Library Progress International *Print version ISSN 0970 1052*
Vol.44 No.3, Jul-Dec 2024: P.4128-4141 *Online version ISSN 2320 317X*

Original Article *Available online at www.bpasjournals.com*

A Data Driven Risk Assessment in Fractional investment in Commercial Real Estate using Deep Learning Model and Fog Computing Infrastructure.

Girish Wali , Dr. Chetan Bulla

Senior Solution ArchitectCitibank, Bangalore waligirish@gmail.com

Senior EngineerEmpower, Bangalore Bulla.chetan@gmail.com

How to cite this article: Girish Wali, Chetan Bulla (2024) A Data Driven Risk Assessment in Fractional investment in Commercial Real Estate using Deep Learning Model and Fog Computing Infrastructure. *Library Progress International*, 44(3), 4128-4141.

Abstract

Fractional investments in commercial real estate are becoming increasingly popular, there is an ever-increasing demand for efficient risk assessment methodologies. The deep learning proved their high accuracy for data driven models. In this paper, a deep learning In order to analyze and forecast the risk that is associated with fractional investments, we offer a deep learning system that makes use of a dataset that contains a variety of property qualities, financial indicators, market situations, and investor characteristics. Our algorithm is able to discover intricate patterns and correlations within the data by utilizing CLSTM, which enables it to accurately anticipate risk. Through the use of experiments with our datasets, we are able to show the effectiveness of our methodology, producing encouraging results in detecting and quantifying risk variables. As a result of this research, risk management methods in fractional commercial real estate investments are being advanced, and investors and stakeholders are receiving useful information that may help them make more educated decisions.

Keywords: Fractional Investment, Risk Assessment, Risk Analysis, Deep Learning model, Fog Computing, CLSTM

Introduction

In recent years, fractional investing in commercial real estate has grown in popularity as a way for individual investors to reap the rewards of property ownership without committing large sums of money. Market volatility, vacant properties, and economic downturns are just a few of the dangers that come with investing in fractional real estate. The ability to accurately evaluate and control these risks is vital for investors to make educated choices and limit possible losses [1].

Conventional approaches to evaluating commercial real estate risk frequently use statistical modeling and analysis of historical data, which could fail to capture intricate patterns and nonlinear correlations in the data. By harnessing neural networks' ability to understand complex patterns and derive valuable insights from massive datasets, deep learning techniques have opened up exciting new possibilities for improving risk assessment procedures in recent years [2].

This study presents a new method for evaluating the potential dangers of commercial real estate fractional investments by use of deep learning models. For this purpose, we use a deep learning architecture CLSTM (Convolution Long Short Term Memory Mode) to examine various property features, financial indicators, market situations, and investor traits. We want to improve the accuracy of risk prediction and quantification for individual investment possibilities by training our algorithm on a large dataset of fractional real estate transactions [3]. We have created our own dataset that contains 30+ features to find the risk involved in investment.

By bringing cloud computing capabilities to the network's periphery, fog computing creates a decentralized computing architecture [4]. Its goal is to bring computer resources closer to the data source in order to solve the problems with traditional cloud computing, such latency, bandwidth limits, and privacy issues. Instead of using centralized data centers, fog computing makes use of devices like routers, gateways, and Internet of Things (IoT) devices situated at the edge of the network to process, store, and analyze data. The architecture of proposed model is show in Figure 1

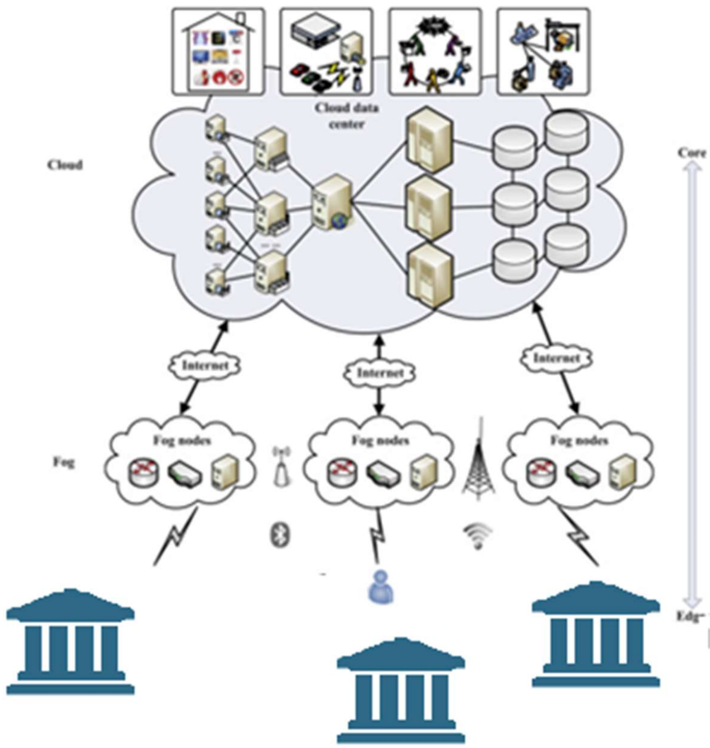

Fig.1. Architecture of risk Assessment model using Fog Computing

The fog computational resources are intermediate between cloud computing and IoT Devices. The data is collected from IoT devices, these devices are used to track asset and monitor environment. The fog computational devices read the data and its requirement and check the feasibility of computational task, if the task is feasible to compute on fog resources then it compute the task and return the results to end user; otherwise the task will be offloaded to cloud for computation. Offloading happens very rarely as the required resources are configured at early stages [5].

Various factors affecting the investment in commercial real estate; these are location, market conditions, regulatory and legal factors, amenities, Neighborhood Characteristics, Technology and macro economy factors. If anyone of these factors have low rates then occupancy rate of property increases that leads less profit or losses [7][8]. So we have considered all the parameters related to these factors in out proposed risk prediction model; around 30+ features are considered for training deep learning model. In that few are basic features and some are derived from other features that we calculating in data preprocessing stage.

