. Literature Review

The area of financial risk management has moved from models that are monolithic, linear, to those that are adaptive, non-linear and powered by Artificial Intelligence (AI). In the past the framework of modern investment theory was built on the basis of fixed distributional assumptions known as Mean-Variance [5]. However, traditional optimization methods do not consider higher order risks, and other methods developed to consider extreme market events have been proposed such as Conditional Value-at-Risk (CVaR) [13]. The advent of AI-powered analytics has transformed the landscape of predictive modeling, allowing for more intelligent investment strategies in the modern world, as it can process large amounts of data on an industrial scale [18].

Traditional time-series predictors and static credit scoring have been constrained and have given rise to the rise of Deep Reinforcement Learning (DRL) in financial decision-making. In DRL, fundamental research has shown that agents can successfully control complex environments at a human level by constantly interacting with data [9]. This feature can be particularly useful when designing automated trading strategies for stock trading and forming portfolio strategies that beat traditional benchmarks [11]. Some specialized libraries such as FinRL, have expanded the scope of how DRL is used to allow for the use of a standardized environment in quantitative finance and automated trading [14].

New research highlights the need for multi-source data fusion and architecture innovation. For example, deep learning models have been able to greatly improve the accuracy of financial investment choices [3]. The potential of deep learning for complex sequence data has been extended to environmental risk assessment [6] and power factor correction [8]. In addition, generative models such as Generative Adversarial Networks (GANs) have been used in high fidelity simulations, such as the modelling of damage in structures that can be applied to financial stress tests [10].

Explainable AI (XAI) is now emerging as a key enabler for strategic risk management, offering accurate and interpretable decision support for institutional risk managers [12]. The recently published literature emphasizes on the use of optimization methods together with reinforcement learning (RL) to improve the control of risk in trading strategies [2][17]. New advanced frameworks combine machine learning and DRL for more proactive decision-making for financial institutions [16]. The systems overcome the dynamism of the global markets and enable to shift the traditional risk modelling into adaptive forecasting systems [22].

The ongoing research landscape shows a change towards end-to-end trainable systems. New ideas apply DRL to dynamic rebalancing in order to achieve a smart risk-return trade-off in smart tangency portfolios [4]. The quantitative models have also been used to analyze individual investment decision factors to optimize the portfolio management and financial outcomes [15][19]. In order to mitigate operational risks, Pareto distribution analysis has been used to prioritize the losses of business operations [20]. In conclusion, the fusion of AI insights with automation is reshaping the landscape of risk mitigation, enabling dynamic solutions to enterprise-wide challenges. Finally, the power of AI-driven insights combined with automation is revolutionizing risk mitigation, facilitating adaptive responses to enterprise-level issues.

Through literature synthesis, there is a clear paradigm shift from retrospective and static risk assessment to prospective and dynamic frameworks for risk mitigation. Multi-source data fusion is the state-of-the-art integration of DRL (PPO and DQN) and it is currently the closest bridge to a high-dimensional feature extraction and a real-time decision-making process. The recent studies seem to agree that if AI models are to be institutionally viable in financial applications, these need to not only outperform traditional financial models in statistical terms, but also have high levels of interpretability and regulatory compliance.

3. Methodology

The ADRL-RM methodology has been organized into a multi-stage and end-to-end trainable pipeline, which aims to span the gap between the processing of high-dimensional data and strategic financial execution. The 34 financial features and macroeconomic signals are first aggregated into a heterogeneous data fusion layer, followed by the Bidirectional LSTM (BiLSTM) encoder and Temporal Pattern Attention (TPA) mechanism to generate state representations based on context. also used simultaneously by a Deep Q-Network (DQN) to classify the risks of a discrete enterprise and a Proximal Policy Optimization (PPO) actor-critic module for continuous portfolio optimization. The system optimizes the allocation of the investment for both risk-aversion parameters (λ) and dynamic rebalancing horizons (h), the system ensures that investment allocations are both responsive to real-time market volatility and aligned with long-term institutional risk thresholds.

Framework Overview

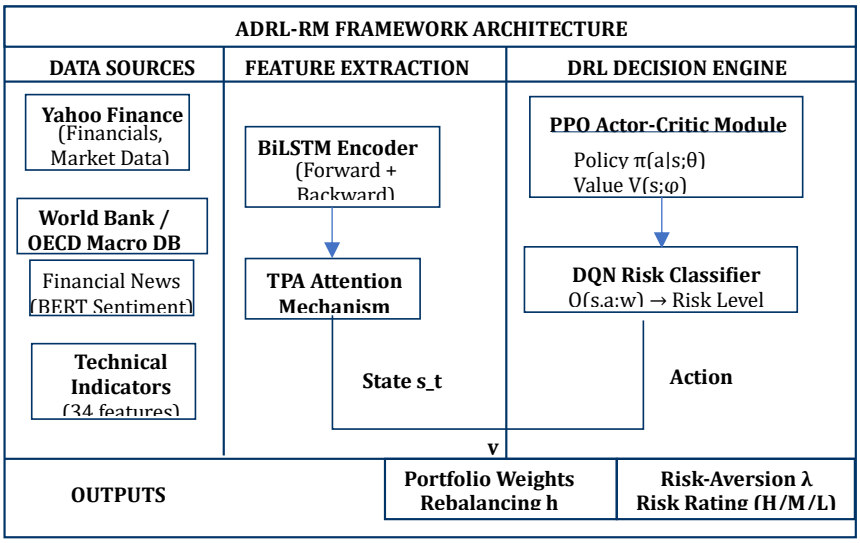

Figure 1: Overall architecture of the ADRL-RM framework

The Adaptive Deep Reinforcement Learning Risk Mitigation (ADRL-RM) framework as shown in figure 1 consists of three tightly coupled modules: (i) a Multi-Source Data Fusion and Preprocessing Pipeline, (ii) a BiLSTM-TPA Feature Extraction Encoder, and (iii) an Actor-Critic DRL Decision Engine implementing PPO and DQN policies. The framework's output are dynamic portfolio weights, risk-aversion indices, rebalancing horizons and categorical enterprise risk ratings for management consumption.

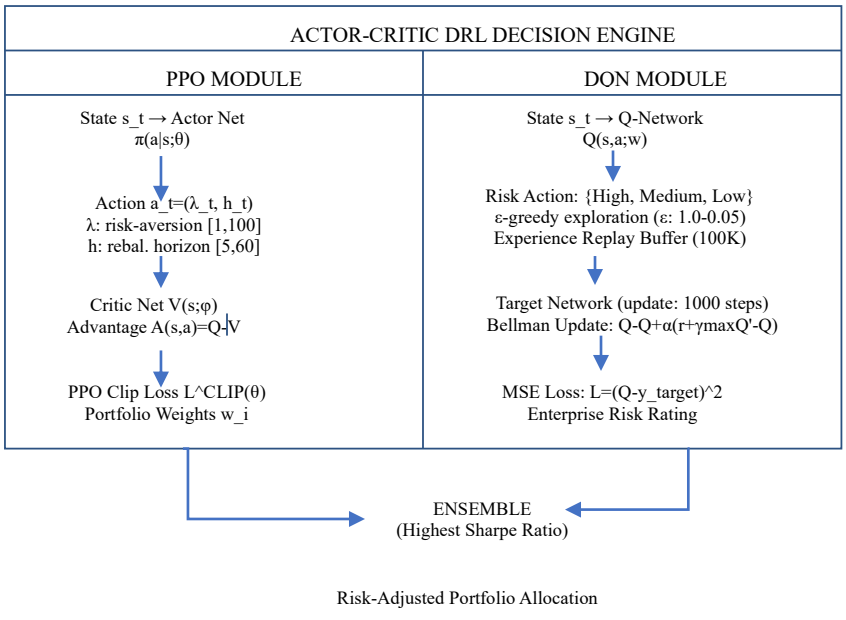

Figure 2: Actor-Critic DRL Decision Engine (PPO + DQN)

Actor-Critic DRL Decision Engine

The Actor-Critic DRL Decision Engine is shown in figure 2, where two reinforcement learning paradigms are seamlessly combined in a complementary way, namely the Proximal Policy Optimization (PPO) and Deep Q-Network (DQN) modules. The PPO module provides continuous action, e.g. risk aversion, rebalancing horizon, based on the current state. It uses a critic network to estimate the state value, and uses PPO clipped loss to ensure stable learning. The module then provides portfolio weights as output. The DQN module, however, chooses the risk actions (High/Medium/Low) by applying ε-greedy exploration. The learning process is stabilized by target network and experience replay buffer, and updated by the Bellman equation to minimize MSE loss. The DQN also provides an enterprise risk rating. Lastly, the PPO and DQN modules' outputs are consolidated in an ensemble decision-making process in which the action having the highest Sharpe ratio is chosen, thus determining the risk-adjusted portfolio allocation.

Mathematical Formulation

State Space and Reward Function

The Markov Decision Process (MDP) underlying ADRL-RM is defined by the tuple (S, A, P, R, γ).

State Space (S): The state vector $s_t \in$ Sis derived from the BiLSTM-TPA output. It integrates 34 features, including technical indicators, financial ratios, macroeconomic variables, and BERT-derived sentiment scores.

Action Space (A): The PPO module outputs a continuous action vector $a_t = (\lambda_t, h_t)$, where:

$\lambda_t \in \{1,2, \dots ,100\}$is the risk-aversion index, determining the portfolio's position on the efficient frontier.

$h_t \in \{5,6, \dots ,60\}$is the rebalancing horizon in trading days.

Reward Function (R): For the portfolio task, the reward is the log return over the holding period h_t is given by equation (1):

$$R_t = \ln\left(\frac{PV_t}{PV_{t-h_t}}\right) \quad (1)$$

where $PV_t$is the portfolio value at time t. For the DQN enterprise risk classification, the reward is the negative cross-entropy loss between predicted and true risk categories. PPO Objective Function

The PPO algorithm updates the policy $\pi_\theta$by maximizing the clipped surrogate objective given by Equation (2):

$$L^{CLIP}(\theta) = \mathbb{E}_t\left[\min\left(r_t(\theta)\hat{A}_t, \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t\right)\right] \quad (2)$$

Where: $r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)}$ is the probability ratio. $\hat{A}_t = Q(s_t, a_t) - V(s_t)$ is the estimated advantage function.$\epsilon = 0.2$ is the clipping hyperparameter.

DQN Bellman Update

The DQN module minimizes the temporal-difference (TD) error for risk classification using the Bellman equation (3):

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha\left[r_t + \gamma \max_{a'} Q'(s_{t+1}, a') - Q(s_t, a_t)\right] \quad (3)$$

Where: $\alpha = 0.001$(Learning rate). $\gamma = 0.99$(Discount factor)

Portfolio Optimisation Objectives

Based on the risk-aversion index λ_t, The system solves one of three convex optimization problems given by equations (4), (5) and (6):

Mean-Variance (MV)

$$\min_{w} \; w^T \Sigma \qquad (4)$$

subject to $w^T \mu \geq \mu(\lambda), \sum w_i = 1, w_i \geq 0$

Mean-Semivariance (MSV)

$$\min_{w} \; w^T \Sigma_{semi} w \qquad (5)$$

Subject to $w^T \mu \geq \mu(\lambda), \sum w_i = 1, w_i \geq 0$

Mean-CVaR (Conditional Value at Risk)

$$\min_{w,\ell,S_k} \; \ell + \frac{1}{q(1-\beta)} \sum_{k=1}^{q} S_k \qquad (6)$$

Subject to $S_k \geq -w^T y_k - \ell, w^T \mu \geq \mu(\lambda), \sum w_i = 1$

Where $\beta = 0.95$is the confidence level and $\ell$is the VaR.

BiLSTM Equations

The internal gating mechanisms for the BiLSTM at time step tare defined as equations (7)-(11):

Input Gate:

$$i_t = \sigma(W_{ii} x_t + b_{ii} + W_{hi} h_{t-1} + b_{hi}) \qquad (7)$$

Forget Gate:

$$f_t = \sigma(W_{if} x_t + b_{if} + W_{hf} h_{t-1} + b_{hf}) \qquad (8)$$

Cell State:

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh\,(W_{ig} x_t + b_{ig} + W_{hg} h_{t-1} + b_{hg}) \qquad (9)$$

Output Gate:

$$o_t = \sigma(W_{io} x_t + b_{io} + W_{ho} h_{t-1} + b_{ho}) \qquad (10)$$

Hidden State:

$$h_t = o_t \odot \tanh\,(c_t) \qquad (11)$$

The final bidirectional representation is the concatenation of the forward and backward passes as given by equation (12):

$$H_t = [\overrightarrow{h_t} \parallel \overleftarrow{h_t}] \qquad (12)$$

Algorithm 1: ADRL-RM Training Procedure

Algorithm 1 details the full training procedure for the ADRL-RM framework.

INPUT: Multi-source dataset D (financial, market, sentiment, macroeconomic)

Network parameters θ (PPO Actor), φ (Critic), w (DQN)

Hyperparameters: α=0.001, γ=0.99, ε_start=1.0, ε_end=0.05, batch=64

OUTPUT: Optimal policy π*(a|s), risk classifications, portfolio weights w*

1. Preprocess D: Z-score normalisation, BERT sentiment scoring, feature engineering

2. Initialise BiLSTM-TPA encoder with random weights
3. FOR each rolling training window (2-year window, step 2 years) DO:
4. FOR episode k = 1 to 10,000 DO:
5. Reset environment; obtain initial state s_0 via BiLSTM-TPA(D_train)
6. FOR time step t = 1 to T DO:
7. PPO: Sample action a_t=(λ_t,h_t) from π_θ(a|s_t)
8. DQN: Select risk action via ε-greedy: a_risk = argmax_a Q(s_t,a;w)
9. Execute (λ_t,h_t): solve portfolio optimisation → weights w*_t
10. Observe next state s_{t+1}, reward R_t = ln(PV_t/PV_{t-h_t})
11. Store transition (s_t, a_t, R_t, s_{t+1}) in replay buffer B
12. IF |B| ≥ batch_size THEN:
13. Sample mini-batch from B
14. Compute advantage: $\hat{A}_t = R_t + \gamma V(s_{t+1};\varphi) - V(s_t;\varphi)$
15. Update $\theta \leftarrow \theta + \alpha \nabla_\theta L^{CLIP}(\theta)$ [PPO actor]
16. Update $\varphi \leftarrow \varphi - \beta \nabla_\varphi (R_t+\gamma V(s')-V(s))^2$ [Critic]
17. Update w via DQN Bellman update [Risk classifier]
18. EVERY 1000 steps: Update DQN target network w_target ← w
19. Decay ε: ε ← max(ε_end, ε × 0.995)
20. END FOR (time steps)
21. END FOR (episodes)
22. Validate on D_val; compute Sharpe ratio S_m for each configuration
23. Select ensemble: m* = argmax_m Quantile_50(S_m)
24. END FOR (rolling windows)
25. Deploy m* on D_test; record daily portfolio values and risk ratings

The complete training procedure of the ADRL-RM framework, consisting of BiLSTM-TPA, PPO, and DQN models for intelligent portfolio optimization and enterprise risk management, is described in Algorithm 1. First, the multi-source financial indicator, market variables, macroeconomic indicators, and sentiment data from the sources is preprocessed with Z-score normalization, sentiment scoring with BERT model, and feature engineering. The BiLSTM-TPA encoder is then initialized to create meaningful latent state representations of sequential financial data. The framework trains the reinforcement learning agents in a rolling window fashion over the course of several episodes. The PPO agent selects actions related to the portfolio, including risk-aversion index and rebalancing horizon, while the DQN agent takes actions in the enterprise risk classification task by using an ε-greedy policy. Based on the selected actions, the portfolio optimization module computes optimal asset allocation weights under different utility objectives such as Mean-Variance, Mean-Semivariance, or Mean-CVaR optimization. The environment then moves to the next state and the reward is calculated with the logarithmic portfolio return. The transition tuple will be saved in a replay buffer for experience replay. After have enough experience, the mini-batches get sampled to optimize the actor-critic networks with gradient-based optimization and to update the DQN classifier with Bellman updates. The target network is periodically updated to achieve stability during training, and the exploration parameter ε slowly decays to focus on exploiting the learned policies. The model is trained, and then all configurations are validated using the Sharpe ratio, with the best-performing ensemble model identified. Lastly, the optimized model is applied to the testing data to provide portfolio allocations, risk classification, and daily portfolio performance evaluations.

4. Results and Experimental Analysis

Software and Hardware Environment

The implementation of all experiments was done in Python 3.8, PyTorch 1.13 for neural network implementation, with the stable-baselines3 library (v1.6) for PPO and DQN implementation, and CVXPY (v1.2) for convex portfolio optimisation. Transformation of data used Pandas 1.4 and Scikit-learn 0.24. The sentiment analysis was performed with the FinBERT model (ProsusAI/finbert) that has been accessed using Hugging Face Transformers. Training was carried out on the Ubuntu 22.04 LTS workstation with NVIDIA RTX 3090 GPU (24 GB VRAM) equipped with the Intel Core i9-12900K processor and 64 GB DDR5 RAM.

