A Study on Predicting House Prices With Deep Learning: An Evaluation of Advanced Neural Network Models

E.Devi Swetha[1], Dr.R.Porkodi[2]
[1]Dept of computer science
[2]Professor, Dept of computer science
[1, 2] Bharathiar University Coimbatore-641046

Abstract- *Everyone dreams of owning a home that aligns with their lifestyle and meets their unique needs, including essential amenities. Factors like area, location, and views play a major role in determining the price of a house. However, predicting house prices can be challenging due to their frequent fluctuations and occasional overpricing. This volatility creates challenges for prospective buyers aiming to make informed decisions and for real estate agencies seeking to invest wisely in properties. As a result, accurately estimating house prices becomes essential to help both buyers and investors navigate the housing market more confidently. To address these challenges, this study explores the use of deep learning models for predicting house prices with the help of a comprehensive dataset of real estate listings from Kaggle. This dataset contains over 10,071 entries and 25 features, including price, area, location, number of bedrooms, and various amenities. Data preprocessing involved one-hot encoding, feature scaling, and outlier removal using the Interquartile Range (IQR) method to prepare the data for analysis. Six different deep learning models were developed and trained for prediction: Artificial Neural Network (ANN), Feedforward Neural Network (FNN), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Gated Recurrent Unit (GRU), and Long Short-Term Memory (LSTM) networks. To evaluate model accuracy, performance metrics such as R^2, Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) were used. Among these models, the LSTM model achieved the highest accuracy, with an R^2 score of 0.9773, followed closely by the GRU and FNN models. These results indicate that LSTM networks are particularly effective in identifying complex patterns in house price data.*

Keywords- House Price Prediction, Deep Learning Models, Real Estate Market, LSTM Networks, Gated Recurrent Units (GRU), Convolutional Neural Networks (CNN), Artificial Neural Networks (ANN)

I. INTRODUCTION

Accurate house price prediction plays a vital role in the real estate market, benefiting homeowners, buyers, investors, and real estate professionals. It allows stakeholders to make informed decisions regarding property investments, pricing strategies, and financial planning. Property prices are impacted by various factors, such as location, size, condition, available amenities, market trends, and broader economic conditions. Historically, traditional models such as linear regression and decision trees were employed to predict house prices. However, these models often struggle with handling the complexity of real estate data, where relationships between features are nonlinear, multifaceted, and difficult to quantify. In recent years, the application of machine learning (ML) and deep learning (DL) techniques has shown significant promise in improving the accuracy and robustness of house price prediction models. Deep learning, particularly, has proven to be effective at capturing complex patterns in large datasets, making it highly suitable for real estate applications where the data can be high-dimensional and intricate. Models like Artificial Neural Networks (ANN), Feedforward Neural Networks (FNN), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Gated Recurrent Units (GRU), and Long Short-Term Memory (LSTM) networks have emerged as powerful tools capable of modeling complex relationships within the data and delivering precise price predictions. The transition from traditional statistical methods to deep learning models represents a significant shift in the way house price predictions are made, as these models can adapt and improve with exposure to larger datasets and more complex features. While traditional regression models are limited by their assumptions of linearity and independence of variables, deep learning models are capable of handling intricate, high-dimensional datasets without the need for explicit feature engineering. This paper explores and compares several deep learning techniques applied to house price prediction using a real estate dataset, focusing on the performance of ANN, FNN, CNN, RNN, GRU, and LSTM models. To evaluate the models based on key performance metrics such as R-squared (R^2), Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). Through this study, we aim to showcase the effectiveness of deep learning in real estate price prediction and highlight the models' potential for improving decision-making within the industry.

II. LITERATURE REVIEW

P.Y. Wang et al. [1] This Research has established that various factors, such as property characteristics (age, area, and materials) and environmental features (proximity to schools, parks, and transit), significantly influence prices. Early models primarily used structured data, which limited prediction accuracy by omitting contextual features. Recent advancements introduced deep learning techniques that leverage heterogeneous data, incorporating unstructured inputs like satellite imagery to capture environmental context. Attention mechanisms and spatial transformer networks (STNs) are now being applied to improve feature extraction, enabling more precise modeling of complex interactions.

Mahdieh Yazdani [2] This article traditionally focused on hedonic pricing models, where the property's attributes—such as structure, neighborhood, and location—serve as core determinants of price (Arvanitidis, 2014). However, traditional methods like hedonic regression can be subjective, often leading to biases, particularly in the United States, where issues of racial bias in appraisals have been reported (Bloomberg CityLab). To address these limitations, machine learning (ML) and deep learning (DL) models, including artificial neural networks (ANN), random forests (RF), and k-nearest neighbors (KNN), have been increasingly explored as alternatives for real estate valuation. Studies by Selim (2011) and Park and Bae (2015) highlight the improved accuracy of ML techniques over traditional methods in various markets, such as Turkey and South Korea.

Md Hasbul Hasan et.al.[3] This study based on traditional models, like Hedonic Price models, used basic property features (e.g., number of rooms, property size) to estimate values but were limited by reliance on manually engineered features, which required domain expertise. Machine learning models later introduced spatial data to capture neighborhood effects, with recent approaches like the Geo-Spatial Network Embedding (GSNE) using nearby amenities, such as schools and transport links, to improve accuracy. Multi-modal deep learning has recently gained traction, leveraging visual and textual data alongside traditional features. Techniques such as BERT for text embeddings and CLIP for joint text-image embeddings are becoming popular for capturing nuanced aspects like aesthetics and detailed descriptions, offering promising results in improving prediction precision across various datasets.

Fatemeh Mostofi et al. [4] House price prediction has been addressed through various machine learning techniques, including deep neural networks (DNN), which have shown high accuracy in applications with large datasets. However, high dimensionality and skewness of real estate data remain significant challenges. Principal Component Analysis (PCA) is commonly applied for dimensionality reduction in fields like stock forecasting and air quality monitoring, helping streamline models while preserving data variance. In real estate, DNNs combined with PCA have demonstrated strong predictive performance, though few studies specifically address skewness reduction. Recent research highlights the potential of transformations like square root, cube root, and logarithmic methods for normalizing data distributions, which could improve DNN performance by reducing errors and overfitting in skewed datasets.

Fan et al. [5] emphasize the challenges posed by large datasets comprised of different data types, which complicates the analysis. Recent literature has shown that combining supervised ML methods, such as Random Forest, with Natural Language Processing (NLP) techniques can improve prediction accuracy. Furthermore, using deep learning architectures, like Bidirectional Long Short-Term Memory (Bi LSTM), for processing textual data and employing transfer learning for image feature extraction has proven effective in enhancing model performance. Despite the robust methodologies available, the application of ML in predicting housing prices specifically in Brazil remains underexplored, highlighting the need for further research in this domain (Moreira de Aguiar et al., 2014; De Souza, 1999).

Rabia Naz et al. [6] This study categorized methodologies into three main groups: machine learning, deep learning, and hybrid models. Machine learning techniques, such as XG Boost and support vector regression, demonstrate strong predictive capabilities for smaller datasets, while deep learning methods excel in managing larger, more complex data sets. Hybrid models have emerged as particularly effective, combining the strengths of both traditional and modern techniques to enhance prediction accuracy. Notably, a review of SCOPUS-indexed literature revealed a trend towards advanced machine learning applications, particularly in papers published between 2020 and 2021. Furthermore, the integration of diverse data types, such as spatial data and images, is becoming increasingly common, showcasing a shift towards innovative predictive methodologies (Geerts et al., 2021).

Ebubekir Ayan et al. [7] The literature on housing price bubbles highlights various definitions and indicators that signal overvaluation in real estate markets. Stiglitz (1990) argues that a bubble forms when investors perceive prices as unsustainable high, while Lind (2009) associates bubbles with sharp declines following prolonged price increases. Oust and Hrafnkelsson (2017) refine these definitions, identifying “large” and “small” bubbles based on specific percentage increases and subsequent declines over designated timeframes. The dynamic nature of contemporary markets challenges traditional definitions, with prices sometimes maintaining high levels without immediate corrections. Furthermore, the financialization of housing, especially evident in developed nations, has intensified academic scrutiny, particularly following the 2008 mortgage crisis, which revealed the vulnerabilities of housing markets and their significant impact on broader economic stability.

