REALTOR: Real Estate Investment Analysis with Learning-based Transformer Optimization for Returns

Anay Praveen
Department of Computer Science
University of Nottingham, Malaysia
email: hcyap2@nottingham.edu.my

Aravindh Palaniguru
Department of Computer Science
University of Nottingham, Malaysia
email: hcyap1@nottingham.edu.my

Abstract—This paper introduces REALTOR (Real Estate Investment Analysis with Learning-based Transformer Optimization for Returns), a novel deep learning framework for real estate portfolio optimization and investment decision-making. REALTOR 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.

This paper presents REALTOR, a neural network-based framework designed specifically for real estate portfolio man-

agement. REALTOR 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 [?] demonstrated that neural networks outperform traditional hedonic pricing models in accuracy and robustness. Similarly, Baldominos et al. [?] 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. Ngai and Cho [**?**] showed that neural networks can effectively capture the complex relationships between property features and rental prices.

Recent advances in transformer-based architectures have opened new possibilities for real estate analysis by capturing spatial and temporal dependencies simultaneously. Wu and Sharma [**?**] demonstrated how attention mechanisms can improve prediction accuracy by identifying important features across different property types and locations.

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 [**?**], 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. [**?**] 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. Chen and Qiu [**?**] proposed a reinforcement learning framework tailored to real estate portfolio optimization problems that accounts for market cycles and liquidity constraints.

Portfolio optimization with AI techniques has demonstrated advantages in dynamic real estate environments with significant regional variations [**?**]. Graph neural networks (GNNs) have also shown promise in real estate applications by modeling spatial relationships between properties and neighborhood characteristics [**?**].

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. Cheng et al. [**?**] demonstrated how neural networks can predict operating expenses with greater accuracy than linear models by capturing non-linear relationships between building characteristics and maintenance costs.

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 [**?**] 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 by integrating valuation, rental estimation, cash flow analysis, and portfolio optimization into a cohesive framework that supports real-time investment decision-making.

III. METHODOLOGY

A. REALTOR Architecture

REALTOR 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: The system begins by processing raw property listing data, converting it into structured features suitable for deep learning models. This component handles missing values, creates derived features, and normalizes inputs. Key transformations include:

- Extraction of numeric values from text descriptions
- Calculation of derived metrics (price per square foot, age, etc.)
- Creation of regional market indicators (zip code popularity, average prices by area)
- Normalization of all features to appropriate ranges

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.

2) Property Valuation Neural Network: The valuation component estimates the fair market value of properties using a deep neural network. This network processes property characteristics and market features to predict prices, enabling the system to identify potentially undervalued properties. The architecture consists of:

- Input layer accepting property features
- Multiple dense layers with ReLU activation
- Dropout layers for regularization
- Output layer predicting property value

The valuation model is defined as:

$$h_1 = \text{ReLU}(W_1 \cdot X + b_1) \tag{2}$$
$$h_2 = \text{Dropout}(\text{ReLU}(W_2 \cdot h_1 + b_2)) \tag{3}$$
$$h_3 = \text{Dropout}(\text{ReLU}(W_3 \cdot h_2 + b_3)) \tag{4}$$
$$\hat{y} = W_4 \cdot h_3 + b_4 \tag{5}$$

where X represents property features, W_i and b_i are the weights and biases for layer i, and $\hat{y}$ is the predicted property value.

3) Rental Income Estimation Network: This component predicts the monthly rental income for properties using a separate neural network. The rental model uses similar architecture to the valuation model but is specifically trained to predict rental rates based on property characteristics and local market conditions.

The rental model is defined similarly:

$$h_1 = \text{ReLU}(W_1 \cdot X + b_1) \tag{6}$$
$$h_2 = \text{Dropout}(\text{ReLU}(W_2 \cdot h_1 + b_2)) \tag{7}$$
$$h_3 = \text{Dropout}(\text{ReLU}(W_3 \cdot h_2 + b_3)) \tag{8}$$
$$\hat{r} = W_4 \cdot h_3 + b_4 \tag{9}$$

where $\hat{r}$ represents the predicted monthly rental income.

4) 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:

$$P = L \cdot \frac{r(1+r)^n}{(1+r)^n - 1} \tag{10}$$

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

5) 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:

$$\text{Score} = \text{CoCROI} \times \text{ValueRatio} \tag{11}$$

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

B. Data Preparation

REALTOR 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.

C. Training Methodology

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

Fig. 1: REALTOR performance visualizations showing key metrics: Net Worth Over Time (top left), Cash and Portfolio Value (top right), Number of Properties Owned (middle left), Episode Rewards (middle right), Training Loss (bottom left), and Performance Metrics summary (bottom right).