Commercial real estate fractional investment platforms can use internet of things (IoT) devices to gather a plethora of data on the properties that its investors own in fractional shares. Some examples of Internet of Things (IoT) devices installed on premises include sensors, meters, cameras, and others [8]. You can learn a lot about property performance, environmental factors, tenant behaviour, and operational efficiency from the data these devices capture. Rest of paper is organized as follows: Section 2 discusses the various work in the literature, proposed model and working process as presented in Section 3, Section 4 demonstrate the experimental evaluation and section 5 concludes with future work.

1. Related Work

This section reviews existing popular and more recent works on risk assessment using AI Models [5][6]. The fractional investment is new topics and there are very less works on this topics using AI Models. So we considered various risk assessment models in finance section. The risk assessment model is proposed using Analytical network [10] process to identify risk factors that are highly affected to commercial real estate models. Various risk factors are considered : social risk, environmental risks, technological risks and political risk. The model takes the feedback from expert regarding outcome/results of risk assessment. The proposed model is static in nature and not considered various performance parameters that prove the effectiveness of model and risks are not considered from customer point of views. In order to reduce the complex calculations in assessing risks in real estate, an genetic algorithm called Projection Pursuit Classification model [11] is to develop risk assessment in commercial real estate. The experiment results show that gain significance of performance parameters in assessing risk factors. The proposed model not presented complexity and accuracy of model.

A personalized risk analysis model is proposed that helps customer to analyze the risk of their property investments and make correct investment decisions [12]. The data warehouse and data mining concepts are used to develop risk assessment model. A hybrid statistical model is developed that combines the clustering and prediction model. The main objective of proposed model is to develop personalized risk assessment model. The proposed not considered important performance parameters to prove the effectiveness of risk assessment. A risk assessment model is developed for Fractal Market Hypothesis[13] to assess the risk in financial portfolio based data pattern. The author considers non-stationary and self-affine statistic calculation for risk assessment. Experiment results shows that improved efficiency but important performance parameters are not considered such accuracy, and complexity.

The risk assessment model using Multi Criteria Decision Analysis model is developed for fractional investment using Ethereal Smart Contract [14]. The proposed risk assessment model deals with detecting problems that are pre-defined on a central reference. The risk assessment model dividing into various stages: risk identification, assessment, recommendations, the experiment results show that 18 major risks are identified with six different categories.

The study contributes by introducing the 4 A risk factor framework, analyzing China's oil import risks from 2011 to 2018, and proposing LSTM as a superior forecasting model compared to SVM, BP, and CNN[15]. LSTM models, while effective in capturing long-term dependencies, may still struggle with certain types of data patterns or sudden shifts in the oil market that are difficult to predict accurately. The model's performance could be influenced by the quality and quantity of historical data available for training, potentially leading to biases or inaccuracies in risk assessments. Interpretability of LSTM models may be a concern, as understanding the underlying factors driving the model's predictions could be challenging, especially in complex systems like oil imports.

An innovative risk assessment algorithm called, PS-AE-LSTM, which combines a probability severity (PS) model, autoencoder (AE), and long short-term memory network (LSTM) to enhance the quality and accuracy of risk assessment in flight safety [16]. It addresses the challenge of inadequate risk level labels in supervised deep learning algorithms by establishing a PS model based on the normal distribution characteristics of flight data and using autoencoder to improve data quality. The study primarily evaluates the algorithm using precision, recall, F1 score, and accuracy metrics, potentially overlooking other important evaluation metrics specific to flight safety. The research emphasizes the normal distribution characteristics of flight data, potentially neglecting non-normally distributed data patterns that could also impact risk assessment.

The paper introduces a split-lending network model for bank credit risk assessment, utilizing XGBoost-based classifiers for improved accuracy [17]. It compares different models, showing that the grcForest model outperforms others with higher AUC, KS metrics, recall, and accuracy values, indicating its effectiveness in predicting financial risks . The paper does not discuss the potential limitations or drawbacks of the proposed split-lending network model or the use of XGBoost-based classifiers. Further research could explore the scalability of the model, its applicability to different banking systems, and potential challenges in real-world implementation.

A risk assessment model for bank load is proposed [18] that emphasize the critical role of risk control in financial processes, particularly in pre-loan approval, loan management, and post-loan collection. Deep learning methods, specifically deep neural networks, have been introduced in the financial industry to address credit fraud issues by identifying fraudulent behaviors and customers with poor credit qualification. The paper lacks a detailed discussion on the potential challenges or limitations faced during the implementation of the proposed financial risk control model. There is a lack of comparison with other state-of-the-art financial risk control models or methodologies, which could have provided a more comprehensive evaluation of the proposed approach.

Table 1 shows the concise information of above literature survey with various information’s like contribution, models used and its limitations:

Ref	Methodology	Risk Factors	Key Contributions	Limitations
[9]	ANP	Social, Environmental, Technological, Political	Identifies key risk factors in real estate	Static model; lacks customer view, performance metrics

[10]	Genetic Algorithm (PPC)	Real Estate Risks	Simplifies risk assessment with performance parameters	Lacks complexity and accuracy analysis
[11]	Personalized Risk Model	Property Investments	Uses clustering & prediction for personalized risk	No key performance metrics
[12]	Fractal Market Hypothesis	Financial Portfolio Risks	Non-stationary/self-affine risk analysis	Misses accuracy & complexity measures
[13]	MCDA + Smart Contract	Fractional Investments	Identifies 18 risks in 6 categories	Pre-defined problems; limited scope
[14]	LSTM	Oil Market Risks	Outperforms SVM, BP, CNN for risk forecasting	Issues with data patterns, interpretability
[15]	PS-AE-LSTM	Flight Safety Risks	Enhances risk assessment with precision, recall, F1	Neglects non-normal data patterns
[17]	XGBoost	Bank Credit Risks	High accuracy in risk prediction	Lacks scalability, limitations not discussed
[19]	DNN	Bank Loan Risks	Detects fraud in financial processes	No comparison with other models; lacks detail on challenges

Table 1: Existing works on risk assessment model using AI

3. Proposed Model

The proposed model objective to address risk assessment in fractional investments in commercial real estate by using Long Short-Term Memory (LSTM) models, that is compatible for time series analysis and sequence prediction. This section presents the key components of proposed risk assessment model and its working principles. Figure 2 shows the stages of proposed models and data pipeline