Hyun-Soo Kim et al. [8] The relationship between public rental housing (PRH) and nearby property values has been a contentious topic, often driven by community concerns that PRH negatively affects real estate prices. Previous studies utilizing hedonic pricing and regression analyses have shown mixed results, indicating that the impact of affordable housing can vary significantly based on local context and the characteristics of both the housing and neighborhoods. Traditional methods struggle with complexities such as nonlinearity and spatial dependencies, which can obscure the true effects of PRH. In contrast, machine learning techniques, particularly long short-term memory (LSTM) models, offer a more nuanced approach, capturing temporal dynamics and accommodating diverse data structures. Research has demonstrated the effectiveness of these models in enhancing predictive accuracy in housing price forecasts.

Hussin Ragab et al. [9] The accurate prediction of house prices is crucial for both buyers and sellers, influenced by various factors such as location, amenities, and property characteristics. Traditional regression methods have been widely used for this purpose, yet they often struggle with the complexities of real estate data. Recent advancements in machine learning, particularly hybrid models that combine Gated Recurrent Units (GRU) and Long Short-Term Memory (LSTM) networks, have shown promise in improving prediction accuracy. These models can effectively capture temporal dependencies and complex relationships in housing datasets. Comparative studies indicate that such hybrid approaches outperform conventional techniques, offering enhanced precision in estimating house prices.

F.U. Sosyal Bilimler Dergisi et al. [10] Recent studies have highlighted the significant impact of various economic factors on housing prices, particularly in the wake of the Covid-19 pandemic. The increased demand for larger living spaces, combined with supply chain disruptions and rising costs, has contributed to inflated housing prices in urban areas like Istanbul and Ankara. Machine learning and deep learning models, such as Recurrent Neural Networks (RNN), including Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), have emerged as effective tools for forecasting housing price trends. Comparisons with traditional methods, like Random Forest, have shown that these advanced models often yield more accurate predictions. Understanding these dynamics is crucial for stakeholders in the housing market to navigate future trends and make informed decisions.

III. METHODOLOGY

Dataset

Dataset collected from the Kaggle website. The dataset appears to focus on real estate property listings, containing 10,071 instances and 40 attributes. Key variables include:

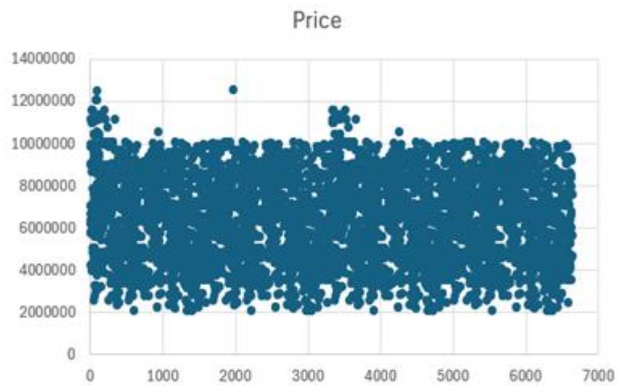
Price: The property price in local currency.
Area: The property's size measured in square feet.
Location: A categorical variable for the property's location (e.g., "Nizam pet," "Hitech City").
Number of Bedrooms: Total bedrooms within the property.
Resale: A binary indicator of whether the property is for resale.
Amenities: Numerous binary columns indicating the presence of amenities, such as: Maintenance Staff, Gymnasium, Swimming Pool, Landscaped Gardens, Jogging Track, and more.

These features suggest that the dataset could support research on property pricing factors, trends in amenity availability, or comparisons across different locations. Fig.1 outlines the key steps in a model workflow, from data collection to model evaluation and results interpretation.

Data Preprocessing

Preprocessing includes steps such as data cleaning, normalization, and handling outliers, which can significantly impact the model's performance and interpretation of results. In particularly Outliers, are extreme values that can skew the dataset and introduce biases in statistical analyses. Detecting and managing these outliers is thus a critical step before moving forward with any data analysis or machine learning model

Before Removing Outliers



After Removing Outliers

The study focused on identifying and removing outliers in the `Price` attribute using the Interquartile Range (IQR) method. The IQR method is a widely used approach that leverages the distribution of the data to detect anomalies. Specifically, it calculates the 1st (Q1) and 3rd (Q3) quartiles of the data and defines the interquartile range (IQR) as the difference between these two quartiles (i.e., IQR = Q3 - Q1). Any values falling outside the range [Q1−1.5×IQR, Q3+1.5×IQR] are flagged as outliers.

Applying this method allowed us to identify and remove unusually high or low values in the `Price` attribute that might otherwise distort the analysis. Initially, our dataset contained 10,072 instances. After applying the IQR-based filtering process to remove outliers from the `Price` attribute, the dataset was reduced to 6,610 instances. This reduction reflects the effective identification of extreme values while preserving the core structure and variability of the data. By filtering these outliers, the dataset's integrity and representativeness were maintained, supporting a more accurate and robust analysis in subsequent stages of the research. Fig.1 and Fig.2 Shows Price Distribution with and without Outliers to indicate that one image shows the data before and after outlier removal.

Model Selection

Model selection is a crucial step in machine learning to achieve accurate predictions. Choosing the right model is essential for obtaining good accuracy. In this study, we apply four different regression models Artificial Neural Networks (ANN), Feedforward Neural Networks (FNN), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Gated Recurrent Units (GRU), and Long Short-Term Memory (LSTM) networks on the housing dataset. To assess model performance, we split the dataset into two subsets: 80% for training and 20% for testing. The models are trained on the training dataset and evaluated on the test dataset to measure their accuracy.

IV. PREDICTION TECHNIQUES FOR HOUSE PRICE PREDICTION USNG DEEP LEARNING MODELS

Artificial Neural Networks (ANN)

An Artificial Neural Network (ANN) is a machine learning model inspired by biological neural networks, used in applications like classification and regression. ANNs consist of an input layer, hidden layers, and an output layer, with each layer containing neurons connected to the next by weights. The process in each neuron involves a weighted sum of inputs, given by:

$$Z = \sum_{i=n}^{n} w_i x_i + b \qquad [1]$$

where are weights, are inputs, and b is the bias. This sum z is then passed through an activation function, f(z), which introduces non-linearity:

$$f(z)= \text{activation}(z) \qquad [2]$$

Each layer's output becomes the input for the next layer, and this forward propagation continues until reaching the output layer, which produces the network’s final output. During training, weights and biases are optimized using backpropagation to minimize prediction error. Unlike CNNs, which specialize in capturing spatial patterns, ANNs are more general-purpose, often applied to structured data where features contribute independently to predictions [2]. ANN model Predicted versus actual prices were visualized in Fig.3 for accuracy comparison.

Feedforward Neural Networks (FNN)

A Feedforward Neural Network (FNN) is a basic type of artificial neural network where connections between nodes do not form cycles, and data flows in one direction—from the input layer through hidden layers to the output layer. FNNs are commonly used for tasks like classification and regression, processing information in a straightforward manner without feedback loops. In each neuron of an FNN, the inputs are multiplied by weights and summed with a bias term. The operation for each neuron can be represented as:

$$Z = \sum_{i=n}^{n} w_i x_i + b \quad (3)$$

where represents weights, are the input values, and b is the bias. This sum, z, is then passed through an activation function, such as ReLU or sigmoid, to introduce non-linearity:

$$f(z)= \text{activation}(z) \quad (4)$$

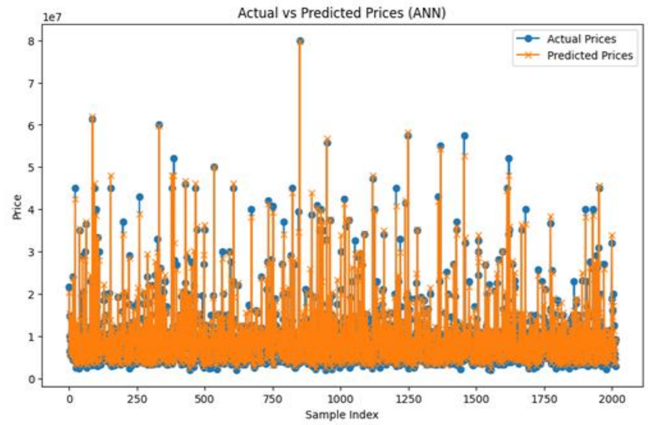
The outputs from each layer serve as inputs to the next, allowing patterns to be learned as data propagates forward. Finally, the output layer provides the network's prediction or classification. During training, weights and biases are optimized through backpropagation, adjusting them to reduce the prediction error. FNNs are versatile for structured data, though they generally lack the spatial pattern-detection capabilities found in Convolutional Neural Networks (CNNs). Finally, actual vs. predicted prices of FNN are visualized in Fig.3.

