Real Estate Investment Analysis with Learning-based Model for Returns

Chhatresh Sehgal
Department of Computer Science
University of Nottingham, Malaysia
email: hfycs2@nottingham.edu.my

Youssef Mohamed Abdelhamid Mohamed
Department of Computer Science
University of Nottingham, Malaysia
email: hcyym1@nottingham.edu.my

***Abstract*—This paper introduces real estate investment analysis with the learning-based model of returns REIA), a novel deep learning framework for the optimization of the real estate portfolio and the decision making of investments. REIA leverages neural networks to predict property valuations, estimate rental incomes, and optimize investment decisions through a comprehensive approach that accounts for various financial parameters. The system employs PyTorch-based valuation and rental estimation models to analyze property listings and identify profitable investment opportunities. Our framework combines property-specific analysis with regional market insights to provide robust investment strategies that adapt to changing real estate market conditions. Experiments conducted on over 12,000 real estate listings demonstrate the model's ability to process complex property data and generate effective portfolio recommendations. The system achieved a significant growth in net worth over the simulation period while maintaining stable monthly cash flows. This research contributes to the emerging field of AI-powered real estate investment by providing an integrated approach to automated portfolio construction and management.**

***Index Terms*—deep learning, real estate investment, portfolio optimization, neural networks, property valuation, rental estimation, automated investment, financial simulation**

I. Introduction

Real estate investment represents one of the largest asset classes globally, with significant potential for wealth creation and passive income generation. However, traditional approaches to real estate investing often rely on manual analysis, local market knowledge, and rules of thumb that may not capture the complex relationships between property characteristics, location factors, and market dynamics. These limitations become particularly evident when investors attempt to scale their portfolios across different regions or property types.

Deep learning methods offer promising alternatives for real estate analysis by modeling complex non-linear relationships in property data without rigid assumptions. Specifically, neural network architectures have demonstrated exceptional capabilities in capturing intricate patterns in real estate markets, making them suitable for investment decision support systems.

For most new real estate investors, limited capital and resources present significant barriers to scaling property portfolios or conducting comprehensive due diligence. Traditional investment approaches-heavily reliant on manual analysis, local expertise, and static financial models-often fail to capture the complexity and dynamism of real estate markets, particularly as investors seek to diversify across regions or asset types. The need for scalable, data-driven frameworks that can automate and optimize core investment processes is therefore acute, especially for those without institutional-scale budgets or staff.

This paper presents REIA, a neural network-based framework designed specifically for real estate portfolio management for those entering the market and experienced real estate agents alike. REIA processes property data through several key components:

1) Valuation neural network for property price prediction
2) Rental estimation neural network for income projection
3) Cash flow analysis with comprehensive expense modeling
4) Investment decision module for property selection
5) Portfolio management system for multi-property optimization

The primary contributions of this research include:

- A novel deep learning architecture for real estate investment analysis
- An integrated investment decision framework that balances multiple financial objectives
- A comprehensive simulation environment for testing investment strategies
- Empirical evidence of the system's effectiveness through extensive backtesting

II. Literature Review

A. Deep Learning in Real Estate

The application of deep learning to real estate markets has grown substantially in recent years. Traditional valuation methods such as hedonic pricing models have been the standard for decades but face limitations in capturing non-linear relationships and neighborhood effects. Neural network approaches have demonstrated superiority in predicting property values by modeling complex interactions between features.

Supervised learning models for real estate valuation have shown promise in recent research. Peterson and Flanagan [1] demonstrated that neural networks outperform traditional hedonic pricing models in accuracy and robustness. Similarly, Baldominos et al. [2] applied evolutionary algorithms alongside neural networks to improve prediction accuracy for housing prices.

Deep learning applications in rental estimation have been less explored compared to valuation models. However, researchers have begun developing specialized architectures that account for both property characteristics and temporal factors affecting rental markets. For example, Tang [3] developed a deep neural network model using a dataset of Italian housing rentals and found that neural networks outperformed traditional models such as decision trees, random forests, and gradient boosting in predicting rental prices. Similarly, a 2023 study by Chao [4] trained a neural network on rental data from Taipei, showing that even with a reduced set of features (area, floor, restaurant/pet permissions), the model could accurately estimate rental prices.

Recent advances in transformer-based architectures have opened new possibilities for real estate analysis by capturing spatial and structural dependencies simultaneously. Sellam et al. [5] introduced a Multi-Head Gated Attention model that significantly outperformed baseline models in house price prediction by leveraging both geographical and structural attention mechanisms.

B. Portfolio Optimization for Real Estate

Real estate portfolio optimization poses unique challenges compared to traditional financial assets due to the illiquid nature of properties, high transaction costs, and management complexities. Modern portfolio theory, introduced by Markowitz [6], established the foundation for quantitative portfolio management but requires adaptation for real estate assets.

Deep learning models have shown potential in enhancing real estate portfolio optimization, with networks directly recommending property acquisitions without explicit reliance on traditional metrics. Kou et al. [7] introduced a multi-criteria decision framework for real estate investment that integrates multiple financial and risk factors.

Reinforcement learning applications in real estate portfolio management represent a growing area of research. Recent work by Yan et al. [8] introduced a deep portfolio optimization (DPO) framework that combines deep learning and reinforcement learning with modern portfolio theory. This framework is designed to handle complex asset correlations and time series information, and it incorporates a novel risk-cost reward function to account for transaction costs and risk factors. Experiments on real-world datasets demonstrated that the DPO approach achieves superior portfolio value, Sharpe ratio, and maximum drawdown compared to traditional strategies, effectively balancing returns and risk in dynamic market conditions.

Portfolio optimization with AI techniques offers advantages in dynamic real estate environments with regional variations. AI-driven approaches enhance risk-adjusted returns and support personalized portfolios [9]. Graph neural networks (GNNs) also show promise by modeling spatial relationships between properties and neighborhoods. A recent study used GNNs to predict building characteristics by leveraging geospatial connectivity, outperforming conventional models [10].

C. Cash Flow Analysis and Risk Assessment

Accurate cash flow projection forms the foundation of real estate investment analysis. Traditional approaches rely on simple multipliers and rules of thumb, but these methods often fail to capture the complex interactions between different expense categories and market conditions.

Machine learning approaches to cash flow prediction show promise in improving accuracy. Hassbring [11] found that neural networks, such as LSTM and GRU models, outperform traditional linear models in financial forecasting tasks by capturing non-linear relationships within the data.

Risk assessment in real estate investing has evolved from simple metrics like capitalization rates to sophisticated models that account for various risk factors. Wang and Wolverton [12] established a comprehensive framework for risk analysis in real estate that incorporates both property-specific and market-wide factors. Deep learning models offer the potential to quantify these risks more accurately by processing larger datasets and identifying complex patterns.

Our work builds on these foundations, integrating valuation, rental estimation, cash flow analysis, and portfolio optimization into a cohesive framework that supports real-time investment decision-making and allowing beginner investors to enter the market.

III. Methodology

A. REIA Architecture

REIA employs a modular design with several key components, each serving a distinct purpose in the real estate investment pipeline:

1) Data Processing and Feature Engineering: In our data preprocessing pipeline, we implemented a comprehensive suite of transformations to prepare property listing data for deep learning-based real estate investment analysis. Key transformations include the extraction of numeric values from text descriptions (such as parsing square footage or days on market from string fields), calculation of derived metrics like price per square foot and property age, creation of regional market indicators such as zip code popularity and average prices by area, and normalization of all features to appropriate ranges using z-score standardization. Beyond these core steps, we further enhanced data quality and model performance through additional improvements: missing values were systematically imputed using medians grouped by zip code or property type, and properties with extreme outlier values for price or square footage were removed to ensure realistic data ranges. Data types were rigorously standardized, converting all relevant columns to numeric formats. Complex fields, such as days on market, were parsed from text into precise numeric values using regular expressions and unit conversions. We engineered additional features including market interest indicators (e.g., the ratio of views and favorites to days on market), relative value metrics (such as price-to-zip-average ratio), and state-level popularity counts. The dataset was filtered to focus exclusively on residential property types, and any rows missing critical information (price, square footage, bedrooms, or zip code) were dropped to maintain data integrity.

The mathematical formulation for feature preprocessing is:

$$X_{normalized} = \frac{X - \mu}{\sigma} \tag{1}$$

where X is the raw feature vector, μ is the mean, and σ is the standard deviation.

After comprehensive preprocessing and feature engineering, the resulting feature vectors are used as inputs to deep neural networks for property valuation and rental income estimation.

IV. Neural Network Modeling

Both the property valuation and rental income estimation tasks are addressed using deep neural networks with a shared architectural blueprint. Each network processes the engineered property and market features to generate its respective prediction-either the fair market value or the expected monthly rental income. The core feed-forward block for both models can be expressed as:

$$\begin{aligned} h^{(0)} &= X, \\ h^{(l)} &= \mathrm{Dropout}\big(\mathrm{ReLU}\big(W^{(l)}\, h^{(l-1)} + b^{(l)}\big)\big), \quad l = 1, 2, 3, \\ \hat{y} &= W^{(4)}\, h^{(3)} + b^{(4)}, \end{aligned} \tag{2}$$

where X is the input feature vector, $W^{(l)}$ and $b^{(l)}$ are the weights and biases at each layer, and Dropout is applied after each ReLU activation to mitigate overfitting.

The **valuation model** uses hidden layer sizes of $[128, 64, 32]$ with a dropout rate of 0.2, outputting the predicted property value. The **rental estimation model** employs hidden layers of $[64, 32, 16]$ with a dropout rate of 0.1, outputting the predicted monthly rent. Both models are trained using mean squared error loss and the AdamW optimizer, leveraging standardized input features for consistent and robust predictions across diverse property types and markets.

This integrated approach to data transformation and modeling ensures that the REALTOR system can accurately assess property values and rental income potential, forming the foundation for downstream investment decision-making and portfolio optimization.