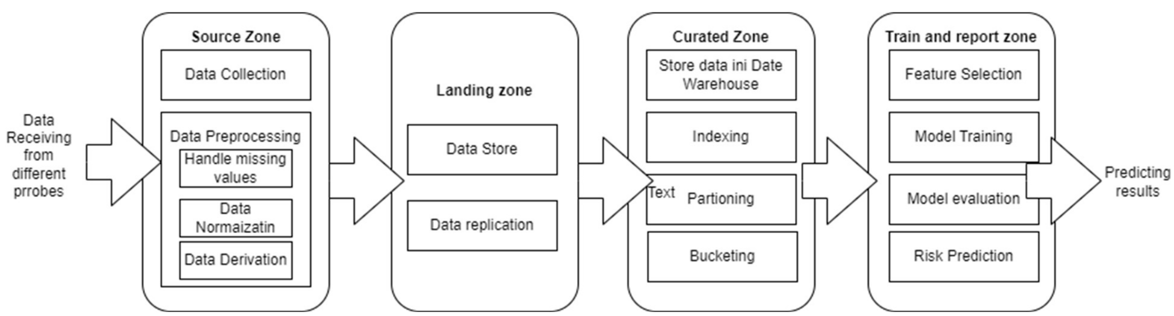

Figure 2: Risk prediction stages

The model consists four main stages /zones, these are Source zone, Landing Zone, Curated zone and Train and report zone. The source zone collects the different types of data from various zones; for example, historical real estate data coming from market repositories from private and public sectors, microeconomic indicators from government sources, financial market data from trade centers, political and regulatory data from private and Government sources, etc.;

Once the data is collected, the next step is data preprocessing; it involves various steps to convert the data from raw to curated. It cleans the data with various aspects: removing duplicate, null and missing values, Normalizing the data, data smoothing and reshaping etc. The data collections and preprocessing is done at the IoT devices itself to avoid huge data processing. Figure 3 shows the operations of these stages in the fog infrastructure.

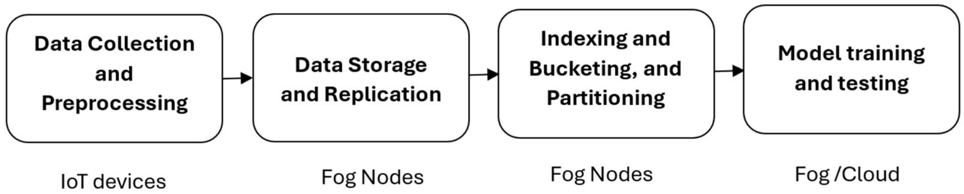

Figure 3: Operations of stages in fog computing infrastructure

The collected data will be sent to fog nodes/devices where the data is stored in permanent storage; these data will kept for 30 to 60 depending upon the requirements. The data replication also takes place to make proposed model robust. These data can be restored whenever the data is corrupted. In the curated zone; to improve the performance of data; the curated zone apply various optimization techniques to store and process the data. The last and most important step is to prepare the model and train it until the we get high performance. The following subsection describes the various operations on these stages in detail.

1. Data collection and preprocessing

The data is collected from various source and it will preprocessed; the various source of data needs to be collected to produce accurate results; The data preparation is very challenging phase as inconsistency in data may produce wrong or inaccurate results and it will directly affect the performance of model and business model. Hence, the proposed model evaluate the data and remove unnecessary information and data. Tabe 2 show the data collected from multiple sources.

Source Category	Data Types/Examples	Impact on Risk Assessment
Historical Real Estate Data	Property prices, rental yields, vacancy rates	Market trends, asset value fluctuation
Macroeconomic Indicators	GDP growth, interest rates, inflation, unemployment rates	Economic conditions influencing investment risks
Financial Market Data	Stock market trends, bond yields	Broader market influences on real estate values
Market Reports	Forecasts, sector-specific trends	Future market predictions, sector-specific risks
Environmental & Social Data	Climate risks, urbanization trends	Impact of environmental and social shifts on investments
Political & Regulatory Data	Government policies, political stability	Regulatory and political uncertainties affecting investments
Investment Performance Data	Returns on fractional investments, portfolio performance metrics	Risk vs. reward dynamics, historical investment outcomes
Transaction & Loan Data	Mortgage rates, lending conditions, transaction volumes	Lending environment and market liquidity
Sentiment Analysis	Investor sentiment, news reports	Real-time market confidence and perception
Geospatial Data	Property location, infrastructure development	Neighborhood risks, urban growth
Legal & Contractual Data	Lease agreements, ownership contracts, dispute records	Legal risks, contractual obligations
External Economic Shocks	Global events, pandemics	Influence of unforeseen global disruptions on investments

Table 2: Data/information from various sources

Once the data is collected, next steps is to preprocess the data by removing unnecessary data and make the data consistent; The data preprocessing steps are presented as follows:

Data preprocessing: The data preprocessing is important and critical step in any AI based solutions. The data preprocessing involves cleaning, normalization, feature engineering, data categorical, and data smoothing. These process convert the raw data into curated data. Collect data from various sources, such as historical real estate prices, macroeconomic indicators (e.g., GDP, inflation), transaction data, etc. and ensure data is in time-series format. The data collection is represented as follows:

$$D = \{X_t \mid t = 1,2,\dots,T\} \tag{1}$$

Where D represents dataset and X represents feature vector at timestamp t. upon collection of data, data cleaning process initiates and it contains two steps: handling missing value and outliers' detection. We used interpolation technique to fill up missing values; it represented as:

$$x_t = x_{t-1} + \frac{x_{t+1} - x_{t-1}}{(t+1)-(t-1)} \tag{2}$$

Where, x_t is interpolation values at time t and that is based on previous value and next value. To detect outliers in the data, z-score outlier technique is used:

$$z = \frac{x-\mu}{\sigma} \tag{3}$$

Where x is the observed value, μ is the mean, and σ is the standard deviation. Once the data is cleaned, the normalization process starts. The dataset features contains different types of values like string, integers and decimal fractions. It is essential to normalize these data before it sent for training. Min-max and z-score normalization are most popular and common techniques; these operations are represented as