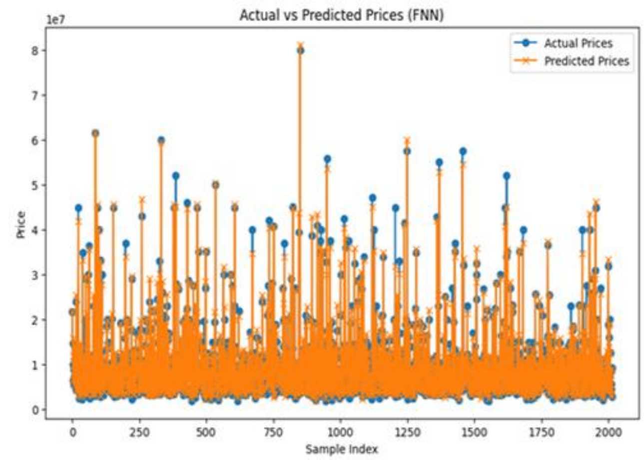
Recurrent Neural Networks (FNN)

A Recurrent Neural Network (RNN) is a type of neural network designed for sequential data, such as time series or language, by utilizing connections that loop back to previous layers, creating a form of memory. This structure enables RNNs to retain information from earlier steps, making them suitable for tasks where context across time is essential. At each time step t, the hidden state is calculated based on the current input and the hidden state from the previous time step as follows:

$$h_t = f(W_h x_t + U_h h_{t-1} + b_h) \quad [5]$$

where: is the hidden state at time t, is the current input, and are weights, and is a bias term. The activation function f (e.g., tanh or Relu) introduces non-linearity. This recurrent connection allows RNNs to recognize temporal patterns, but traditional RNNs can struggle with long sequences due to issues like vanishing gradients. Advanced RNN types, such as Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs), address this by adding mechanisms to retain or forget information over longer sequences. Fig.4 shows a visual comparison of actual vs. predicted prices was plotted.





Comparision of actual and predicted price using ANN and FNN

Convolutional Neural Networks (CNN)

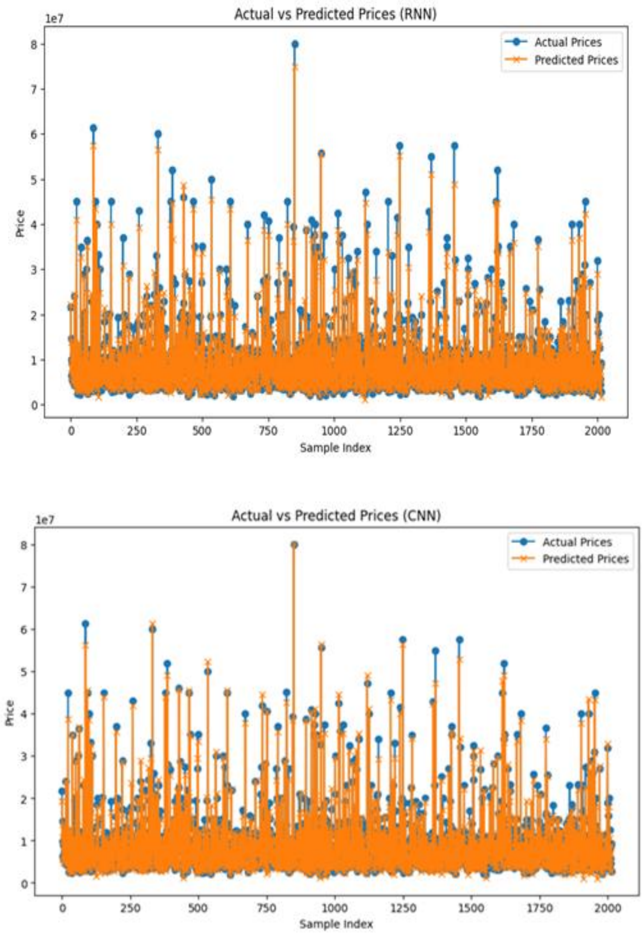
A Convolutional Neural Network (CNN) is an advanced neural network model designed to capture spatial and temporal patterns, especially in image and time series data. CNNs use convolutional layers, pooling layers, and fully connected layers to extract and process important features across varying scales. In each convolutional layer, a filter (or kernel) slides over the input data, performing an element-wise multiplication and summation to produce a feature map. This operation can be described as:

$$f(i,j) = \sum_m \sum_n W_{m,n} . x(i+m, j+n) + b \quad [6]$$

Where, is the output feature at location, W is the filter matrix, are the input values within the filter region, b is the bias. A pooling layer, commonly max pooling, follows to reduce the feature map's dimensions while retaining key features, typically by taking the maximum value in each small region, as in:

$$p(i,j) = max_{pool} f(i+m, j+n) \quad [7]$$

Finally, the feature maps pass through fully connected layers, which generate the network's final output. In time series applications, CNNs can successfully predict values by learning temporal dependencies in the data [13].



Comparision of actual and predicted price using RNN and CNN

Gated Recurrent Units (GRU)

The Gated Recurrent Unit (GRU), developed by Kyunghyun Cho et al [11] and colleagues, is an enhanced version of the Recurrent Neural Network (RNN). GRUs incorporate two types of gates the update gate and the reset gate to control information flow within the network. The update gate manages which information should be retained in memory, while the reset gate determines what information is allowed to exit the memory. Essentially, these gates are vectors that help decide the relevant information to pass to the output.

The update gate is determined as follows:

$$Z_t = \sigma\left(W^{(Z)} x_t + U^{(Z)} h_{t-1}\right) \quad [8]$$

The reset gate rt is determined as:

$$r_t = \sigma\left(W^{(r)} x_t + U^{(r)} h_{t-1}\right) \quad [9]$$

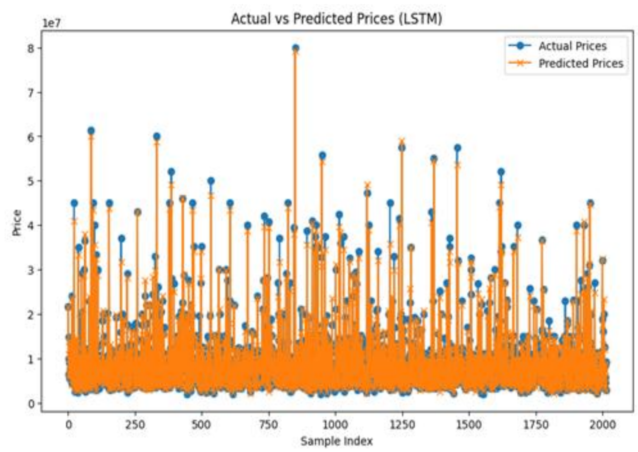
the current memory content is computed as:

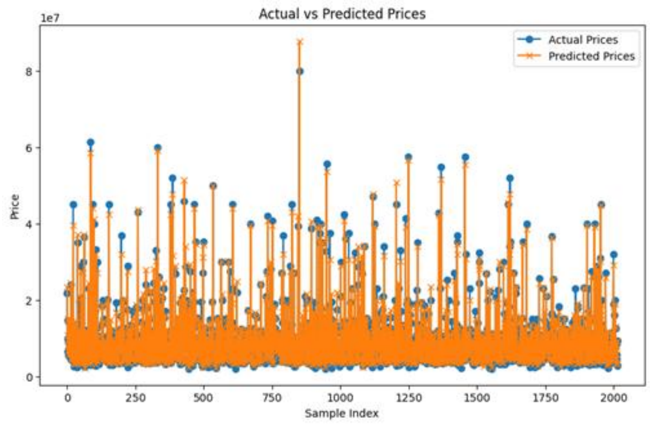
$$h'_t = \tanh(W x_t + r_t \odot U h_{t-1}) \quad [10]$$

Finally, the memory at the current time step h t is obtained as:

$$h_t = Z_t \odot h_{t-1} + (1 - Z_t) \odot h'_{t} \, [11]$$

In these equations x t represents the input vector at the current time step, h t-1 denotes the previous hidden state, W and U are weight matrices, represents the sigmoid activation function and is the Hadamard product (element-wise multiplication). The result of the GRU model shown in Fig. 5 indicates that its predicted prices generally align with the actual prices.