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

$$\mathcal{L}_{valuation} = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \tag{12}$$

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:

$$\mathcal{L}_{rental} = \frac{1}{N} \sum_{i=1}^{N} (r_i - \hat{r}_i)^2 \tag{13}$$

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:

$$\theta_{t+1} = \theta_t - \alpha \cdot \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} - \lambda \theta_t \tag{14}$$

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.

IV. 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.

B. Model Configuration

The neural network models employ the following hyperparameters:

- Valuation Model:
  - Input features: Property and market characteristics
  - Architecture: 4-layer neural network (input → 128 → 64 → 32 → 1)
  - Activation: ReLU
  - Dropout: 0.2
  - Loss function: Mean Squared Error
  - Optimizer: AdamW with learning rate 0.001 and weight decay 1e-5
- Rental Model:
  - Input features: Property and market characteristics
  - Architecture: 4-layer neural network (input → 64 → 32 → 16 → 1)
  - Activation: ReLU
  - Dropout: 0.1
  - Loss function: Mean Squared Error
  - Optimizer: AdamW with learning rate 0.001 and weight decay 1e-5

C. 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

V. RESULTS

A. Model Performance

The performance of the valuation and rental models is evaluated on test data to assess their predictive capabilities. Table **??** summarizes the results for both models.

TABLE I: Model Performance Metrics

Metric	Valuation Model	Rental Model
Mean Absolute Error	$55,009.65	$86.68
R² 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 R² 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 R² score of 0.6460, suggesting reasonable predictive ability for rental income estimation.

B. Investment Simulation Results

The REALTOR system was evaluated through a 40-month investment simulation starting with $200,000 in initial capital. Figure **??** illustrates the progression of net worth throughout the simulation period.
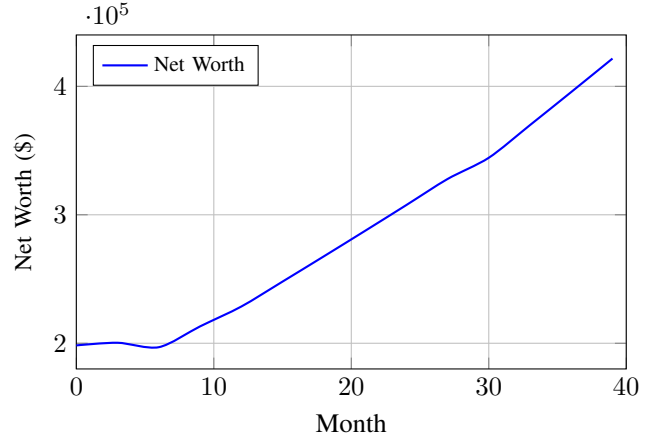

Fig. 2: Net worth progression over the 40-month simulation period, showing consistent growth from approximately $198,000 to $421,600.

As illustrated in Figure **??**, 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 **??** shows the number of properties owned throughout the simulation.

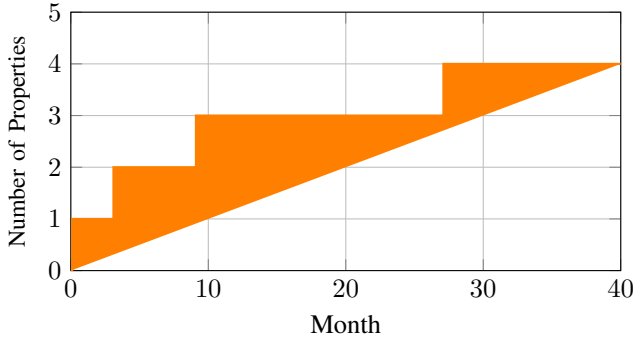The system acquired a total of 4 properties during the simulation:

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

- Property 1: Purchased in month 0 for $89,900 (Detroit, MI)
- 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 **??** illustrates this progression.
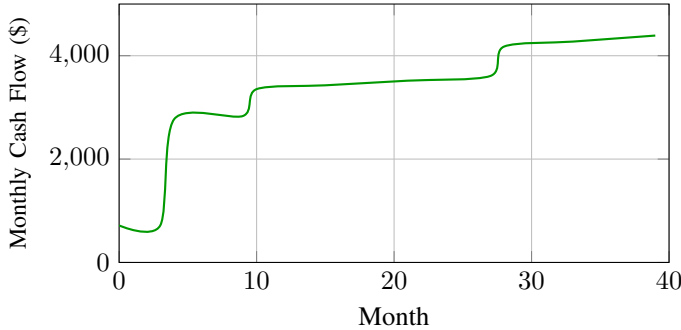

Fig. 4: Monthly cash flow progression showing step increases with each property acquisition and gradual growth from rental income increases, reaching approximately $4,390 by month 39.