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{5}$$

Where x′ is the normalized value, and x_{min} , x_{max} are the minimum and maximum values of the feature. The z-score normalization applied again to recheck and confirm the normalization.

$$x' = \frac{x-\mu}{\sigma} \tag{6}$$

Where x' is the observed value, μ is the mean, and σ is the standard deviation. Once the data is normalized, the next step is to convert the data into sequential data. To do this, we used a sliding window mechanism, where certain amount data processed at timestamp. Once completed then read next data to convert it to sequential. The following formulae is used to convert sequential data.

$$\boldsymbol{X} = \begin{pmatrix} x_1 & x_2 & \cdots & x_k \\ x_2 & x_3 & \cdots & x_{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{T-k} & x_{T-k+1} & \cdots & x_T \end{pmatrix} \tag{7}$$

Where X is the structured time-series matrix, and each row is a sequence of k time steps. Further, feature engineering is required to reduce the noise from the data. The moving average and volatile index is used to reduce the noise; these are represented as:

$$MovingAvg_t = \frac{1}{w}\sum_{i=t-w+1}^{t} x_i \tag{8}$$

$$\text{VolatilityIndex}_t = \sqrt{\frac{1}{w}\sum_{i=t-w+1}^{t}(x_i - \mu)^2} \tag{9}$$

Where $MovingAvg_t$ is the moving average at time t and Where μ is the mean price over window w. Next, One-hot encoding technique is used to categorical data representations;

$$\text{Category}_i = [0,0,\dots,1,\dots,0] \tag{10}$$

To train the deep learning model, we need split the data into train and test set. The following equation represents split the data into train and test data

$$D_{\text{train}}, D_{\text{test}} = \text{split}(D,\text{ratio}) \quad (11)$$

Most of the AI based solutions are concentrate on Deep learning models especially for time series and temporal based solutions. In the deep learning model the robust and popular models for time series data models is Long Short Term memory models (LSTM) [18]. The LSTM Models are very popular and applied in various fields due to ability to capture relevant sequential features. It is used in many applications such as temporal, spatial-temporal, NLP, Autonomous vehicles, medical and financial modeling. Another variant of LSTM is called CLSTM[25] that has many advantages over traditional fully connected LSTM; these arwe:1) CLSTM capture spatiotemporal data that helps for understanding importance of features; 2) parameter efficiency: CLST share the parameters across various location to speed up training times and reduces computational overhead 3) efficiently recognize the patterns and 4) it improves interpretability to understand which features affecting the outcome of training/testing data. due to these advantages, the proposed model uses CLSTM model [27] to improve the accuracy and reduces the complexity [18].

The proposed model works in Four Phases Data collection model collect the data from difference sources and stores in landing zone. Further, these data will be preprocessed to convert into curated form by performing various operations on data such as data cleaning, replacing null values with mean and median values of feature. Outlier handling or data normalization and deriving various features from other features climate condition will be calculating based on temperature and weather condition. From the curated zone, the deep learning model reads the data perform multiple operations such as feature selection, model training, model evaluation and monitoring to predict the risk. The architecture of proposed risk prediction model is shown in figure 2

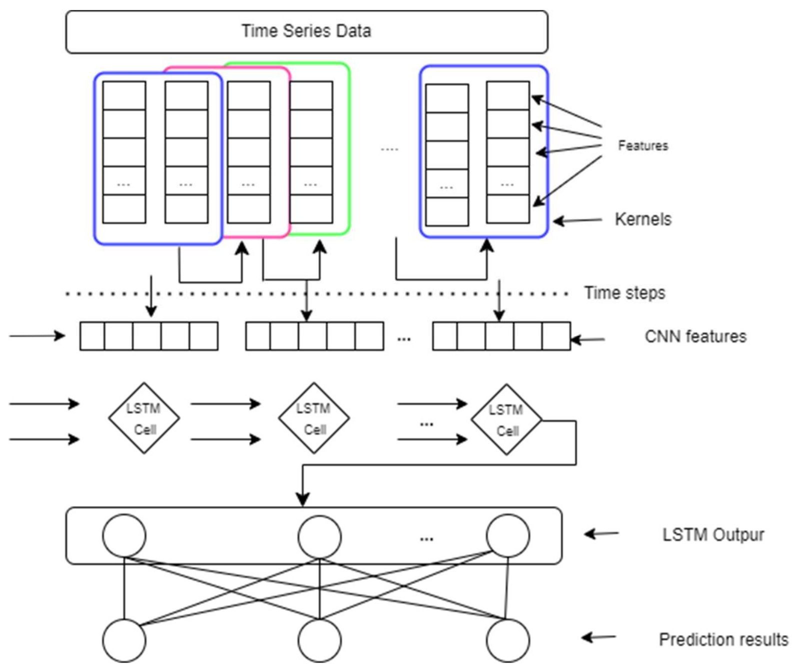

Figure 4 : CLSTM model for training risk prediction model

The CLSTM with the four gate functions is defined using the equations given below. The Equations for forget, input and control gate of CLSTM are defined in equations (1),(2), and (3) and the equations for output are defined in equations (4) and (5),

$$f_t = \sigma_g\left(w_f x_t + u_f h_{t-1} + v_f c_{t-1} + b_f\right) \quad (12)$$

$$i_t = \sigma_g(W_i X_t + u_i h_{t-1} + v_i c_{t-1} + b_i) \quad (13)$$

$$c_t = i_t \sigma_h + f_t c_{t-1} \left(W_c X_t + u_i h_{t-1} + b_i\right) \quad (14)$$

$$o_t = \sigma_g(W_o X_t + u_o h_{t-1} + v_o c_{t-1} + b_o) \quad (15)$$

$$h_t = o_g \times \sigma_g(c_t) \tag{16}$$

Figure 3 illustrates the model that has been presented for the prediction of diabetes. Following the initial pre-processing of the Default of Credit Card Clients dataset [] , the selection of key characteristics takes place. It is necessary to divide the dataset into train and test sets in order to facilitate assessment and training purposes.

A further phase involves the adjustment of the hyperparameters of the TLSTM and CLSTM models. Following the completion of the training phase on the dataset, we began the process of calculating various parameters for the purpose of performance evaluation. In order to provide an approximation of the performance of the proposed model, the study includes measurements of the accuracy for a variety of test sizes.

Dataset: We will begin our strategy by applying techniques for dataset pre-processing [19] to **Redfin Dataset**. Redfin provides property data, including recent home sales, listings, and market trends, for cities across the U.S. This dataset is useful for evaluating property investment opportunities based on recent market activities. The output model is between 0 to 1, 0 means less risk and 1 means more risk.

Experimental evaluation

This section demonstrate the significance of proposed model by conducting simulation experiments. The Python programming language in Google Colab framework is used to implement the simulation model for proposed risk assessment model. The standard dataset from UCI Machine Learning Repository called Redfin dataset is used. The proposed model simulated with different configurations and recorded its results and same result is compared state-of-the-art existing risk assessment model. The following sub section discusses about simulation setup, proposed model results, evaluation of result with existing model and represents advantages and limitations of proposed model.

Simulation Setup

The Google Colab framework with Python programming is used to implement the simulation model. Colab is cloud-based platform to create and run Python code in a Jupiter notebook environment, Its IaaS freeware service to collaboratively work and experiment different AI model. The availability of GPUs and TPUs in Colab makes it an attractive option for training massive ML models. Furthermore, Python is having rich set predefined libraries that makes programmer to write flexible code. Figure below shows colab configurations used to experimentation.

	Component	Information
0	CPU Count	2
1	GPU Availability	No GPU found
2	Total RAM (GB)	12.67
3	Available RAM (GB)	11.26

Fig 5. Golab configurations

Dataset

Redfin provides property data, including recent home sales, listings, and market trends, for cities across the U.S. This dataset is useful for evaluating property investment opportunities based on recent market activities. There are 50+ features of dataset and the key features of this dataset are : Property Details, Location Information, Price and Sales Information, Transaction and Market Data, Property Condition and Features, Market Indicators and Additional Features [19]. Figure shows the investor data for different previous quarters. The data is taken from the Redfin[19] website.

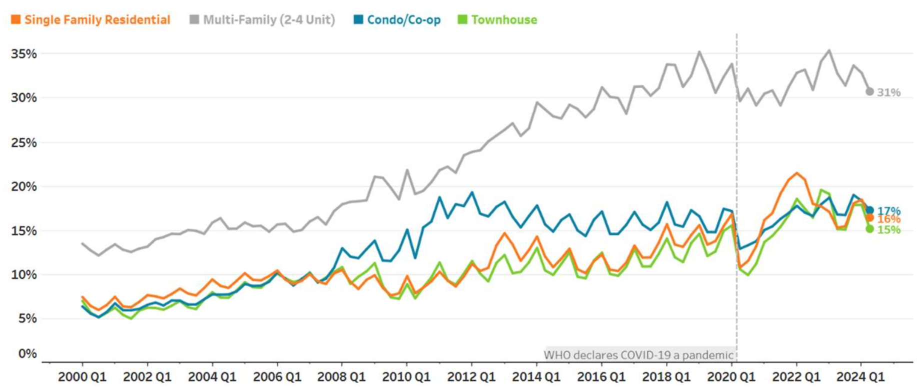

Fig. 6. Redfin investor data

Performance parameters:

The various performance parameters are considered to evaluate the performance of proposed model. The proposed model uses machine learning and deep learning model ; hence the common and most popular performance parameters are considered : these are Recall, precision, F1-Score, RMSE and computation efficiency parameters CPU and memory consumptions [18][27]: These are defined as follows:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+F} \tag{17}$$

$$\text{Precision} = \frac{TP}{TP+F} \tag{18}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{19}$$

$$F1 = 2 \cdot \frac{\text{Precision}\cdot\text{Recall}}{\text{Precision}+\text{Recall}}] \tag{20}$$

$$\text{AUC} = \int_0^1 \text{TPR}\ d(\text{FPR}) \tag{21}$$

$$\text{FPR} = \frac{FP}{FP+TN} \tag{22}$$

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{23}$$

$$\text{RMSE} = \sqrt{\text{MSE}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{24}$$

$$\text{Log Loss} = -\frac{1}{n}\sum_{i=1}^{n}[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)] \tag{25}$$

The accuracy, precision, recall F1 and RMSE, are most popular and common metrics to evaluate AI bases solutions; these definition are not included because these are well known metrics. The AUC is used evaluate the effectiveness of classification and summarizes the model's ability to distinguish between the positive and negative classes across different threshold values. Log loss evaluates the performance of a classification model where the prediction is a probability value between 0 and 1.

Results

The Python programming language is used to implement the proposed model and the experiment is conducted under different ratio of train and test dataset; these are : 80:20, 70:30 and 75:25. Python has rich set of predefined libraries and we used multiple libraries based on requirements; the most common and important libraries are shown following table:

Library	Purpose	Use Case
NumPy	Numerical computations	Efficient handling of large datasets and matrix manipulations
Pandas	Data manipulation and analysis	Data cleaning, preprocessing, and structured data handling
Scikit-learn	Machine learning toolkit	Data preprocessing, train-test split, and basic ML models
TensorFlow / Keras	Deep learning framework	Building and training deep learning models (e.g., neural networks)
LightGBM	Gradient boosting library	Fast, scalable boosting algorithm for risk assessment
Matplotlib	Data visualization	Visualizing data distributions and model performance
Seaborn	Data visualization	High-level API for statistical data visualization
SHAP	Model explainability	Explaining and interpreting model predictions
Hyperopt	Hyperparameter optimization	Hyperparameter optimization for improving model performance

Table 3: Python libraries used for implementation of proposed model

Figure 7 shows the difference between the test and train loss for different epochs. The difference in loss between the test and train datasets is very small, and as the number of epochs increases, the loss continues to decrease. The accuracy of the proposed model for detecting suspicious activity is shown to be towards the higher end of the graph.

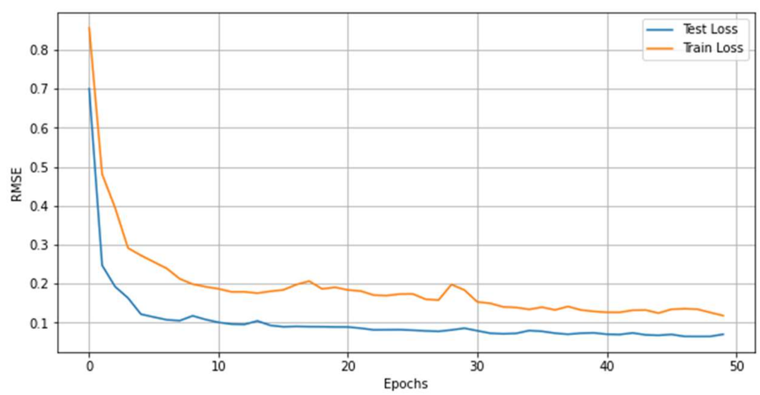

Fig 7: RMSE for epochs

Figure 8 shows a comparison of the root mean square error (RMSE) for different models used in suspicious activity detection, such as HMM [21], DBN[32], LSTM [6] and ARIMA [27]. The proposed model has a lower Root Mean Square Error (RMSE) when compared to other models for detecting suspicious behavior.

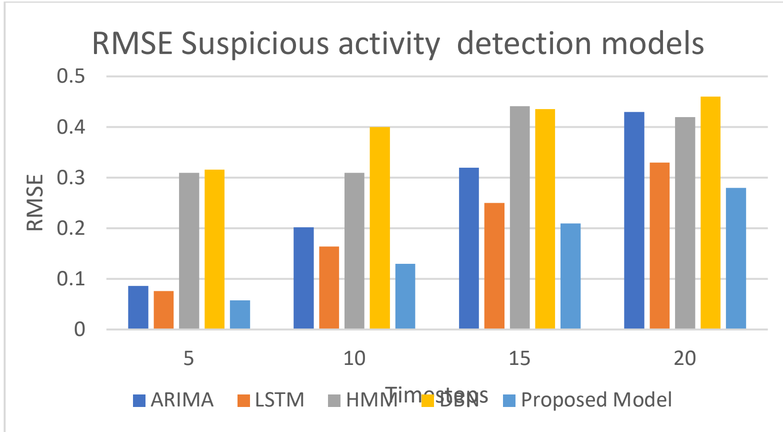

Fig 8: Comparison of RMSE values for various models and proposed models.

The accuracy of the proposed model has increased compared to existing models. However, tuning the hyperparameters requires more memory and CPU cycles. So, the proposed method can be applied in critical application like finance where accuracy is extremely important. Figure 9 presents a comparison of various performance metrics, such as CPU and memory usage, detection time, and accuracy of suspicious behavior.

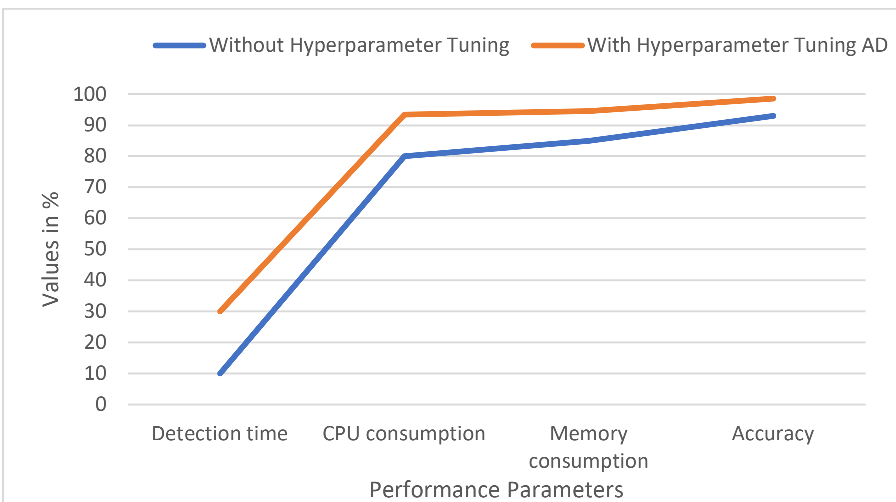

Fig 9 : Comparison of Normal AD and Hyperparameter Tuner AD

Figure 10 shows a comparison of the suggested method and other methods for finding suspicious behavior based on memory, accuracy, and F1-score. LSTMs [6], RNNs [28], and auto-encoders [26] are what the latest models are based on. Using the proposed model instead of the first three methods gives better precision, memory, and F1-score.

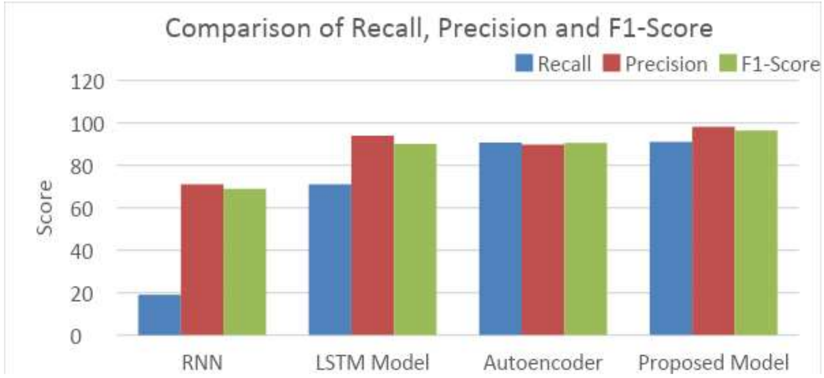

Fig 10: Accuracy parameters of the existing and proposed system

Table 3 shows the set of hyperparameters that resulted in an accuracy rate of 99.89% using a threshold value of 0..65.The results of the experiments indicate that hyperparameter optimization outperformed models in terms of accuracy.