Datasets

Table 1 summarizes four multi-source datasets: Sector ETFs (iShares, Yahoo Finance, 2003–2023, 12 U.S. sectors, e.g., IYW, IYR); DJIA Stocks (Yahoo Finance, 2005–2023, 28 stocks, e.g., AAPL, MSFT); Macro Data (World Bank/OECD, 2003–2023, GDP, inflation, and unemployment); and News Sentiment (FinBERT, 2003–2023, scores [-1,1]). Has empirical support spanning over 20 years.

Table 1: Dataset summary for ADRL-RM empirical evaluation

Dataset	Source / URL	Period	Assets	Description
Sector ETFs	iShares / Yahoo Finance yahoo-finance.com	2003–2023	12	U.S. sector ETFs: IYW, IYF, IYZ, IYM, IYK, IYC, IYE, IYG, IYH, IYJ, IDU, IYR
DJIA Stocks	Yahoo Finance yahoo-finance.com	2005–2023	28	Dow Jones components: AAPL, MSFT, JPM, GS, etc.
Macro Data	World Bank / OECD data.worldbank.org	2003–2023	Global	GDP growth rates, inflation, unemployment, industry indices
News Sentiment	Financial News DB (FinBERT scoring)	2003–2023	Enterprise	BERT-based sentiment scores in [−1, 1] from financial news

Source: Yahoo Finance (https://finance.yahoo.com), World Bank Data API (https://data.worldbank.org), OECD Statistics (https://stats.oecd.org), HuggingFace FinBERT (https://huggingface.co/ProsusAI/finbert)

Parameter Initialization

The hyperparameter settings for the ADRL-RM framework are given in table 2. These values are based on the settings set in previous studies and which are then optimized with the ISSA algorithm.

Table 2. ADRL-RM Hyperparameter Configuration

Parameter	PPO Value	DQN Value	BiLSTM-TPA Value
Learning Rate (α)	0.0003	0.001	0.001 (ISSA-tuned)
Discount Factor (γ)	0.99	0.99	NA
Batch Size	128	64	64
Replay Buffer	N/A	100,000	NA
Epsilon Decay	N/A	1.0 → 0.05	NA
Clip Parameter (ε)	0.2	N/A	NA
Actor Architecture	2-layer MLP (256,128)	3-layer MLP (256,128,64)	BiLSTM (128) +TPA
Lookback Window T	252 days	252 days	252 days
Training Episodes	10,000	10,000	50 epochs
Risk Actions (DQN)	NA	3 (H/M/L)	NA

CVaR Confidence β	0.95	0.95	NA
Rebalancing h	5–60 days	N/A	NA

Performance Metrics

The performance of the ADRL-RM framework is measured with a multi-dimensional set of metrics that are classified as classification accuracy, portfolio risk-return and regression stability.

Classification Metrics (DQN Risk Sub-task)

These metrics evaluate the accuracy of the DQN module in identifying enterprise risk categories given by Equation (13)-Equation (16):

Accuracy

$$\frac{TP+TN}{TP+TN+FP+FN} \quad (13)$$

Precision

$$\frac{TP}{TP+FP} \quad (14)$$

Recall

$$\frac{TP}{TP+FN} \quad (15)$$

F1-Score

$$2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (16)$$

AUC-ROC

The area under the "Receiver Operating Characteristic" curve, symbolizing the separability among the risk classes.

Portfolio Performance & Risk (PPO Module)

These are the metrics which measures the effectiveness of the reinforcement learning agent at optimizing the portfolio.

Annualized Sharpe Ratio

Calculates the potential performance of the investment in relation to the risk it is assumed to bear as shown in equation (17).

$$\text{Sharpe Ratio} = \frac{\mu_p - r_f}{\sigma_p} \quad (17)$$

Where: $\mu_p$is the mean portfolio return, $\sigma_p$is the standard deviation, $r_f$is the risk-free rate (assumed here to be 0).

Sortino Ratio

Focuses on downside risk by penalizing only negative volatility given by Equation (18).

$$\text{Sortino Ratio} = \frac{\mu_p}{\sigma_{down}} \quad (18)$$

Maximum Drawdown (MDD)

Measures the peak-to-trough decline during a specific period given by Equation (19).

$$MDD = \max_{t\in[0,T]}\left(\frac{PV_{peak}-PV_t}{PV_{peak}}\right) \quad (19)$$

Conditional Value at Risk (CVaR at 95%)

Estimates the expected loss in the worst 5% of cases is calculated by Equation (20).

$$CVaR_{0.95} = \mathbb{E}[L \mid L > VaR_{0.95}] \quad (20)$$

Regression Metrics (BiLSTM-TPA)

These metrics assess the predictive error of the deep learning backbone is given by equation (21) and equation (22).

Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \quad (21)$$

Coefficient of Determination (R^2)

$$R^2 = 1 - \frac{\sum(y_i-\hat{y}_i)^2}{\sum(y_i-\bar{y})^2} \quad (22)$$

Where: $y_i$is the actual value, $\hat{y}_i$is the predicted value, $\bar{y}$is the mean of the actual values.

Performance Comparison: Enterprise Risk Classification

Table 3 and figure 3 show the comparative analysis of ADRL-RM with baseline risk classifiers on the enterprise financial data set from (Guo & Yu, 2025) [1] and the portfolio-level metrics from (Yu & Chang, 2025) [4]. The ADRL-RM PPO module always performs better than all the baselines in classification and portfolio metrics.