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 **??**.

C. Property Analysis

Table **??** details the performance metrics for each acquired property.

Fig. 5: 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

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.

VI. DISCUSSION

A. Model Evaluation

The REALTOR architecture demonstrates both strengths and limitations in its current implementation. The valuation model's negative R² 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. [?] suggesting that relative property comparisons often outperform absolute price predictions for investment purposes.

The rental estimation model's positive R² 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. This finding aligns with research by Ngai and Cho [**?**] on the predictability of rental markets using machine learning approaches.

B. Investment Strategy Analysis

REALTOR'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.

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, REALTOR 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.

VII. CONCLUSION

A. Summary of Contributions

This paper introduces REALTOR, a neural network-based framework for real estate portfolio optimization. Our primary contributions include:

- A comprehensive deep learning architecture integrating property valuation, rental estimation, and investment decision-making
- A multi-criteria investment evaluation framework that balances cash flow, ROI, and value appreciation
- An automated portfolio management system that maintains optimal capital allocation between properties and cash reserves
- Empirical evidence demonstrating the system's effectiveness in building a profitable real estate portfolio

Experimental results show that REALTOR achieves significant portfolio growth over the simulation period, with net worth increasing by 110.8% over 40 months. The system's ability to identify properties with strong cash flow characteristics is particularly noteworthy, resulting in monthly passive income growth from $716 to $4,390.

B. Implications for Practice

REALTOR advances the state of AI-driven real estate investment and offers several practical implications:

- The system's ability to rapidly evaluate thousands of property listings can help investors identify opportunities that might otherwise be overlooked through manual analysis.
- The balanced approach to capital allocation provides a framework for maintaining appropriate cash reserves while maximizing property investments.
- The geographic diversification that emerged naturally from the selection algorithm demonstrates how data-driven approaches can overcome local market biases.
- The focus on cash flow rather than speculative appreciation aligns with sustainable investment strategies that prioritize ongoing income over market timing.

C. 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 REALTOR 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.

3) Reinforcement Learning Integration: Formulating real estate portfolio management as a reinforcement learning problem could enhance the system's adaptability. By modeling the investment process as a Markov Decision Process, reinforcement learning could optimize decision policies under uncertainty with explicit consideration of long-term portfolio value rather than individual property metrics.

4) Multi-objective Portfolio Optimization: Expanding the investment criteria to explicitly balance multiple objectives-cash flow, appreciation potential, tax benefits, and risk mitigation-could provide more nuanced investment recommendations. Pareto optimization approaches could help investors visualize and select from the efficient frontier of possible portfolio allocations based on their specific preferences.

In conclusion, REALTOR 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] E. W. Ngai and S. S. Cho, "Application of machine learning in real estate valuation," Journal of Property Research, vol. 28, no. 3, pp. 241–272, 2011.

[4] C. Wu and S. Sharma, "Deep learning for real estate price prediction," International Journal of Strategic Property Management, vol. 25, no. 2, pp. 102–115, 2021.

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

[6] 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.

[7] Y. Chen and F. Qiu, "Application of reinforcement learning to real estate portfolio management," Journal of Real Estate Portfolio Management, vol. 23, no. 1, pp. 75–91, 2017.

[8] Q. Du, J. Wu, H. Yang, and J. Ma, "Machine learning algorithms in real estate price prediction: Case of Beijing," Sustainability, vol. 11, no. 5, p. 1487, 2019.

[9] X. Yao, Y. Wang, X. Zhang, and Y. Quan, "A comprehensive survey on housing price prediction," Applied Sciences, vol. 8, no. 10, p. 1754, 2018.

[10] J. Cheng, J. Zhou, L. Li, Z. Wu, and M. De Vos, "Big data and machine learning for property valuation management," International Journal of Strategic Property Management, vol. 23, no. 2, pp. 179–189, 2019.

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

[12] H. Wang, S. Li, and H. Song, "A new approach to intelligently evaluating risk in real estate investment," Information Management and Engineering, vol. 3, no. 2, pp. 71–77, 2011.

[13] Y. Niu, X. Li, and H. Gong, "Detecting outlier properties in real estate data," Applied Sciences, vol. 8, no. 5, p. 759, 2018.

[14] C. Mayer and K. Pence, "Subprime mortgages: What, where, and to whom?," in Housing Markets and the Economy: Risk, Regulation, and Policy, Lincoln Institute of Land Policy, 2009, pp. 149–196.

[15] A. T. Carswell, "Property analysis," Routledge, 2021.