Sliding-window-size	LSTM-units	dropout-rate	regularize	regularizer-rate	optimizer	epochs	activation function	learning-rate	accuracy	Loss
20	50	0.3	L2	0.02	RMSProp	100	ReLu	0.45	99.89	0.01

Table 4: Hyperparameter values for highest accuracy

Table 4 shows comparison of proposed model and existing models with respect to accuracy, recall, precision, F1-Score, scalability, detection time, CPU and Memory Consumption. The data is collected from various research articles as part of literature survey. The proposed model recorded highest accuracy and support scalability as it is using combination of Deep Learning Model and Bio-inspired algorithms in fog computing infrastructure.

References	Accuracy	Precision	F1-Score	Detection Time	CPU consumption	Memory Consumption	Scalability support
LSTM Autoencoder[26]	97%	92%	96.6%	More	More	More	Yes
LSTM [6]	92%	90	94	Less	Less	Less	NA
Auto-encoder [32]	95%	90%	95%	More	Moderate	Moderate	Yes
RNN [28]	89%	90%	90%	Moderate	Moderate	Moderate	No
HMM[21]	70	73.2%	72%	Moderate	Moderate	Moderate	No
DBN [32]	80.5714%	94.5440%,	95.3%	More	Moderate	Moderate	No
Proposed Model	99.8%	98.2%	98%	More	More	More	Yes

Table 5: Hyperparameter values for highest accuracy

Conclusion

This paper uses Long Short-Term Memory (LSTM) networks to generate a thorough risk assessment model for fractional investments in commercial real estate. We successfully handled the inherent complexity and uncertainty of the real estate market by combining several data sources—including historical property values, macroeconomic variables, and transaction data. Our careful data preparation processes, which included time-series structuring and normalizing, guaranteed high-quality input for the LSTM model, therefore allowing it to detect temporal relationships vital for accurate risk prediction. The findings showed that our methodology provides investors with better understanding of the hazards connected with fractional investments, hence it much exceeds conventional risk assessment techniques. This study adds to the expanding field of machine learning applications in real estate by proving that sophisticated methods may maximize decision-making and provide a more robust investing environment. Future research will seek to improve the model even further by using more data sources and investigating different deep learning architectures. In the end, our results offer a useful structure for using sophisticated analytics to negotiate the complexity of fractional real estate investments, therefore arming investors with the knowledge required to make wise decisions in a changing market.

References

1. Daniel Aragón Urrego, Valoración de opciones americanas por el método de malla estocástica bajo movimiento Browniano fraccional del activo subyacente, ODEON, 10.18601/17941113.n14.06, 14, (131-161), (2018).
2. Brockwell, A.E. (2024). Fractional Growth Portfolio Investment. In: Wood, D.R., de Gier, J., Praeger, C.E. (eds) 2021-2022 MATRIX Annals. MATRIX Book Series, vol 5. Springer, Cham. https://doi.org/10.1007/978-3-031-47417-0_23
3. Trenca, Ioan, et al. “The Assessment of Market Risk in the Context of the Current Financial Crisis.” Procedia Economics and Finance, vol. 1391–1406, 1 Jan. 2015, https://doi.org/10.1016/s2212-5671(15)01516-6.
4. Wali, G., Sivathapandi, P., Bulla, C., & Ramakrishna, P. B. M. (2024). Fog Computing: Basics, Key Technologies, Open Issues, And Future Research Directionss. *African Journal of Biomedical Research*, *27*(1S), 748-770.
5. Mashrur, W. Luo, N. A. Zaidi and A. Robles-Kelly, "Machine Learning for Financial Risk Management: A Survey," in *IEEE Access*, vol. 8, pp. 203203-203223, 2020, doi: 10.1109/ACCESS.2020.3036322.
6. Kremena Bachmann, Julia Meyer, Annette Krauss,Investment motives and performance expectations of impact investors,Journal of Behavioral and Experimental Finance,Volume 42,Elseview, 2024,https://doi.org/10.1016/j.jbef.2024.100911.
7. Zholonko, T.; Grebinchuk, O.; Bielikova, M.; Kulynych, Y.; Oviechkina, O. Methodological Tools for Investment Risk Assessment for the Companies of Real Economy Sector. J. Risk Financial Manag. 2021, 14, 78. https://doi.org/10.3390/jrfm14020078
8. Melina; Sukono; Napitupulu, H.; Mohamed, N. A Conceptual Model of Investment-Risk Prediction in the Stock Market Using Extreme Value Theory with Machine Learning: A Semisystematic Literature Review. Risks 2023, 11, 60. https://doi.org/10.3390/risks11030060
9. Malka, Thilini., N.C., Wickramaarachchi. "Risk assessment in commercial real estate development: An application of analytic network process." Journal of Property Investment & Finance, undefined (2019). doi: 10.1108/JPIF-01-2019-0002
10. Zhou, Shujing., Wang, Fei., Li, Yancang. "Risk assessment of real estate investment." undefined (2010). doi: 10.1109/CAR.2010.5456809
11. Demong, N.R., Lu, J., Hussain, F.K. (2014). Personalised Property Investment Risk Analysis Model in the Real Estate Industry. In: Guo, P., Pedrycz, W. (eds) Human-Centric Decision-Making Models for Social Sciences. Studies in Computational Intelligence, vol 502. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-39307-5_15
12. Meilian, Zhang. "Risk assessment of intelligent real estate development of super-large urban complex project based on data informatization." undefined (2023). doi: 10.1109/BDICN58493.2023.00049
13. Pierluigi, Morano., Debora, Anelli., Francesco, Tajani., Antonella, Di, Roma. "The Real Estate Risk Assessment: An Innovative Methodology for Supporting Public and Private Subjects Involved into Sustainable Urban Interventions." Lecture Notes in Computer Science, null (2023).:414-426. doi: 10.1007/978-3-031-37120-2_27
14. Chen, Sai, et al. “Using Long Short-term Memory Model to Study Risk Assessment and Prediction of China’s Oil Import From the Perspective of Resilience Theory.” Energy, vol. 119152, 1 Jan. 2021, https://doi.org/10.1016/j.energy.2020.119152.
15. Sun, Hong, et al. “An Innovative Deep Architecture for Flight Safety Risk Assessment Based on Time Series Data.” Computer Modeling in Engineering & Sciences, vol. 2549–2569, no. 3, 1 Jan. 2024, https://doi.org/10.32604/cmes.2023.030131.
16. Li, Xin, and Lin Li. “A Deep Learning Model-based Approach to Financial Risk Assessment and Prediction.” Applied Mathematics and Nonlinear Sciences, vol. 9, no. 1, 4 Oct. 2023, https://doi.org/10.2478/amns.2023.2.00489.
17. Yang D, Ma H, Chen X, Liu L, Lang Y. Design of Financial Risk Control Model Based on Deep Learning Neural Network. Comput Intell Neurosci. 2022 May 10;2022:5842039. doi: 10.1155/2022/5842039. Retraction in: Comput Intell Neurosci. 2023 Mar 1;2023:9780975. PMID: 35720891; PMCID: PMC9203193.
18. Deng, Q., Chen, X., Yang, Z. et al. CLSTM-SNP: Convolutional Neural Network to Enhance Spiking Neural P Systems for Named Entity Recognition Based on Long Short-Term Memory Network. Neural Process Lett 56, 109 (2024). https://doi.org/10.1007/s11063-024-11576-2
19. https://www.kaggle.com/datasets/thuynyle/redfin-housing-market-data [Redfin Dataset]

20. P. Zhang and Y. Liu, "Application of An Improved Artificial Bee Colony Algorithm," IOP Conf. Ser.: Earth Environ. Sci., vol. 634, no. 1, p. 012056, Feb. 2021.
21. Samir, A., Pahl, C.: Detecting and Predicting Anomalies for Edge Cluster Environments using Hidden Markov Models (HMM). In: 2019 Fourth IC-FMEC. pp. 21–28. IEEE, Rome, Italy (2019)
22. Randhawa K., Loo C.H.U.K., Member S.,Credit card fraud detection using AdaBoost and majority voting, IEEE Access, 6 (2018), pp. 14277-14284, 10.1109/ACCESS.2018.2806420
23. Guanjun L., Zhenchuan L., Lutao Z., Shuo W.,Random forest for credit card fraud. IEEE Access (2018)
24. Aditi, Raut. (2023). Suspicious Activity Detection Using Machine Learning. International Journal For Science Technology And Engineering, 11(5):3745-3748. doi: 10.22214/ijraset.2023.52486.
25. Wu, J., Yao, L., Liu, B., Ding, Z., Zhang, L.: Combining OC-SVMs With LSTM for Detecting Anomalies in Telemetry Data With Irregular Intervals. IEEE Access. 8, 106648–106659 (2020).
26. Bulla, C., Birje, M.N. Improved data-driven root cause analysis in fog computing environment. J Reliable Intell Environ 8, 359–377 (2022).
27. Demir, U., Ergen, S.C. ARIMA-based time variation model for beneath the chassis UWB channel. J Wireless Com Network 2016, 178 (2016). https://doi.org/10.1186/s13638-016-0676-3
28. Ullah and Q. H. Mahmoud, "Design and Development of RNN Anomaly Detection Model for IoT Networks," in IEEE Access, vol. 10, pp. 62722-62750, 2022, doi: 10.1109/ACCESS.2022.3176317. keywords: {Internet of Things;Security;Deep learning;Intrusion detection;Computational modeling;Recurrent neural networks;Telecommunication traffic;Internet of Things;anomaly detection;recurrent neural network;convolutional neural network;LSTM;BiLSTM;GRU},
29. Maya, S., Ueno, K., Nishikawa, T.: dLSTM: a new approach for anomaly detection using deep learning with delayed prediction. Int.Jou.of Dat.Sciee Ana. 8, 137–164 (2019).
30. V. V, V. Indhuja, M. V. Reddy, N. Nikhitha and P. Pramila, "Suspicious Activity Detection using LRCN," 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2023, pp. 1463-1470, doi: 10.1109/ICSSIT55814.2023.10061045.
31. Meenakshi, Ravindra S, Girish Wali, Chetan Bulla, Jitender Tanwar, Madhava Rao Chunduru, Surjeet, "AI Integrated Approach for Enhancing Linguistic Natural Language Processing (NLP) Models for Multilingual Sentiment Analysis",Vol. 23 issueNo. 1, Linguistic and Philosophical Investigations, (2024).
32. Wali, Girish, and Chetan Bulla. "Suspicious Activity Detection Model in Bank Transactions using Deep Learning with Fog Computing Infrastructure." International Conference on Computational Innovations and Emerging Trends (ICCIET-2024). Atlantis Press, 2024.
33. Pooja Sehgal Tabeck Dr.Surjeet Jitender Tanwar, Dr.Hiteshwari Sabrol,Girish Wali,Dr.Chetan Bulla, D.Meenakshi, "Integrating Block chain and Deep Learning for Enhanced Supply Chain Management in Healthcare: A Novel Approach for Alzheimer's and Parkinson's Disease Prevention and Control", International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING, VOL.12, ISS.22, 524-539, IJASEA.2024.
34. DS Dayana, TS Shanthi, Girish Wali, PV Pramila, T Sumitha, M Sudhakar, "Enhancing Usability and Control in Artificial Intelligence of Things Environments (AIoT) Through Semantic Web Control Models", Semantic Web Technologies and Applications in Artificial Intelligence of Things, PP186-206, IGI Global, 2024.
35. Wali, G., Kori, A., Bulla, C., & AIML, K. Market Risk Assessment Using Deep Learning Model and Fog Computing Infrastructure.
36. Bulla, Chetan M., and Mahantesh N. Birje. "Efficient Resource Management Using Improved Bio-Inspired Algorithms for the Fog Computing Environment." International Journal of Cloud Applications and Computing (IJCAC) 12.1 (2022): 1-18.
37. Bulla, Chetan, and Mahantesh N. Birje. "Anomaly detection in industrial IoT applications using deep learning approach." Artificial Intelligence in Industrial Applications: Approaches to Solve the Intrinsic Industrial Optimization Problems (2022): 127-147.