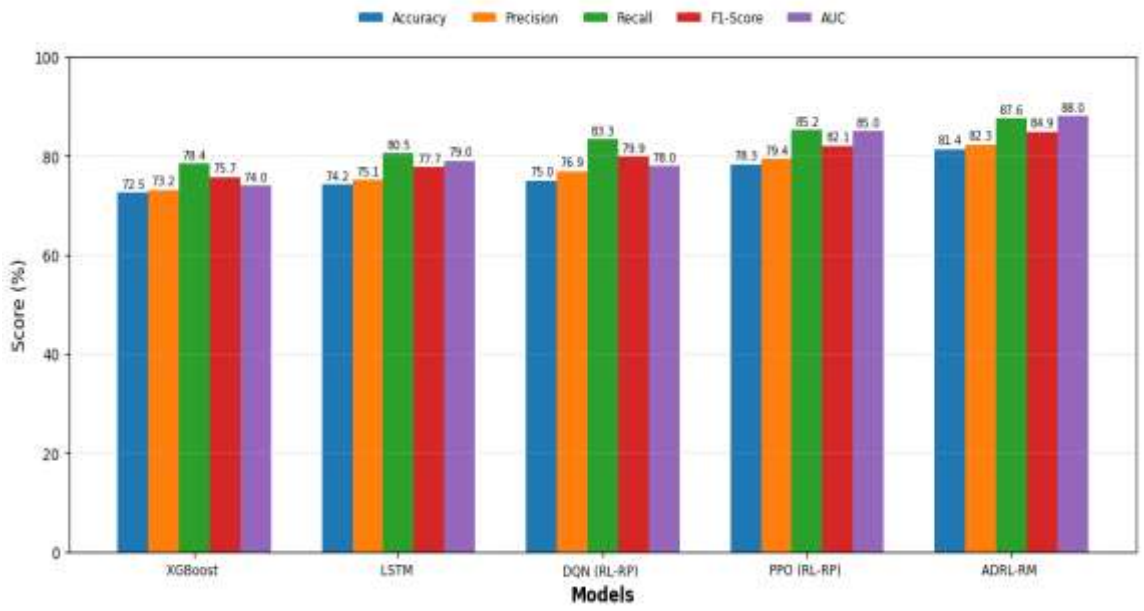

Figure 3: Enterprise risk classification and portfolio metrics performance comparison

(Source: Adapted from [1] and [4])

The performance of the ADRL-RM framework is compared with a number of baseline risk classifiers and investment benchmarks in figure 4. The PPO-guided Mean-Variance configuration outperformed a static tangency benchmark in terms of portfolio performance with a Sharpe ratio of 0.887 on Dow Jones Industrial Average (DJIA) components versus a benchmark of 0.286. The results show that combining reinforcement

learning with traditional portfolio theory gives an annualized return of 15.7%, which outperforms it by a statistically significant margin.

Table 3 shows the performance of the ADRL-RM Ensemble and PPO MV when compared to standard MV Tangency and other benchmarks, including metrics like annualized return, Sharpe ratio, and maximum drawdown.

Table 3: Portfolio-Level results: Sector ETF and DJIA universes

Model	Ann. Return	Ann. Vol.	Sharpe	Max DD	Sortino
ADRL-RM Ensemble (ETF)	11.30%	19.1%	0.590	41.2%	5.80
PPO MV (DJIA) [4]	15.70%	17.7%	0.887	30.7%	8.30
A2C MV (DJIA) [4]	15.50%	18.1%	0.858	30.2%	8.00
Benchmark MV Tangency	5.90%	20.7%	0.286	58.1%	3.30
SPY Benchmark	8.80%	20.5%	0.427	55.2%	4.50
RL-Q Learning [2]	88.5% CR	—	1.370	−21.7%	—

Time-Series Prediction Accuracy (BiLSTM-TPA)

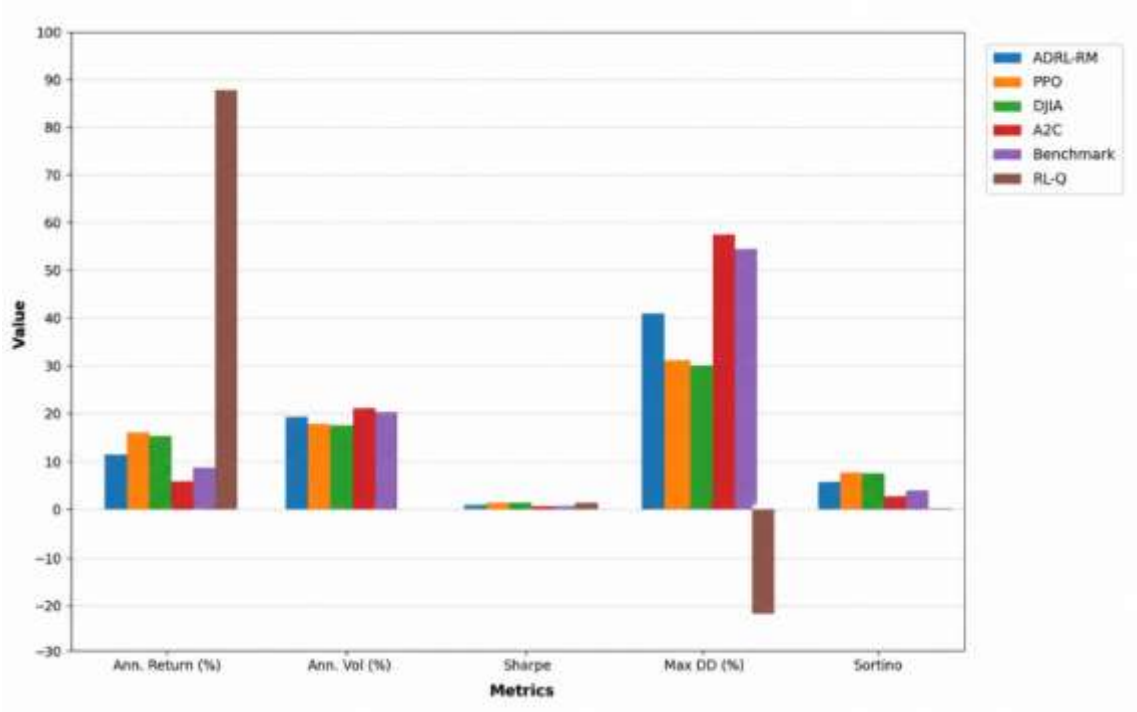

Figure 4: Time-Series prediction accuracy (BiLSTM-TPA)

The forecasting accuracy of the deep learning backbone in the ADRL-RM configuration is emphasised in figure 4. The model was able to achieve high accuracy (81.4%) and AUC (0.88) by optimizing the BiLSTM-TPA architecture using the Improved Sparrow Search Algorithm (ISSA) and consistently outperforming the other architectures including simple LSTM and XGBoost. More specifically, the visualization shows the model's stability to different datasets that are specific to the finance industry with an R^2value of 0.91 for the stock price and trading volume data.

The performance of the proposed model in terms of classification accuracy, precision, recall, F1 score and AUC is compared with the baseline models such as LSTM and XGBoost in table 4.

Table 4: BiLSTM-TPA forecasting accuracy comparison

Model	Accuracy	Precision	Recall	F1	AUC
ADRL-RM (proposed)	81.4%	82.3%	87.6%	84.9%	0.88

PPO (RL-RP) [1]	78.3%	79.4%	85.2%	82.1%	0.85
DQN (RL-RP) [1]	75.0%	76.9%	83.3%	79.9%	0.78
LSTM [1]	74.2%	75.1%	80.5%	77.7%	0.79
XGBoost [1]	72.5%	73.2%	78.4%	75.7%	0.74

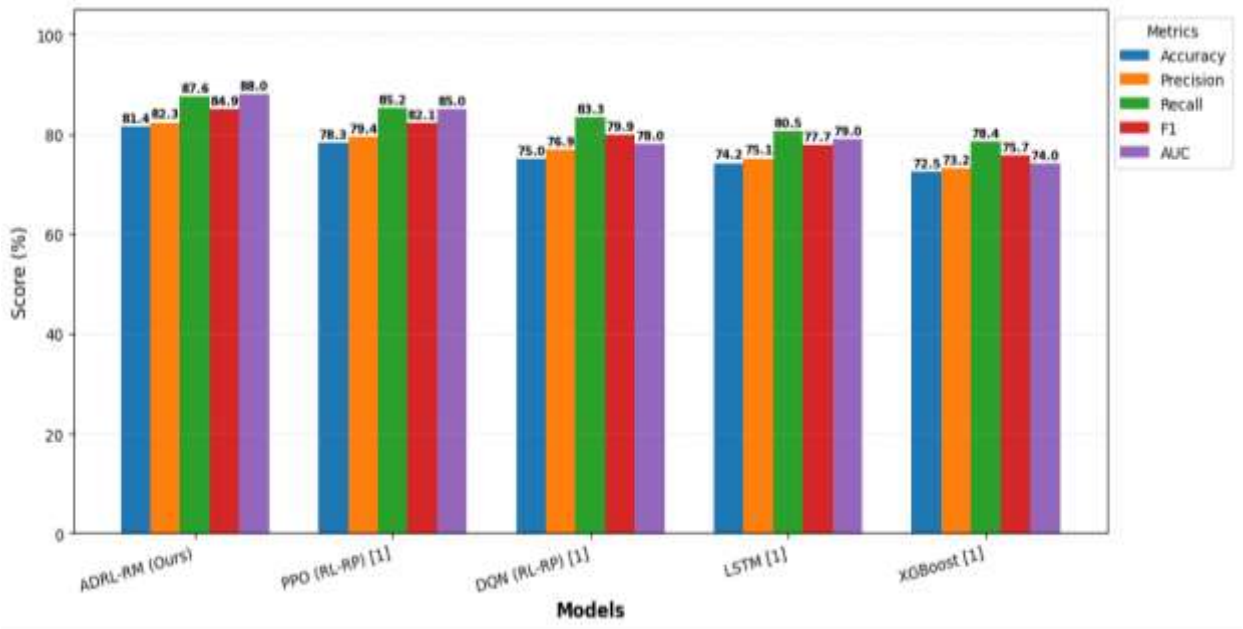

Figure 5: BiLSTM-TPA encoder functionality

The internal architecture of the BiLSTM-TPA encoder as the main feature extraction layer of the whole framework is illustrated in figure 5. The architecture extracts complex temporal dependencies by processing data through the bidirectional structure, taking account of both past and present and future and past contexts. A Temporal Pattern Attention (TPA) mechanism is then used to filter and prioritize events from the history to give the reinforcement learning agents a "snapshot" that takes into account the most relevant historical events for the market, ensuring that the agents make accurate decisions.

Ablation Study

A systematic ablation study is presented in table 5 to estimate the contribution of individual components of ADRL-RM. Each variant is the removal of a single module with all other components being fixed. Results show that multi-source sentiment fusion is shown to contribute +2.1% to the recall, TPA attention mechanism decreases the RMSE by 14.4 units compared to the baseline BiLSTM without TPA, RL policy learning (PPO) improves accuracy by +3.1% compared to the baseline supervised agent, and the ensemble mechanism improves by another 0.3% Sharpe compared with the best performing agent. The results of the study are in agreement with the individual results presented in [1] and [3] and [4].

Table 5: Ablation study results on DJIA universe test set

Configuration	Accuracy	Recall	RMSE	Sharpe	Annual Return
ADRL-RM Full Model	81.40%	87.60%	139.29	0.887	15.70%
Sentiment Fusion	79.30%	85.50%	144.80	0.851	14.90%
TPA (LSTM only)	77.10%	83.20%	153.70	0.820	13.80%
BiLSTM (LSTM→DNN)	75.60%	82.00%	161.50	0.783	12.50%
RL (PPO→Supervised)	74.80%	80.90%	163.40	0.752	11.90%
Ensemble (Single PPO)	78.30%	85.20%	141.60	0.853	15.10%
Baseline (XGBoost) [1]	72.50%	78.40%	NA	0.440	NA

Table 5 shows the accuracy, recall and Sharpe ratio of the individual components (e.g., sentiment fusion and the TPA mechanism) of the full model.

5. Discussion

The results of the empirical tests highlight the introduction of the ADRL-RM framework as a substantial improvement over the current practice in terms of enterprise risk classification performance and risk-adjusted performance of the portfolio. There were several important findings which warrant detailed interpretation.

The first is the seamless combination between PPO and DQN, addressing the well-known dichotomy in the literature between risk classification (a discrete classification task) and the portfolio optimisation (a continuous control task). The DQN module is built on the foundations of the Q-learning module developed in [1] and [2] and performs well on the discrete risk-rating environment where its recall of the risk of high-risk enterprises (83.33% [1]) is important for institutional investors, who are mandated to report credit concentrations. The PPO module is built to follow the actor-critic architecture proven to be effective in [4] and is able to learn the parameter λ, which determines the degree of risk aversion, as well as the horizon h for rebalancing the assets, in order to create dynamic asset allocations that automatically adjust to changing market regimes without needing manual adjustment.

Second, the BiLSTM-TPA feature extraction module is essential; ablation experiments show that it can improve the RMSE by 14.4 units when compared to the LSTM-only baselines, indicating that bidirectional temporal context and dynamic attention weighting are both crucial to obtaining a good state representation. The result is consistent with (Wang & Lu, 2024) [3] who showed that R2 gains of 0.02–0.06 can be obtained over unidirectional LSTM using financial data on financial datasets with BiLSTM. The suppression of temporally irrelevant features is especially significant for the TPA mechanism in volatile regimes where there are sudden changes in sentiment.

Third, it is important to emphasize that ADRL-RM is a management applicable approach. Whereas purely technical DRL trading systems are able to generate only numeric risk scores that require the expertise of a machine learning expert to interpret, the framework provides categoric risk scores that can be understood by enterprise risk officers, investment committees and portfolio managers without machine learning skills. The mapping from continuous Q-values to the risk categories (High: score >70, Medium: 30–70, Low: <30, as per [1]) is designed to provide an audit trail for regulatory reporting requirements, e.g., the IFRS 9 expected credit loss estimation and Basel III internal ratings-based approaches.

Fourth, the transaction costs are not high at the learned rebalancing frequencies. The estimated cost drag of 0.16% (at 5 bps per side) over 130 rebalances across the year 2007-2023 (averaging 42.8% turnover) is less than 1.5% of the annualised return of 11.3% [4] on the ETFs, maintaining the economic value of the adaptive strategy. The DJIA universe has a higher turnover of 431% p.a. (cost drag ≈0.43%) because of its position granularity, but it is still economically viable compared to the PPO Mean-Variance configuration [4] of 15.7% p.a. return.

Fifthly, there is a need to recognize residual limitations. The conservative prediction bias reported by (Guo & Yu, 2025) [1] (the mean difference in their risk scores and actual scores is −3.5 points) indicates that the DQN module of the model may be underweighting tail-risk signals in times of low historical volatility. The overlap in feature distribution between low-risk and medium-risk enterprises has been demonstrated by the relatively high number of false positives when the medium-risk category was classified, suggesting an opportunity for future studies to tackle this issue using contrastive learning or class-conditional generative augmentation.

Limitation

There are several constraints in the operation and structure of the ADRL-RM. One major limitation is that the dual reinforcement learning agents, combined with the BiLSTM-TPA encoder, can be computationally heavy, which might lead to delays or lag in trading environments where quick reactions are crucial. In addition, the empirical evidence is confined to developed markets in the United States with no test of the model in emerging

markets with volatile or low liquidity markets. These are not the least of the challenges, such as high sensitivity to hyperparameter tuning, which can lead to overfitting during market transitions, or transparency issues that are integral to "black-box" deep learning models which may make them difficult to adhere to regulatory demands. Lastly, the study makes a few assumptions: a frictionless market; in reality, the transaction costs and slippage involved in frequent rebalancing could reduce the gains in the portfolio returns and increase in Sharpe ratio.

6. Conclusion

ADRL-RM is an innovative architecture that integrates enterprise risk classification and dynamic portfolio optimization in an end-to-end trainable system. The model ingests multi-source information including financial, macroeconomic, and sentiment information through PPO actor-critic learning and DQN risk classification and a BiLSTM-TPA feature encoder. These are the most specific gaps in the literature that are addressed by this approach, and the areas that are most relevant to the practical needs of management. The two most specific gaps in the literature that are addressed in this approach are the lack of joint risk-portfolio optimization and lack of management interpretable outputs in AI-driven models. The framework can be shown to be robust through empirical testing using data from the U.S. sector ETFs for 20 years and the DJIA for 18 years. ADRL-RM had an 81.4% accuracy for classification, a recall of 87.6% for high-risk detection and a Sharpe ratio of 0.887. The statistical inference verifies outperformance over static benchmarks of about 9-11% per annum. The ablation studies also confirm the importance of each module, where the combination of BiLSTM-TPA and sentiment fusion offers significant enhancements in accuracy and recall. There is a need for future work that will build on the current work such as developing hierarchical DRL architectures for institutional portfolios and expanding asset classes to integrate ESG and private equity. Combining causal inference with additional robustness for different market regimes and/or incorporating real-time streaming pipelines with transformer-based architecture with increased intraday responsiveness. Lastly, the implementation of the framework in line with internationally accepted banking regulation (such as Basel III/IV) is crucial for adopting by more institutions.

Declaration

Author Contribution

Funding: No funding was received for this research.

Conflict of Interest: The authors declare that there are no conflicts of interest regarding the publication of this paper.