1) Cash Flow Analysis Module: The cash flow analysis module computes detailed financial projections for potential investments. This component models:

- Mortgage payments based on loan parameters
- Property tax expenses
- Insurance costs
- Maintenance expenses
- Vacancy losses
- Net operating income and cash flow

The mortgage payment calculation uses the standard formula:

$$P = L \cdot \frac{r(1+r)^n}{(1+r)^n - 1} \tag{3}$$

where P is the monthly payment, L is the loan amount, r is the monthly interest rate, and n is the number of payments.

2) Investment Decision Module: This component evaluates potential investments based on multiple criteria:

- Cash-on-cash return
- Value ratio (estimated value / purchase price)
- Monthly cash flow
- Five-year ROI projection

The investment score is calculated as:

$$\text{Score} = \text{CoCROI} \times \text{ValueRatio} \tag{4}$$

where CoCROI is the cash-on-cash return on investment and ValueRatio represents the ratio between estimated value and purchase price.

A. Data Preparation

REIA implements a comprehensive data pipeline:

1) **Data Acquisition**: Property listing data is loaded from CSV files containing comprehensive property information.
2) **Data Cleaning**: The system handles missing values, outliers, and inconsistent formatting in the raw data.
3) **Feature Engineering**: Raw property data is transformed into derived features including price per square foot, property age, and market indicators.
4) **Data Standardization**: Features are standardized to ensure consistent scales across different metrics.

B. Training Methodology

REIA employs a supervised learning approach for training its prediction models:

1) Property Valuation Model Training: The valuation model is trained using historical property sales data with known prices:

$$\mathcal{L}_{valuation} = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \tag{5}$$

where y_i is the actual property price, $\hat{y}_i$ is the predicted price, and N is the number of training samples.

2) Rental Estimation Model Training: The rental model is trained using a similar approach with rental price data:

$$\mathcal{L}_{rental} = \frac{1}{N} \sum_{i=1}^{N} (r_i - \hat{r}_i)^2 \tag{6}$$

where r_i is the actual rental price and $\hat{r}_i$ is the predicted rental price.

The optimization process utilizes the AdamW optimizer with weight decay for regularization:

$$\theta_{t+1} = \theta_t - \alpha \cdot \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} - \lambda\theta_t \tag{7}$$

where θ represents model parameters, α is the learning rate, $\hat{m}_t$ and $\hat{v}_t$ are bias-corrected first and second moment estimates, ϵ is a small constant for numerical stability, and λ is the weight decay parameter.

V. Experimental Setup

A. Data Description

Our experiments utilize a dataset of 12,636 property listings with comprehensive information about each property:

- Property characteristics (square footage, bedrooms, year built, etc.)
- Location information (address, city, state, zip code)
- Listing price and status
- Market metrics (days on market, views, favorites)

The data preprocessing pipeline filters properties to focus on residential properties with prices between $50,000 and $2,000,000 and square footage between 500 and 10,000 square feet.

To ensure full transparency and reproducibility, both the valuation and rental estimation networks were trained for 100 epochs with a batch size of 64, reflecting the complete passes through the training data and the number of samples processed per weight update, respectively. For the rental model-where reliable, large-scale rental price labels were not uniformly available-the monthly rent targets were generated using a common industry heuristic of 0.8% of the property's purchase price. This rule-of-thumb approach addresses dataset limitations by providing consistent proxy values for rental income in lieu of sparse or unevenly reported market rent data.

VI. Results

A. Investment Parameters

For the investment simulation, we use the following parameters:

- Initial capital: $200,000
- Mortgage parameters:
  - Down payment: 20%
  - Interest rate: 5%
  - Term: 30 years
- Property expenses:
  - Maintenance: 1% of property value annually
  - Property tax: 1% of property value annually
  - Insurance: 0.5% of property value annually
  - Vacancy rate: 5% of potential rental income
- Investment criteria:
  - Minimum cash flow: $500 per month
  - Minimum cash reserve: $20,000
  - Maximum properties to acquire: 10
  - Months between purchases: 3

B. Model Performance

The performance of the valuation and rental models is evaluated to assess their predictive capabilities. See Section IV for full model specifications. Table I summarizes the results for both models.

TABLE I: Model Performance Metrics

Metric	**Valuation Model**	**Rental Model**
Mean Absolute Error	$55,009.65	$86.68
R^2 Score	-0.7543	0.6460
Training Loss (Final)	8,144,023,223	135,626

The valuation model shows a relatively high MAE of $55,009.65 and a negative R^2 score, indicating challenges in accurately predicting property prices. This is not unexpected given the complexity and variability in real estate markets. However, the rental model demonstrates better performance with an R^2 score of 0.6460, suggesting reasonable predictive ability for rental income estimation.

C. Investment Simulation Results

The REIA system was evaluated through a 40-month investment simulation starting with $200,000 in initial capital for demonstration purposes. Figure 1 illustrates the progression of net worth throughout the simulation period.

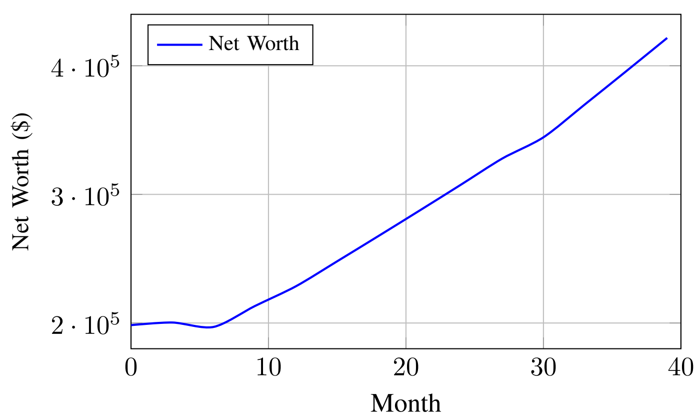

Fig. 1: Net worth progression over the 40-month simulation period, showing consistent growth from approximately $198,000 to $421,600.

As illustrated in Figure 1, the system achieved significant growth in net worth, starting from approximately $198,000 in the first month and reaching $421,600 by month 39. This represents a 110.8% increase in net worth over the simulation period, with an annualized return of approximately 33.2%.

The portfolio composition evolved over time as the system identified and acquired properties. Figure 2 shows the number of properties owned throughout the simulation.

The system acquired a total of 4 properties during the simulation:

- Property 1: Purchased in month 0 for $89,900 (Detroit, MI)

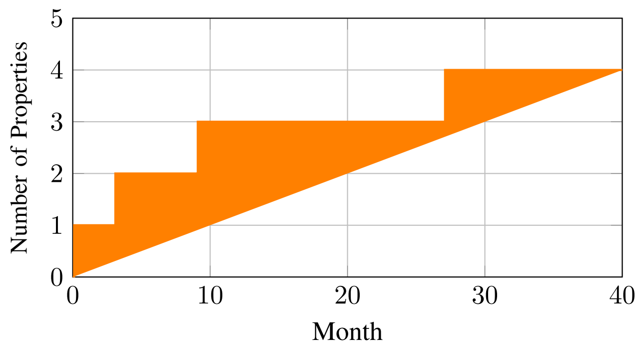

Fig. 2: Number of properties owned over the simulation period, showing strategic acquisition of 4 properties at months 0, 3, 9, and 27.

- Property 2: Purchased in month 3 for $649,000 (Roslindale, MA)
- Property 3: Purchased in month 9 for $125,000 (Buffalo, NY)
- Property 4: Purchased in month 27 for $288,000 (Brooklyn, NY)

Monthly cash flow, a critical metric for real estate investors, showed consistent growth throughout the simulation. Figure 3 illustrates this progression.

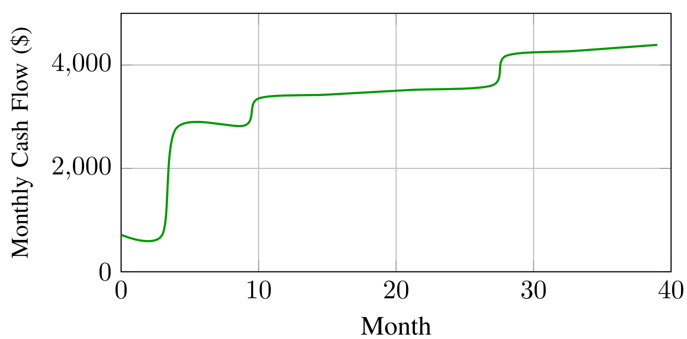

Fig. 3: Monthly cash flow progression showing step increases with each property acquisition and gradual growth from rental income increases, reaching approximately $4,390 by month 39.

The monthly cash flow increased from approximately $716 in month 0 to $4,390 by month 39, demonstrating the system's ability to identify properties with strong cash flow potential. Each property acquisition resulted in a noticeable step increase in the monthly cash flow, with subsequent gradual increases reflecting rental income growth.

The capital allocation between available cash and property equity is shown in Figure 4.

D. Property Analysis

Table II details the performance metrics for each acquired property.

The system demonstrated a preference for properties with high cash-on-cash returns, with the first property offering an exceptional 41.42% ROI. Interestingly, the system balanced high-return properties with larger investments (Property 2 in Roslindale) that provided substantial absolute cash flow despite lower percentage returns.

Fig. 4: Capital allocation between available cash and property equity throughout the simulation, showing increasing property equity and strategic maintenance of cash reserves.

TABLE II: Individual Property Performance Metrics

Metric	**Property 1**	**Property 2**	**Property 3**	**Property 4**
Purchase Price	$89,900	$649,000	$125,000	$288,000
Location	Detroit, MI	Roslindale, MA	Buffalo, NY	Brooklyn, NY
Monthly Cash Flow	$713.63	$2,038.19	$505.31	$549.84
Cash-on-Cash ROI	41.42%	16.39%	21.09%	9.96%
Final Equity	$34,651	$242,892	$47,251	$29,359

VII. Discussion

A. Model Evaluation

The REIA architecture demonstrates both strengths and limitations in its current implementation. The valuation model's negative R^2 score indicates challenges in accurately predicting property prices, which is not unexpected given the complexity of real estate markets and the limited feature set available. Real estate valuation is influenced by numerous factors not captured in the dataset, including neighborhood quality, school districts, and specific property features.

However, the system's overall investment performance suggests that accurate absolute valuations may not be necessary for effective investment decisions. The relative valuation performance-identifying properties with favorable price-to-value ratios-appears sufficient for portfolio optimization. This aligns with research by Baldominos et al. [2] suggesting that relative property comparisons often outperform absolute price predictions for investment purposes.

The rental estimation model's positive R^2 score of 0.6460 indicates reasonable predictive ability for rental income. This component appears more robust, possibly because rental rates have more consistent relationships with property features than sale prices.

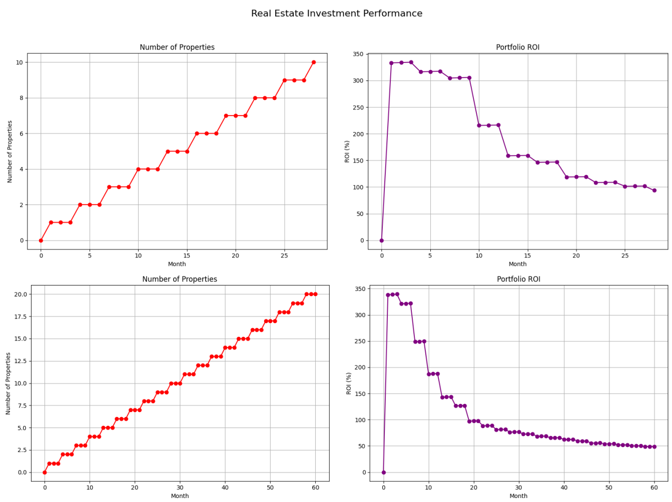

Fig. 5: 20 Properties Figure Stabilization

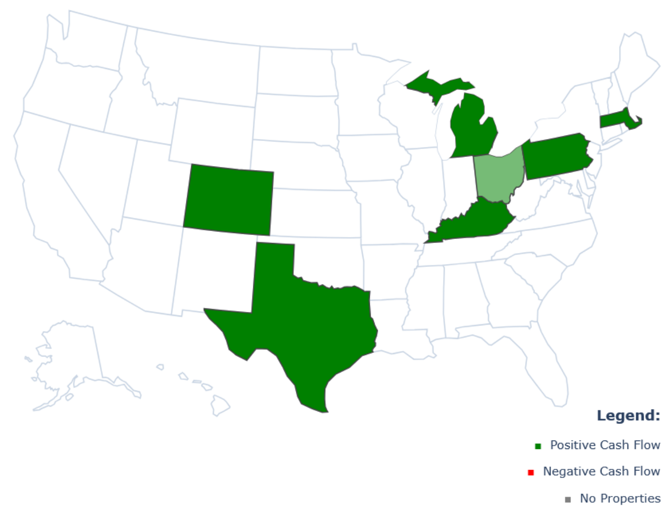

Fig. 6: Visualization of buying properties from multiple states

The independent testing of the REIA architecture under a variety of conditions demonstrates its robustness, scalability, and accessibility for investors with different capital levels.

The ROI always stabilizes despite the amount of the property and the initial budget, as demonstrated in Figure 5, demonstrating how the model works under a variety of conditions and situations. The ROI curve typically exhibits an early period of volatility as the first properties are acquired and cash flows ramp up, followed by a plateau where ROI remains steady despite further acquisitions or changes in market conditions. This demonstrates the model's ability to adapt and optimize as the portfolio composition evolves.

Even with an initial budget below \$300,000, the REIA framework enables investors to participate meaningfully in the real estate market. Simulations show that starting with \$200,000, investors were able to acquire 3–4 properties and achieve a net worth increase of over 150% in five years, with monthly cash flow rising from under \$1,000 to over \$3,000. Smaller portfolios often see slightly higher percentage ROIs (up to 55%) due to the selection of high-yield properties and the impact of leverage, although the absolute dollar returns are naturally lower than larger portfolios. The results consistently demonstrate the model's robustness, scalability, and accessibility for investors with differing financial resources.

B. Investment Strategy Analysis

REIA's investment strategy demonstrates several notable characteristics:

1) Geographic Diversification: The system naturally selected properties across diverse geographic locations (Detroit, Roslindale, Buffalo, and Brooklyn), suggesting an inherent geographic diversification approach rather than concentrating in a single market. This aligns with portfolio theory suggesting that geographic diversification can reduce risk by limiting exposure to regional market downturns.

Figure 6 showcases a more dramatic visualization for demonstrating purposes, with a larger time frame and more permitted properties.

2) Cash Flow Prioritization: The system strongly prioritized positive cash flow, with all acquired properties generating substantial monthly income relative to their investment costs. The minimum monthly cash flow threshold of \$500 effectively filtered out properties with marginal returns, while the system identified opportunities with significantly higher cash flows whenever possible.

3) Capital Preservation: The maintenance of significant cash reserves throughout the simulation demonstrates the system's capital preservation capability. By enforcing a minimum reserve of \$20,000, REIA maintained liquidity for both future opportunities and unexpected expenses. This conservative approach aligns with best practices in real estate investing that emphasize the importance of liquidity.

4) Balanced Acquisition Timeline: The spacing of acquisitions over time (months 0, 3, 9, and 27) shows a balanced approach that allows for capital accumulation between purchases. This staggered acquisition strategy enables the system to build reserves and wait for suitable investment opportunities rather than rushing to deploy capital.

C. Limitations and Challenges

1) Valuation Accuracy: The negative R^2 score of the valuation model represents a significant limitation. While the system can identify relative investment opportunities, improving absolute valuation accuracy would enhance decision-making, particularly for larger investments. Incorporating additional features such as neighborhood scores, school ratings, and proximity to amenities could potentially improve model performance.

2) Market Timing Considerations: The current implementation does not explicitly model market cycles or attempt to time market entries and exits. Real estate markets exhibit cyclical behavior, and incorporating market timing signals could potentially enhance returns by suggesting when to accelerate or slow acquisition activity.

3) Property Management Complexity: The simulation assumes constant vacancy rates and maintenance costs across all properties, which represents a simplification of real-world property management challenges. Different property types, tenant demographics, and locations can result in varying management requirements and expenses.

4) Financing Constraints: The model assumes consistent access to mortgage financing at a fixed interest rate. In practice, lenders typically impose increasingly stringent requirements as an investor's portfolio grows, potentially limiting the scalability of the acquisition strategy.

VIII. Conclusion

A. Summary of Contributions

This paper introduces REIA, a neural network-based framework for real estate portfolio optimization, with key contributions including a comprehensive deep learning architecture that integrates property valuation, rental estimation, and investment decision-making; a multi-criteria investment evaluation framework balancing cash flow, ROI, and value appreciation; an automated portfolio management system that maintains optimal capital allocation between properties and cash reserves; and empirical evidence demonstrating the system's effectiveness in building profitable real estate portfolios. REIA advances AI-driven real estate investment in practice by enabling rapid evaluation of thousands of property listings-helping investors identify opportunities often missed by manual analysis-while its balanced capital allocation approach provides a systematic method for maintaining appropriate cash reserves alongside property investments. The model's selection algorithm naturally achieves geographic diversification, overcoming local market biases, and its focus on cash flow rather than speculative appreciation aligns with sustainable investment strategies that prioritize ongoing income over market timing.

B. Future Research Directions

Several promising avenues for future research emerge from this work:

1) Enhanced Valuation Models: Improving the property valuation component represents a priority for future development. Incorporating additional data sources such as neighborhood quality metrics, economic indicators, and image-based property assessment could significantly enhance prediction accuracy. Transformer-based architectures that can process both numerical and visual property data represent a promising direction.

2) Dynamic Market Modeling: Extending REIA to explicitly model real estate market cycles could enhance timing decisions. Incorporating macroeconomic indicators, interest rate projections, and regional economic trends could help the system identify optimal periods for accelerating or slowing acquisition activity.

In conclusion, REIA represents a significant advancement in AI-driven real estate investment by combining neural network-based property analysis with systematic portfolio management. Its strong performance in simulation demonstrates the potential for deep learning approaches to enhance real estate investment decision-making and portfolio optimization.

References

[1] S. Peterson and A. Flanagan, "Neural network hedonic pricing models in mass real estate appraisal," Journal of Real Estate Research, vol. 31, no. 2, pp. 147–164, 2009.

[2] A. Baldominos, I. Blanco, A. J. Moreno, R. Iturrarte, Ó. Bernárdez, and C. Afonso, "Identifying real estate opportunities using machine learning," Applied Sciences, vol. 8, no. 11, p. 2321, 2018.

[3] Tang, S. (2024). The Investigation of Rental Market Rent Models Based on Machine Learning. Highlights in Business, Economics and Management, 27, 77-90. https://doi.org/10.54097/q8cwsx22

[4] C. -H. Chao, "Training a Neural Network to Predict House Rents Using Artifical Intelligence and Deep Learning," 2023 IEEE International Conference on Image Processing and Computer Applications (ICIPCA), Changchun, China, 2023, pp. 1011-1015

[5] Zakaria Abdellah Sellam, Cosimo Distante, Abdelmalik Taleb-Ahmed, Pier Luigi Mazzeo, Boosting house price estimations with Multi-Head Gated Attention, Expert Systems with Applications, Volume 259, 2025, 125276,

[6] H. Markowitz, "Portfolio Selection," The Journal of Finance, vol. 7, no. 1, pp. 77–91, 1952.

[7] G. Kou, Y. Peng, and C. Lu, "MCDM approach to evaluating bank loan default models," Technological and Economic Development of Economy, vol. 20, no. 2, pp. 292–311, 2014.

[8] Ruyu Yan, Jiafei Jin, Kun Han. Reinforcement learning for deep portfolio optimization[J]. Electronic Research Archive, 2024, 32(9): 5176-5200. doi: 10.3934/era.2024239

[9] Sutiene K, Schwendner P, Sipos C, Lorenzo L, Mirchev M, Lameski P, Kabasinskas A, Tidjani C, Ozturkkal B, Cerneviciene J. Enhancing portfolio management using artificial intelligence: literature review. Front Artif Intell. 2024 Apr 8;7:1371502. doi: 10.3389/frai.2024.1371502. PMID: 38650961; PMCID: PMC11033520.

[10] Lei, B, Liu, P, Milojevic-Dupont, N, Biljecki F (2024): Predicting building characteris- tics at urban scale using graph neural networks and street-level context. Computers, Environment and Urban Systems, 111: 102129.

[11] I. Hassbring, 'Machine learning embedded automation in financial forecasting: A quantitative case study at Ericsson', Dissertation, 2022.

[12] K. Wang and M. L. Wolverton, "Real estate valuation theory," Springer Science & Business Media, vol. 8, 2002.

Appendix

A. Valuation Neural Network

The valuation neural network employs a feed-forward architecture with the following components:

Listing 1: Valuation Neural Network Architecture

```
class ValuationModel(nn.Module):
    def __init__(self, input_size):
        super(ValuationModel, self).__init__()
        self.fc1 = nn.Linear(input_size, 128)
        self.fc2 = nn.Linear(128, 64)
        self.fc3 = nn.Linear(64, 32)
        self.fc4 = nn.Linear(32, 1)
        self.relu = nn.ReLU()
        self.dropout = nn.Dropout(0.2)

    def forward(self, x):
        x = self.relu(self.fc1(x))
        x = self.dropout(x)
        x = self.relu(self.fc2(x))
        x = self.dropout(x)
```

```
        x = self.relu(self.fc3(x))
        x = self.fc4(x)
        return x
```

B. Rental Estimation Neural Network

The rental estimation neural network uses a similar architecture with smaller layer sizes:

Listing 2: Rental Estimation Neural Network Architecture

```
class RentalModel(nn.Module):
    def __init__(self, input_size):
        super(RentalModel, self).__init__()
        self.fc1 = nn.Linear(input_size, 64)
        self.fc2 = nn.Linear(64, 32)
        self.fc3 = nn.Linear(32, 16)
        self.fc4 = nn.Linear(16, 1)
        self.relu = nn.ReLU()
        self.dropout = nn.Dropout(0.1)

    def forward(self, x):
        x = self.relu(self.fc1(x))
        x = self.dropout(x)
        x = self.relu(self.fc2(x))
        x = self.dropout(x)
        x = self.relu(self.fc3(x))
        x = self.fc4(x)
        return x
```

C. Mortgage Payment Calculation

The monthly mortgage payment is calculated using the standard amortization formula:

For a loan amount L, monthly interest rate r (annual rate divided by 12), and term in months n:

$$P = L \cdot \frac{r(1+r)^n}{(1+r)^n - 1} \tag{8}$$

If the interest rate is 0 (for testing purposes), the formula simplifies to:

$$P = \frac{L}{n} \tag{9}$$

D. Cash Flow Analysis

The monthly cash flow for a property is calculated as:

$$CF = ER - (MP + PT + I + M) \tag{10}$$

where:

- CF = Monthly cash flow
- ER = Effective rental income (after vacancy)
- MP = Mortgage payment
- PT = Monthly property tax
- I = Monthly insurance
- M = Monthly maintenance

The effective rental income accounts for vacancy:

$$ER = GR \times (1 - V) \tag{11}$$

where GR is the gross rental income and V is the vacancy rate.

In addition to core cash flow metrics, the model explicitly incorporates closing costs, property appreciation, and rental income growth for a more realistic investment projection. Closing costs-such as appraisal fees, legal fees, and escrow charges-are typically estimated at 3% of the purchase price and are included in the initial investment calculation, ensuring that up-front capital requirements are accurately modeled. Property appreciation is modeled as an annual percentage increase in value, reflecting both natural market growth and potential forced appreciation from improvements; this is compounded over the investment horizon to project future property values. Rental income growth is also factored in, with annual rent increases applied to estimate future cash flows, aligning with national trends that show rents tend to rise over time-often at rates of 2% to 3.5% per year.

E. Return on Investment Calculations

The cash-on-cash return on investment is calculated as:

$$\text{CoCROI} = \frac{ACF}{I} \times 100\% \tag{12}$$

where ACF is the annual cash flow (monthly cash flow × 12) and I is the initial investment (down payment plus closing costs).

The five-year ROI projection includes both cash flow and property appreciation:

$$\text{5-Year ROI} = \frac{FE + ACF_5 - I}{I} \times 100\% \tag{13}$$

where FE is the projected equity after 5 years and ACF_5 is the accumulated cash flow over 5 years.

F. Property Selection Algorithm

Algorithm 1 IdentifyInvestmentOpportunities(properties, available_capital, min_cash_flow)

Input: properties, available_capital, min_cash_flow
max_price ← available_capital / *down_payment_percentage*
potential_investments ← FILTER(properties, price ≤ max_price)
results ← [ ]
for property **in** potential_investments **do**
  cash_flow ← CALCULATECASHFLOW(property)
  if cash_flow ≥ min_cash_flow **then**
    APPEND(results, (property, cash_flow))
if LENGTH(results) = 0 **then**
  return None
else
  best_property ← SELECTBEST(results)
  success ← PURCHASEPROPERTY(best_property.id)
  if success **then**
    UPDATEPORTFOLIO(months = 1)
    simulation_months ← simulation_months + 1
  return best_property

🆖 Github Link: https://github.com/Chhat2206/Wallstreetbets

The code running for Figure 6 is interactive and located in the github.

```
============================================================

============================================================
REAL ESTATE INVESTMENT SUMMARY
============================================================

PORTFOLIO SUMMARY:
Starting Capital: $2,000,000.00
Current Capital: $3,262,262.32
Current Net Worth: $4,632,293.33
Total Equity: $1,370,031.00
Total Debt: $4,213,480.70
Monthly Cash Flow: $96,324.98
Annual Cash Flow: $1,155,899.78

RETURN ON INVESTMENT:
Total Gain: $2,632,293.33
Total Return: 131.61%
Simulation Period: 2.30 years
Annualized Return: 44.11%
```

PROPERTY PORTFOLIO:

id	address	city	state	zip	purchase_price	current_value	equity	monthly_cash_flow
442969945.0	500 S Pennsylvania Street Bldg Unit 51	Denver	CO	80209.0	$1,300,000.00	$1,510,101.82	$555,083.00	$93,856.79
88287138.0	6533 E Jefferson Ave APT 137	Detroit	MI	48207.0	$89,900.00	$103,650.03	$37,276.85	$913.99
2094447714.0	1946 Ewald Cir	Detroit	MI	48238.0	$425,000.00	$486,346.19	$171,028.15	$825.36
442759580.0	4459 Washington St #7	Roslindale	MA	2131.0	$649,000.00	$737,136.88	$253,304.41	$1,879.24
2054071321.0	3301 Three N 16th St	Philadelphia	PA	19140.0	$815,000.00	$918,772.34	$308,304.75	$1,500.87
131838106.0	1386 S 6th St	Louisville	KY	40208.0	$475,000.00	$531,484.70	$174,032.20	$835.66
102517146.0	6402 TAHOKA ST	San Antonio	TX	78227.0	$1,235,000.00	$1,371,547.88	$437,912.73	$2,138.33
73588489.0	4902 Greenview Ct	Louisville	KY	40216.0	$579,000.00	$638,218.57	$198,534.41	$982.77
88395603.0	3470 Cambridge Ave	Detroit	MI	48221.0	$590,000.00	$645,490.33	$195,468.56	$947.63
7514689.0	3113 W Madison St	Phoenix	AZ	85009.0	$469,000.00	$509,280.94	$149,993.93	$828.19
296620292.0	1333-1339 Barnett Rd	Columbus	OH	43227.0	$450,000.00	$485,002.47	$138,794.58	$711.22
16911307.0	4750 Noyes St APT 123	San Diego	CA	92109.0	$525,000.00	$561,613.58	$156,003.48	$773.25
442771646.0	71-27 162nd St UNIT 2I	Fresh Meadows	NY	11365.0	$448,000.00	$475,667.16	$128,113.10	$646.88
10488926.0	5403 Woodcrest Ave	Philadelphia	PA	19131.0	$525,000.00	$553,262.60	$144,313.71	$748.00
20782669.0	532 N Rossmore Ave APT 412	Los Angeles	CA	90004.0	$555,000.00	$580,512.86	$146,463.42	$752.05
16961815.0	4714 Muir Ave	San Diego	CA	92107.0	$799,000.00	$829,492.57	$202,155.18	$1,037.39
11287151.0	6849 Penn Ave	Pittsburgh	PA	15208.0	$425,000.00	$437,926.78	$102,943.02	$592.84
33337418.0	2557 Blatt Ct	Cleveland	OH	44109.0	$505,000.00	$516,476.79	$116,919.15	$580.30
26926302.0	2220 Canton St APT 208	Dallas	TX	75201.0	$599,800.00	$608,853.42	$132,508.96	$744.06
88426137.0	18073 Fairfield St	Detroit	MI	48221.0	$465,000.00	$468,496.23	$97,842.75	$551.88

TOTAL: 20 properties

```
MONTHLY CASH FLOW BREAKDOWN:
Total Rental Income: $187,894.92
Total Expenses: $76,048.24
Net Cash Flow: $111,846.68

PERFORMANCE METRICS:
Simulation Start Date: 2025-05-27
Simulation End Date: 2030-04-01
Net Worth Growth: 337.24%
Number of Properties Acquired: 20
```