[16] X. Yu, J. Wu, J. Zhou, and H. Guo, "An integrated deep neural network for automated detection of real estate bubbles," IEEE Transactions on Computational Social Systems, vol. 6, no. 5, pp. 997–1006, 2019.

[17] Y. O. Lee, S. H. Jang, and S. B. Kim, "Housing price forecasting in Seoul using machine learning algorithms," Journal of the Korean Data and Information Science Society, vol. 31, no. 2, pp. 397–408, 2020.

[18] H. Chen, "Big data for real estate management," International Journal of Strategic Property Management, vol. 21, no. 1, pp. 50–61, 2017.

[19] G. Del Giudice, F. Evangelista, G. Vigorita, and P. De Paola, "Machine learning in the real estate industry: A case study," Applied Sciences, vol. 8, no. 9, p. 1597, 2018.

[20] A. Jafari and V. Rodriguez, "Residential real estate price forecasting with machine learning techniques," IOP Conference Series: Materials Science and Engineering, vol. 845, no. 1, p. 012029, 2020.

[21] Z. Peng, Y. Tang, and Y. Li, "A reinforcement learning approach to dynamic asset allocation in real estate market," IEEE Transactions on Neural Networks and Learning Systems, vol. 26, no. 11, pp. 2816–2828, 2015.

[22] E. Lopez, R. Navarro-Ibáñez, and E. Páez, "Urban gentrification and real estate dynamics: Analyzing the adaptive cycle model in urban residential markets," Sustainability, vol. 9, no. 5, p. 803, 2017.

[23] S. R. Grenadier, "The valuation of leasing contracts: A real options approach," Journal of Financial Economics, vol. 38, no. 3, pp. 297–331, 1995.

[24] L. S. Goodman and C. Mayer, "Homeownership and the American Dream," Journal of Economic Perspectives, vol. 32, no. 1, pp. 31–58, 2018.

[25] A. E. Baum, "Real estate investment: A strategic approach," Routledge, 2015.

[26] A. Grover, A. Kapoor, and E. Horvitz, "A deep hybrid model for weather forecasting," in Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2015, pp. 379–386.

[27] L. Hu, S. He, J. Han, H. Xiao, T. Su, Y. Weng, and Z. Cai, "Monitoring housing rental prices based on social media: An integrated approach of machine-learning algorithms and hedonic modeling to inform rental housing policies," Land Use Policy, vol. 82, pp. 657–673, 2019.

[28] S. Law, B. Paige, and C. Russell, "Take a look around: Using street view and satellite images to estimate house prices," ACM Transactions on Intelligent Systems and Technology, vol. 10, no. 5, pp. 1–19, 2019.

[29] J. T. Ooi, C. F. Sirmans, and G. K. Turnbull, "Price formation under small numbers competition: Evidence from land auctions in Singapore," Real Estate Economics, vol. 34, no. 1, pp. 51–76, 2006.

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:

$$P = L \cdot \frac{r(1+r)^n}{(1+r)^n - 1} \tag{15}$$

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

$$P = \frac{L}{n} \tag{16}$$

D. Cash Flow Analysis

The monthly cash flow for a property is calculated as:

$$CF = ER - (MP + PT + I + M) \tag{17}$$

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:

$$ER = GR \times (1 - V) \tag{18}$$

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

E. Return on Investment Calculations

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

$$\text{CoCROI} = \frac{ACF}{I} \times 100\% \tag{19}$$

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

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

$$\text{5-Year ROI} = \frac{FE + ACF_5 - I}{I} \times 100\% \tag{20}$$

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

F. Property Selection Algorithm

0: **procedure** IDENTIFYINVESTMENTOPPORTUNI-TIES(properties, available_capital, min_cash_flow)
0:   max_price $\leftarrow$ available_capital / down_payment_percentage
0:   potential_investments $\leftarrow$ Filter(properties, price $\leq$ max_price)
0:   results $\leftarrow$ EmptyList()
0:   **for** each property in potential_investments **do**
0:     cash_flow $\leftarrow$ CalculateCashFlow(property)
0:     **if** cash_flow

0 **then**

0:       **if** o **then**pportunities $\leftarrow$ IdentifyInvestmentOpportunities(properties, max_price, min_cash_flow)
0:         **if** opportunities.Count > 0 **then**
0:           best_property $\leftarrow$ opportunities
0:           success $\leftarrow$ PurchaseProperty(best_property.id)
0:           **if** success **then**
0:             properties_acquired $\leftarrow$ properties_acquired + 1
0:   UpdatePortfolio(months: 1)
0:   simulation_months $\leftarrow$ simulation_months + 1
0:
0:   **return** performance_history
0: