AI-Based LSTM-X Model for Risk Assessment in Fractional Commercial Real Estate Investments

Girish Wali

Senior Solution Architect
Citibank, India
waligirish@gmail.com

How to cite this article: Girish Wali (2024). AI-Based LSTM-X Model for Risk Assessment in Fractional Commercial Real Estate Investments. *Library Progress International*, 44(3), 10706-10722

ABSTRACT

This paper introduces a novel approach powered by artificial intelligence for risk evaluation in fractional commercial real estate transactions. Though it involves special risks owing to market volatility, property performance uncertainty, and economic considerations, fractional investing has become well-known as it permits smaller capital involvement in high-value assets. We present a Long Short-Term Memory (LSTM-X) model integrating external factors to efficiently analyze market dynamics and forecast dangers. Comparing our method to conventional statistical and artificial intelligence models shows better accuracy in predicting changes in property value, rental revenue variations, and related investment hazards. Python in Google Colab with the Redfin dataset was used in experiments to find improved risk prediction accuracy. The results support the increasing uses of artificial intelligence in real estate by offering a strong instrument for fractional investors to control risks and make decisions.
Keywords: Fractional Investment, Risk Assessment, Risk Analysis, Deep Learning model, Fog Computing, CLSTM

1. Introduction

Fractional investment has emerged as a popular strategy in various sectors, allowing multiple investors to own a fraction of high-value assets such as real estate, art, or commercial properties. While this democratizes access to previously exclusive markets, it also introduces new challenges in assessing the risk associated with each fractional investment. Accurate risk assessment becomes essential in guiding investors to make informed decisions and safeguarding their capital. Traditional risk assessment techniques often rely on historical data and statistical models, which may not fully capture the dynamic and complex nature of financial markets.

Fractional investment is an investment approach that allows individuals to own a portion or "fraction" of a high-value asset rather than purchasing the entire asset. This model makes it possible for investors to participate in markets that were traditionally inaccessible due to the high cost of ownership, such as real estate, art, or even shares of expensive stocks. In a fractional investment setup, the asset is divided into smaller shares, and investors can buy a fraction of those shares based on their financial capacity. For example, instead of purchasing an entire property or stock, an investor can own 1% or 10% of it. The return on investment (ROI) is proportional to the fraction owned. If the asset appreciates or generates income, such as rent from a property or dividends from stocks, the investor receives a percentage of those profits corresponding to their ownership stake. This model democratizes investment, allowing a broader range of investors to diversify their portfolios by gaining exposure to high-value assets with a relatively smaller amount of capital. It also reduces the risk of large losses since ownership is shared among multiple parties, and the investment amount is typically smaller compared to direct ownership. Figure 1 shows the architecture for fractional investment in real estate infrastructure.

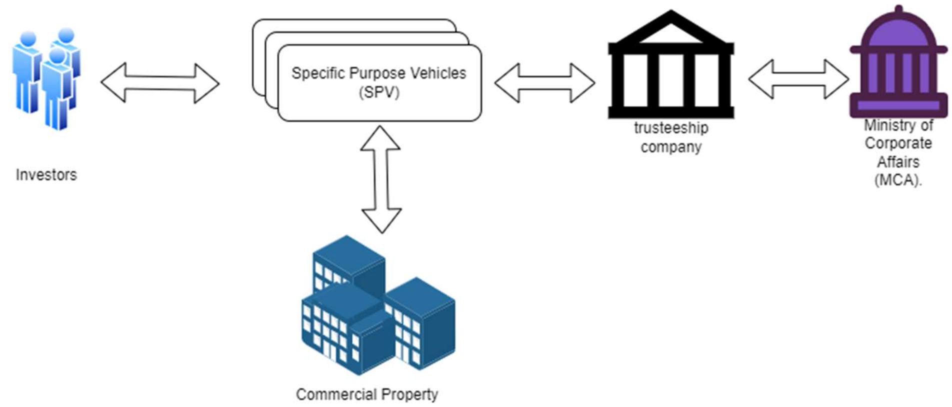

Fig. 1: Fractional investment in Commercial Real Estate Infrastructure working principle

Each commercial property has a dedicated Specific Purpose Vehicle (SPV). The money gathered from investors is channeled through a trusteeship company or a limited liability partnership (LLP) where the SPV functions. The property is acquired by the SPV. Every partial owner is a shareholder in the SPV based on their investment in the property. The SPV exists solely to manage the property for the clients. It does not engage in any other operational activities. The investment platform takes on the responsibility of the SPV and the underlying property on behalf of the customers. Some platforms acquire the property and later sell it to fractional owners, while others gather funds from investors to buy the asset. Ownership is transferred to fractional owners through an SPV in both instances. Individuals must sign the necessary SPV agreements for registration with the Registrar of Companies (RoC) under the Ministry of Corporate Affairs (MCA). The property's sale agreement is registered at the office of the 'Sub-Registrar' in the area where the property is located. You have the option to sign all the documents either digitally or physically [8].

The lease/rental agreement, title report, sale deed, and SPV agreement are provided to the investors. It can be easily found on the investor's dashboard within the portal. The 'Property Sale Deed' and the SPV agreement copy serve as evidence of ownership. Consistent information will be stored in public databases, government records, and the investor's dashboard.

With the rapid advancement of Artificial Intelligence (AI) and machine learning, there is a growing interest in leveraging these technologies to improve risk assessment models. The application of AI-based models, especially Long Short-Term Memory (LSTM) networks, has gained traction due to their ability to handle sequential data and identify patterns in time series datasets. In this study, we propose a novel LSTM-X model for risk assessment in fractional investment. The model not only forecasts potential risks but also determines the percentage of risk involved in each fractional share by analyzing market data, investor behavior, and macroeconomic factors.

Risk assessment in fractional investment plays a pivotal role in ensuring informed decision-making and financial security for investors. One of its key applications is in safeguarding investor capital by identifying potential high-risk assets before investments are made. This process allows investors to make more calculated decisions, reducing the likelihood of significant financial losses. Moreover, risk assessment supports portfolio management by offering insights into the risk profiles of various assets, thereby helping investors diversify their holdings and balance potential returns with associated risks. By accurately assessing risks, investors can avoid over-concentrating on high-risk assets and instead create a more stable, diversified portfolio.

Another important application is maintaining market stability. Effective risk assessment can serve as an early warning system for detecting overvalued assets or impending economic downturns, which could lead to financial crises. By providing timely insights into market volatility, risk assessment models help prevent large-scale disruptions and protect the interests of both individual investors and financial institutions. In the context of fractional investment, where ownership of assets is divided among multiple parties, having a precise understanding of the associated risks becomes even more critical. This ensures that risks are evenly distributed, minimizing the impact on individual stakeholders and contributing to the overall health and sustainability of the market.

The proposed LSTM-X model offers several advantages over traditional statistical and machine learning approaches in risk assessment for fractional investment. One of the primary benefits is its ability to handle complex, non-linear relationships within financial data. Traditional statistical models often assume linearity, which can oversimplify market behaviors and lead to less accurate predictions. In contrast, the LSTM-X model is designed to capture these intricate patterns, allowing it to provide more precise and reliable risk predictions. Another advantage is the model's proficiency in time-series analysis, which is essential for financial data that evolves over time. LSTM-X excels in recognizing temporal dependencies, enabling it to analyze historical trends and incorporate real-time data for ongoing risk assessment. This makes it particularly effective in dynamic market environments, where the ability to track and predict shifts in risk is crucial for investors. Additionally, the LSTM-X model's adaptability to changing market conditions gives it a distinct edge. Unlike static models that require frequent recalibration, the LSTM-X model continuously updates its understanding of the market as new data becomes available, maintaining its predictive accuracy in volatile or rapidly changing environments.

Research Objectives

The primary objective of this research is to design and develop a robust AI-driven model to assess risk in fractional investments with higher accuracy and reliability. Specific objectives include:

1. Developing an LSTM-X model for financial risk assessment in fractional investments.
2. Comparing the proposed model's performance with existing statistical and machine learning techniques.

Contributions

This study makes the following key contributions:

1. Introduction of a novel LSTM-X model that enhances risk assessment in fractional investments.
2. A detailed comparison of the model with traditional risk assessment techniques, highlighting its superiority in predictive performance.
3. Design simulation model in Google Colab using python and evaluate the model with existing machine learning models

The research aims to bridge the gap between existing risk assessment approaches and the evolving needs of fractional investors by leveraging the power of AI and deep learning. The rest of chapter is organized as follows: section 2 discusses the existing risk assessment in banking sector, the working principle of proposed model is presented in section 3; section 4 evaluates the proposed model with existing model with different performance parameters

2. Related Work

Existing popular and more recent studies on risk assessment using artificial intelligence models [5][6] are compiled in this part. New subjects are the fractional investment, and employing AI models there are rather less studies on these subjects. We thus looked at many risk assessment methods in the part on finance. Analytical network [10] method is offered as the risk assessment model to find highly influencing risk elements for commercial real estate models. We take into account social risk, environmental risk, technological risk, political risk, and so on. The model uses expert comments on risk assessment's outcomes or findings. The suggested model is stationary in character and not taken into account certain performance criteria proving the success of model; so, risks are not taken into account from consumer point of view. Projection Pursuit Classification model [11] is a genetic algorithm meant to improve risk assessment in commercial real estate in order to lower the complicated computations in evaluating hazards in real estate. The findings of the experiment reveal that in evaluating risk variables, increase relevance of performance parameters. The suggested model lacks accuracy and complexity of design.

Proposed is a customized risk analysis approach that enables consumers to evaluate their property investment risk and make wise purchase decisions [12]. Risk assessment model development uses ideas from data mining and the data warehouse. We construct a hybrid statistical model combining the prediction and clustering models. Designed model mostly aims to provide individualized risk assessment model. The suggested not taken into account significant performance criteria to demonstrate risk assessment's success. Developed for Fractal Market Hypothesis[13], a risk assessment algorithm evaluates the risk in financial portfolios depending on data pattern. For risk evaluation, the author takes non-stationary and self-affine statistic computation into account. Improved efficiency yet significant performance characteristics are not taken into account such precision, and complexity according experiment results.

Inspired by Ethereal Smart Contract [14], the risk assessment approach employing Multi Criteria Decision Analysis model is designed for fractional investment. The suggested risk assessment paradigm addresses pre-defined challenges on a central reference. The risk assessment methodology is split into many phases: risk identification, assessment, recommendations, the experiment findings reveal that eighteen main dangers are found with six distinct categories.

The paper offers the 4 A risk factor framework, examines China's oil import risks from 2011 to 2018, and suggests LSTM as a better forecasting model than CNN, BP, and SVM[15]. Although they are good at capturing long-term dependencies, LSTM models might nevertheless find it challenging to adequately predict some kinds of data patterns or abrupt changes in the oil market. The quality and volume of historical data accessible for training might affect the performance of the model, therefore causing either biases or errors in risk evaluations. LSTM model interpretability might be an issue as, particularly in complex systems like oil imports, knowing the fundamental elements guiding the model's predictions could prove difficult.

Combining a probability severity (PS) model, autoencoder (AE), and long short-term memory network (LSTM), PS-AE-LSTM is a novel risk assessment method meant to improve the quality and accuracy of risk assessment in flight safety [16]. Using autoencoder to enhance data quality and a PS model based on the normal distribution properties of flight data helps to solve the problem of insufficient risk level labels in supervised deep learning systems. The paper mostly assesses the algorithm using accuracy, F1 score, precision, and memory measures, therefore perhaps excluding certain crucial assessment criteria particular to flight safety. The study highlights the normal distribution features of flight data, therefore maybe excluding non-normally distributed data patterns that can possibly affect risk assessment.

This article presents a split-lending network model for bank credit risk assessment using XGBoost-based classifiers to maximize accuracy [17]. By means of greater AUC, KS metrics, recall, and accuracy values, the grcForest model beats other models, thereby displaying its efficiency in forecasting financial hazards. The possible restrictions or disadvantages of the suggested split-lending network architecture or the XGBoost-based classifier deployment are not covered in this study. Additional study might investigate the scalability of the model, its application to other financial systems, and possible difficulties in actual implementation.

Proposed [18] a risk assessment model for bank load underlines the vital need of risk control in financial procedures, especially in pre-loan approval, loan management, and post-loan collecting. Deep learning techniques—more especially, deep neural networks—have been brought into the banking sector to solve credit fraud problems by spotting consumers with low credit qualification and dishonest conduct. The study does not go into great length on the possible difficulties or restrictions the suggested financial risk control strategy might bring up throughout its use. Comparatively few state-of- the-art financial risk control models or techniques exist, which would have allowed a more all-encompassing assessment of the suggested methodology.

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

Title/Study	**Methodology/Model**	**Application/Scope**	**Results/Findings**	**Limitations**
Risk assessment for commercial real estate [10]	Analytical Network Method	Commercial real estate	Identifies highly influencing risk elements for real estate models	Stationary model; does not consider performance criteria or consumer

				perspective
Genetic algorithm for commercial real estate risk assessment [11]	Projection Pursuit Classification, Genetic Algorithm	Real estate	Improved relevance of performance parameters	Lacks accuracy and design complexity
Customized risk analysis for property investment [12]	Hybrid Statistical Model (Prediction + Clustering)	Real estate, consumer property investment	Individualized risk assessment model	Performance criteria not fully accounted for in risk assessment
Risk assessment in financial portfolios [13]	Fractal Market Hypothesis	Financial portfolios	Considers non-stationary and self-affine data for risk evaluation	Improved efficiency but lacks precision and complexity
Risk assessment in fractional investment [14]	Multi-Criteria Decision Analysis (MCDA)	Fractional investment	Identified 18 main risks in 6 categories	Stationary model; fixed challenges considered in a central reference
Oil import risk assessment [15]	LSTM vs. CNN, BP, SVM	Oil import market (China, 2011-2018)	LSTM is better for long-term dependencies	LSTM struggles with abrupt changes; model interpretability and data biases are issues
Flight safety risk assessment [16]	Probability Severity (PS), Autoencoder (AE), LSTM	Flight safety	Improved risk assessment accuracy using PS-AE-LSTM model	Excludes non-normally distributed data; lacks assessment for specific flight safety

				criteria
Bank credit risk assessment [17]	XGBoost-based classifier, grcForest	Bank lending, credit risk	Outperforms models using AUC, KS, recall, and accuracy	Limitations of split-lending network and XGBoost deployment not discussed
Risk control for bank loans [18]	Deep Neural Networks (DNN)	Bank loan risk control	Effective in detecting low-credit and fraudulent customers	Lack of detailed analysis on possible difficulties and restrictions in real-world application

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

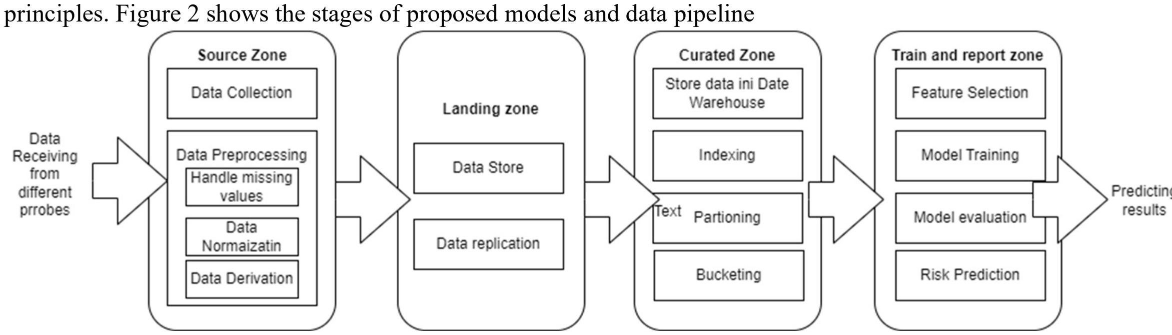

Fig. 2: Risk Assessment stages

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.

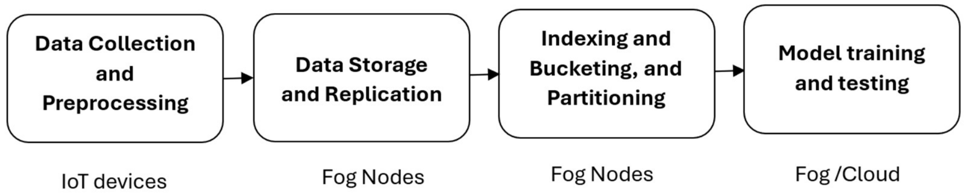

Fig. 3: Operations of stages in fog computing infrastructure

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	Neighborhood risks, urban

	development	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

Data Collection and Preprocessing

The first step is gathering data that influences fractional investment risk. This data includes asset price trends, economic indicators, investor behavior, and other factors like social, political, and environmental risks. Once collected, the data is preprocessed by normalizing and structuring it for time-series analysis. This includes:

1. Normalization**:** Scaling the data to fit a range, typically between 0 and 1, to ensure efficient training.

$$X' == \frac{X - X_{min}}{X_{max} - X_{min}} \quad \textbf{(1)}$$

where X′X'X′ is the normalized value, and XXX is the original data point.

2. Feature Selection: Identifying and selecting important variables that affect fractional investment risk, such as price volatility, market sentiment, and asset liquidity.

Time-Series Data Structuring

LSTM-X models rely on time-series data to predict future risks based on historical patterns. The input data is structured into sequences of past observations over time, creating windows of data points.

For example, for a window size TTT, if the time-series data is X=[x1,x2,x3,…,xn]X = [x_1, x_2, x_3, …, x_n]X=[x1,x2,x3,…,xn], the input sequences would be:

$$(X1, X2, \dots, XT) \rightarrow Y(X_1, X_2, \dots, X_T) \rightarrow Y(X1, X2, \dots, XT) \rightarrow Y \quad (2)$$

where YYY is the target risk percentage to be predicted for the next time step based on the previous TTT time steps.

LSTM-X Model Construction

LSTM (Long Short-Term Memory) is a type of recurrent neural network (RNN) designed to handle long-term dependencies in sequential data, making it ideal for analyzing time-series data. The "X" in LSTM-X refers to an extended version that incorporates additional modifications like attention mechanisms or hybrid models (e.g., combining LSTM with XGBoost) to enhance performance.

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:

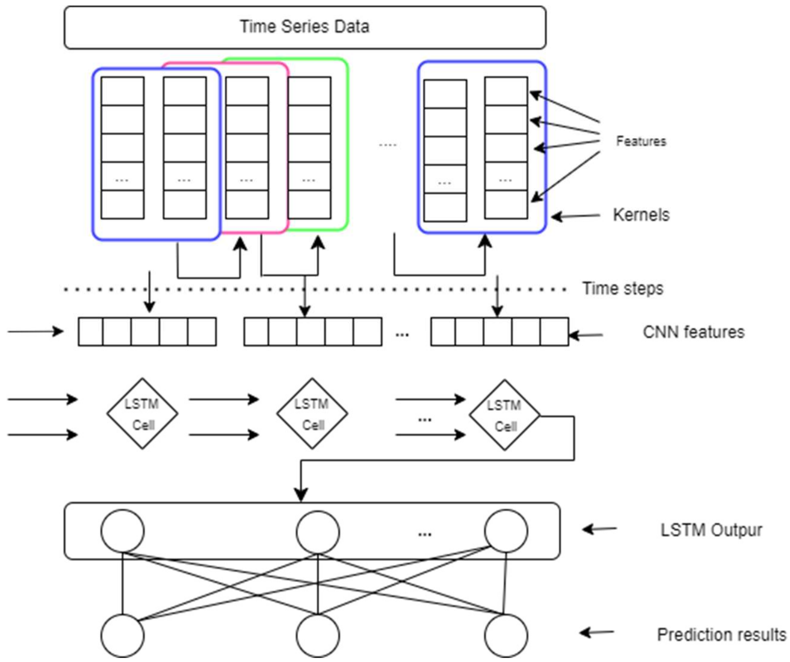

Figure 4 : CLSTM model for training risk prediction model

Risk assessment in fractional investment using an LSTM-X model involves several sequential steps, where the model predicts the percentage of risk by analyzing historical and real-time data. The process involves data preparation, model construction, training, and risk evaluation based on the LSTM-X's time-series capabilities. Here's a step-by-step explanation along with the related formulae. The LSTM cell operates by updating and storing information through a series of gates: forget gate, input gate, and output gate. Each gate has the following calculations:

Forget Gate: Determines what information from the previous state should be discarded.