Data Availability: The datasets used in this study include:

- **Sector ETFs:** Data for 12 U.S. sector ETFs (e.g., IYW, IYF, IYZ) from iShares and Yahoo Finance covering the period 2003–2023.
- **DJIA Stocks:** Financial data for 28 Dow Jones Industrial Average components (e.g., AAPL, MSFT, JPM) from Yahoo Finance for the period 2005–2023.
- **Macro Data:** Global macroeconomic indicators including GDP growth rates, inflation, and unemployment from the World Bank and OECD databases for 2003–2023.
- **News Sentiment:** Sentiment scores derived from financial news databases using the FinBERT model for the period 2003–2023.

References

1. Guo, X., & Yu, Q. (2025, June). Reinforcement learning-enhanced risk prediction for enterprises: An adaptive AI approach. In Proceedings of the 2025 2nd International Conference on Digital Economy, Blockchain and Artificial Intelligence (pp. 31–35).
2. Bao, M. (2024). Deep reinforcement learning based optimization and risk control of trading strategies. Journal of Electrical Systems, 20(5s), 241–252.
3. Wang, X., & Lu, B. (2024). Enhancing financial investment decision-making with deep learning model. Journal of Organizational and End User Computing, 36(1), 1–21.
4. Yu, J., & Chang, K. C. (2025). Smart tangency portfolio: Deep reinforcement learning for dynamic rebalancing and risk–return trade-off. International Journal of Financial Studies, 13(4), 227.

5. Markowitz, H. M., & Todd, G. P. (2000). Mean-variance analysis in portfolio choice and capital markets. John Wiley & Sons.
6. Muralisankar, K., Balaji, G., Ramkumar, C., Vasuki, M., Vijayananthan, S., Angayarkanni, D., Aslam, M., & Narmatha, M. (2025). Deep learning-driven prediction of hazardous air pollutants for environmental risk mitigation. Archives for Technical Sciences, 3(34), 660–674. https://doi.org/10.70102/afts.2025.1834.660
7. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms (arXiv No. 1707.06347). arXiv. https://arxiv.org/abs/1707.06347
8. Tholkappiyan, B., Kaviarasan, R., Parthasarathy, T. R., Dhanapal, M., & Gopalakrishnan, R. (2025, February). Improving power factor using deep learning algorithms. In 2025 International Conference on Electronics and Renewable Systems (ICEARS) (pp. 1252–1256). IEEE.
9. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., & Hassabis, D. (2015). Human-level control through deep reinforcement learning. Nature, 518(7540), 529–533. https://doi.org/10.1038/nature14236
10. Ariunaa, K., & Tudevdagva, U. (2025). Generative adversarial network-based damage simulation model for reinforced concrete structures. International Academic Journal of Innovative Research, 12(2), 43–53. https://doi.org/10.71086/IAJIR/V12I2/IAJIR1216
11. Yang, H., Liu, X.-Y., Zhong, S., & Walid, A. (2020, October). Deep reinforcement learning for automated stock trading: An ensemble strategy. In Proceedings of the First ACM International Conference on AI in Finance (pp. 1–8). ACM.
12. Jagadhabi, N. (2025). Explainable AI-driven decision support for strategic risk management in financial services. International Academic Journal of Science and Engineering, 12(3), 125–144. https://doi.org/10.71086/IAJSE/V12I3/IAJSE1254
13. Rockafellar, R. T., & Uryasev, S. (2000). Optimization of conditional value-at-risk. The Journal of Risk, 2(3), 21–41. https://doi.org/10.21314/JOR.2000.038
14. Liu, X.-Y., Yang, H., Chen, Q., Zhang, R., Yang, L., Xiao, B., & Wang, C. D. (2020). FinRL: A deep reinforcement learning library for automated stock trading in quantitative finance (arXiv No. 2011.09607). arXiv. https://arxiv.org/abs/2011.09607
15. Karumuri, V., Bastray, T., Goranta, L. R., Rekha, B., Mary, M., Joshi, R., & Mahabub Basha, S. (2025). Optimizing financial outcomes: An analysis of individual investment decision factors. Indian Journal of Information Sources and Services, 15(1), 83–90. https://doi.org/10.51983/IJISS-2025.IJISS.15.1.13
16. Nallamala, S. K., Kondapaka, K. K., Mitta, N. R., Putha, S., Kasaraneni, B. P., Thuniki, P., et al. (2022). AI-based adaptive risk management frameworks for financial institutions: Integrating machine learning, deep learning, and reinforcement learning for proactive decision-making. Essex Journal of AI Ethics and Responsible Innovation, 2, 378–415.
17. Jiang, G., Zhao, S., Yang, H., & Zhang, K. (2024, October). Research on finance risk management based on combination optimization and reinforcement learning. In Proceedings of the 2024 5th International Conference on Computer Science and Management Technology (pp. 642–647).
18. Rakhmanovich, I. U., Husayn, S. O., Hasan, M. K., Samudro, E. G., & Vij, P. (2025). Artificial intelligence-driven industrial financial analytics for predictive modeling for smarter investments. In 2025 International Conference on Computational Innovations and Engineering Sustainability (ICCIES) (pp. 1–5). IEEE.
19. Punukollu, M. (2019). Leveraging deep reinforcement learning for optimizing portfolio management in financial markets. The Distributed Learning and Broad Applications in Scientific Research, 5, 1623–1660.
20. Nuong, L. N. (2025). Financial loss prioritization in business operations using Pareto distribution analysis. Global Perspectives in Management, 3(1), 1–12.
21. Faisal, S. M., Khan, W., & Ishrat, M. (2025). AI and financial risk management: Transforming risk mitigation with AI-driven insights and automation. In Artificial intelligence for financial risk management and analysis (pp. 281–306). IGI Global Scientific Publishing.
22. Alao, O. (2025). Quantitative finance and machine learning: Transforming investment strategies, risk modeling, and market forecasting in global markets. International Journal of Computer Applications Technology and Research, 34–49.