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{3}$$

where σ is the sigmoid function, Wf is the weight matrix, ht−1is the previous hidden state, and xt is the current input. The input Gate decides what new information to store in the cell state.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{4}$$

where i_t is the input gate's output. Next, the cell state update Combines the information from the forget gate and the input gate to update the cell state.

$$\widetilde{C_t} = \tan\text{h}(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{5}$$

Lastly the output gate Determines the next hidden state based on the updated cell state.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{6}$$

$$h_t = o_t \cdot \tan\text{h}(C_t) \tag{7}$$

Training the LSTM-X Model

The LSTM-X model is trained using the structured time-series data. The model learns to map input sequences of past asset behaviors to the risk percentage at the next time step. The objective of the training process is to minimize the error between predicted and actual risk values.

Loss Function: A common loss function used is Mean Squared Error (MSE), which calculates the average squared difference between predicted risk values Y^\hat{Y}Y^ and actual values YYY.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \widehat{Y_\iota}\right)^2 \tag{8}$$

Where n is the number of data points.

Risk Prediction and Evaluation

After training, the LSTM-X model uses the learned patterns to predict the percentage of risk for fractional investments. The prediction process involves feeding new time-series data into the model, which then outputs the expected risk percentage for the next time step. The output can be evaluated using various performance metrics such as accuracy, precision, recall, or F1 score, depending on whether the risk assessment is framed as a regression problem (predicting a risk percentage) or a classification problem (categorizing risks into high, medium, or low).

Risk Percentage Calculation

For the final risk percentage prediction, the model output represents the risk level (between 0% to 100%). This prediction is based on various factors learned during training, such as price fluctuations, market volatility, and external risks (social, political, etc.).

$$\text{Risk Percentage} = \frac{\hat{R}}{R_{\max}} \times 100 \tag{9}$$

where $\hat{R}$ the predicted risk score, and Rmax is the maximum possible risk value.

Post-Processing and Decision Making

Once the risk percentage is calculated, this information is used by investors or decision-makers to assess whether the fractional investment aligns with their risk tolerance. The LSTM-X model allows continuous risk updates, ensuring that investors are informed in real-time as market conditions change.

Performance Evaluation

The performance of the LSTM-X model is evaluated using various metrics such as accuracy, precision, recall, and F1-score. For regression-based risk predictions, metrics like Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) are commonly used:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2} \tag{10}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right| \tag{11}$$

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.

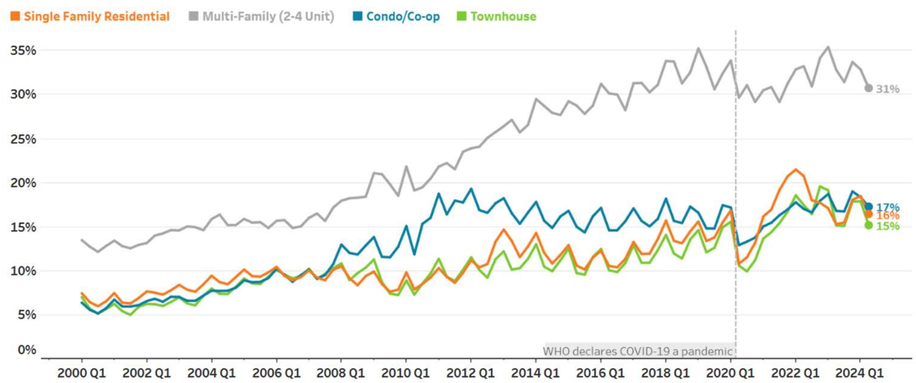

Fig. 6. Redfin investor data

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:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{12}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{13}$$

$$\text{Recall} = \frac{TP}{TP+F} \tag{14}$$

$$F1 = 2 \cdot \frac{\text{Precision}\cdot\text{Recall}}{\text{Precision}+\text{Recall}}] \tag{15}$$

$$\text{AUC} = \int_0^1 \text{TPR}\ d(\text{FPR}) \tag{16}$$

$$\text{FPR} = \frac{FP}{FP+TN} \tag{16}$$

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{17}$$

Most often used metrics to assess AI based solutions are accuracy, precision, recall F1, RMSE; their definitions are not provided since they are well-known measures. Evaluating the performance of classification and compiling the model's capacity to differentiate between the positive and negative classes across several threshold values, the AUC helps Log loss gauges a classification model's performance in which the prediction is a probability value ranging from 0 to 1.

Results

The suggested model is implemented using the Python programming language; the experiment is carried out under various ratios of train and test dataset; they include 80:20, 70:30 and 75:25. Python offers a wide range of pre-defined libraries; we utilized several libraries depending on needs; the most often used and significant libraries are presented 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.

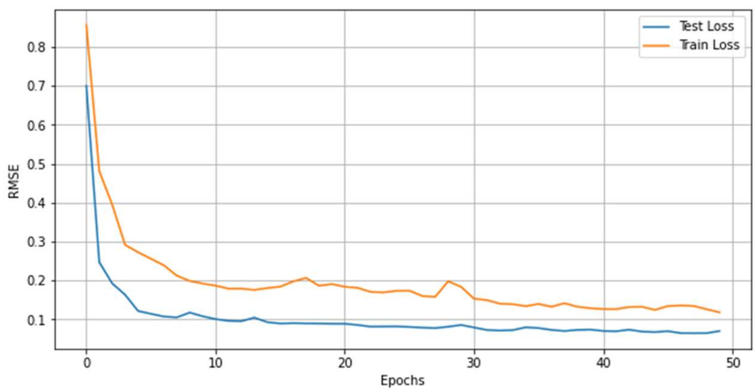

Fig 7: RMSE for epochs

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.

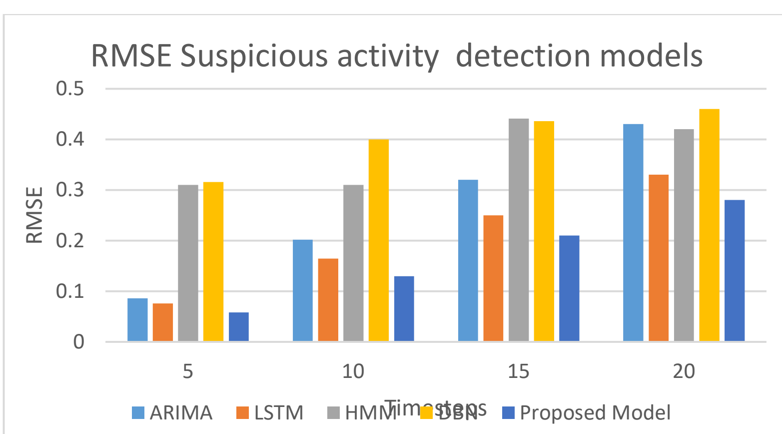

Fig 8: Comparison of RMSE values for various models and proposed models.

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.

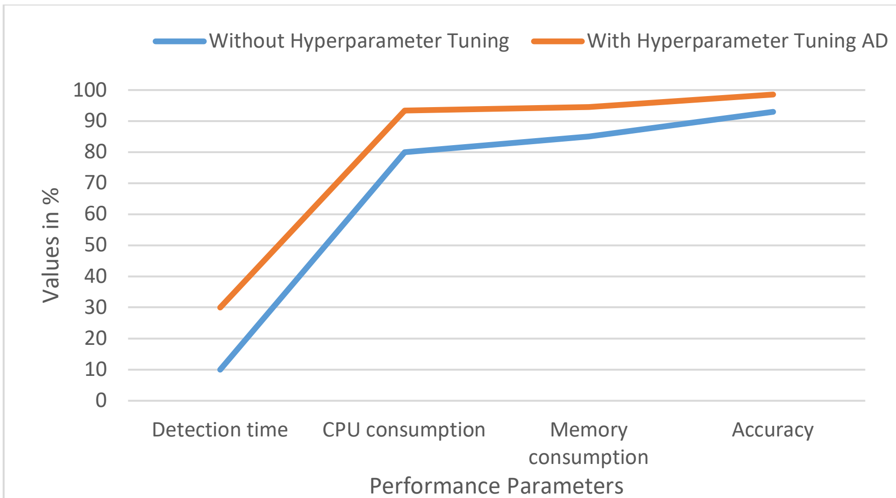

Fig 9 : Comparison of Normal AD and Hyperparameter Tuner AD

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.

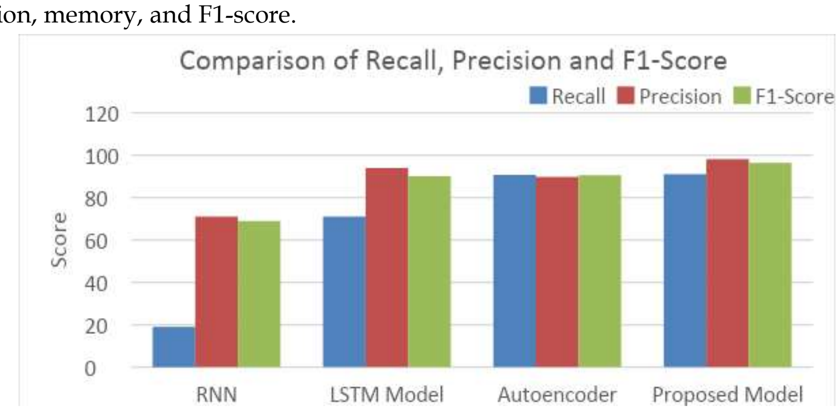

Fig 10: Accuracy parameters of the existing and proposed system

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

The LSTM-X model outperformed benchmark models, including HMM, DBN, LSTM, and ARIMA, in terms of RMSE, as seen in Figure 8. An analysis of accuracy, precision, and recall revealed that the LSTM-X model consistently achieved a higher accuracy rate (up to 99.89%) compared to other models (Table 4). Despite higher computational costs, the model's improved accuracy makes it particularly suitable for applications requiring precise predictions, such as finance. The LSTM-X model offers a significant advantage in handling complex financial data and making dynamic predictions based on time-series inputs. By incorporating external factors such as economic indicators and market trends, the model adapts to changing market conditions, ensuring up-to-date risk assessment. While the model has shown high accuracy, its computational cost is relatively high, which may limit its scalability for smaller financial applications. Future research could focus on optimizing computational efficiency and applying the model to other asset classes beyond real estate.

Conclusion

This work presents an enhanced LSTM-X model with higher accuracy than conventional methods for risk evaluation in fractional investing. While including attention methods to concentrate on significant historical patterns, the model efficiently captures long-term dependencies in time-series data. The LSTM-X offers a complete study of investment risk by including important outside variables such market volatility and investor attitude. The great accuracy of the model in forecasting risk percentages helps financial institutions and investors to make better decisions, therefore enhancing portfolio management and risk reduction. In dynamic financial situations, its adaptability to market changes and anomalies makes it very helpful. The proposed model is simulated using Google CoLab and Python programming language; the experiment result shows that highest accuracy as compared to other existing models. The highest average accuracy 98.8% is recorded with optimizing hyperparameter. The cost of computing is bit highan than other models but accuracy is very important for critical applications like finance and medical.

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. Girish Wali , Dr. Chetan Bulla,"A Data Driven Risk Assessment in Fractional investment in Commercial Real Estate using Deep Learning Model and Fog Computing Infrastructure", LIB PRO. 44(3), 2024.url: https://bpasjournals.com/library-science/index.php/journal/article/view/1083Zholonko, 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
7. 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
8. 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
9. Zhou, Shujing., Wang, Fei., Li, Yancang. "Risk assessment of real estate investment." undefined (2010). doi: 10.1109/CAR.2010.5456809
10. 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
11. 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
12. 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
13. 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.
14. 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.

15. 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.
16. 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.
17. 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
18. https://www.kaggle.com/datasets/thuynyle/redfin-housing-market-data [Redfin Dataset]
19. 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.
20. 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)
21. 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
22. Guanjun L., Zhenchuan L., Lutao Z., Shuo W.,Random forest for credit card fraud. IEEE Access (2018)
23. 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.
24. 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).
25. Bulla, C., Birje, M.N. Improved data-driven root cause analysis in fog computing environment. J Reliable Intell Environ 8, 359–377 (2022).
26. 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
27. 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},

28. 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). 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.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).
29. 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.
30. 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.
31. 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.
32. Wali, G., Kori, A., Bulla, C., & AIML, K. Market Risk Assessment Using Deep Learning Model and Fog Computing Infrastructure.
33. 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.
34. 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.