Comparision of actual and predicted price using GRU and LSTM

Long Short - Term Memory (LSTM)

The LSTM model, introduced by Hochreiter and Schmidhuber in 1997 improves upon the standard RNN structure by addressing its limitations in retaining information over long sequences. LSTM units leverage a unique cell state and carefully designed gates to store and manipulate information effectively. Key to this mechanism is the absence of an activation function within the cell state, which prevents gradual compression of values and mitigates gradient vanishing issues during backpropagation through time (Chen et al., 2017) [12]. The forget gate controls the information to discard from the previous cell state, given the current input and previous hidden state, processed by the activation function as shown in Equation 12. Equation 13 computes intermediate values and by applying activation functions to and, using weight parameters like and.

$$i_t = \sigma_g(W_i X_t + U_i h_{t-1} + b_i) \quad [12]$$

$$C_t' = \sigma_c(W_c X_t + U_c h_{t-1} + b_c) \quad [13]$$

$$C_t = f_t x C_{t-1} + i_t C_t' \quad [14]$$

$$o_t = \sigma_g(W_o X_t + U_o h_{t-1} + b_o) \quad [15]$$

$$h_t = o_t \cdot \tanh(C_t) \quad [16]$$

In Equation 14, the updated cell state combines the forget gate output with the prior cell state, and the input gate's output with the candidate cell state. The output vector is derived by applying the activation function to and , as seen in Equation 15. The final hidden state, given in Equation 16, is obtained by multiplying with the tanh-activated cell state

V. RESULT AND DISCUSSION

Previous studies have explored house price prediction using various machine learning algorithms, including Random Forest, Lasso Regression, Linear Regression, Ridge Regression, Decision Tree, and Gradient Boosting. Among these, Random Forest has demonstrated strong predictive performance, achieving an R2 score of 0.91, indicating a high level of accuracy in capturing the variance in housing prices. These models have proven useful in understanding the influence of different variables on property values. The current study builds on this by applying neural network models, with results shown in the Table 1. This Model evaluates the performance of six models ANN, FNN, CNN, RNN, GRU, and LSTM based on four key metrics R2 score, Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). These metrics help compare the effectiveness of each model in predicting the target variable. The R2 score measures how well the model predictions match the actual data, with values closer to 1 indicating a better fit. The LSTM model achieved the highest R2 score at 0.9773, suggesting it provides the most accurate predictions. GRU and FNN closely followed with scores of 0.9757 and 0.9759, respectively. This indicates that LSTM, GRU, and FNN can better capture the underlying data patterns than the other models tested.

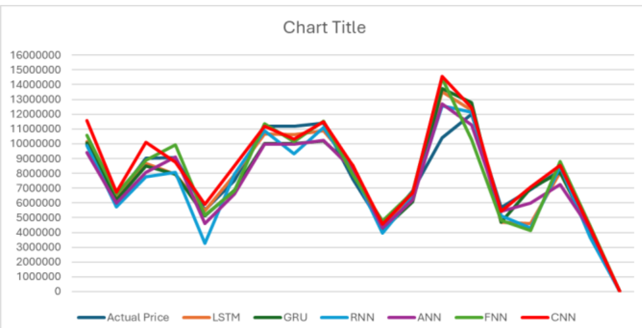
Performance Comparison of Different Neural Network Models

Model	R^2 Score	MAE	MSE	RMSE
ANN	0.9679	2.2377	5.4543	2.3354
FNN	0.9759	1.9440	4.1035	2.0257
CNN	0.9714	2.2572	4.8601	2.2045
RNN	0.9586	2.3086	7.0358	2.6525
GRU	0.9757	1.8573	4.1396	2.0346
LSTM	0.9773	1.8052	3.8565	1.9638

The LSTM model had the lowest MAE of 1.8052, demonstrating its strength in minimizing absolute prediction errors. GRU and FNN also showed low MAE values of 1.8573 and 1.9440, respectively. In contrast, CNN and RNN had higher MAE values (2.2572 and 2.3086), indicating these models are less effective in minimizing errors. Consistent with other metrics, LSTM showed the best performance with the lowest MSE at 3.8565, indicating high accuracy. GRU and FNN followed with MSE values of 4.1396 and 4.1035, respectively. The RNN model displayed the highest MSE (7.0358), showing a lower fit to the data. RMSE values also supported the previous results. LSTM achieved the lowest RMSE at 1.9638, closely followed by FNN (2.0257) and GRU (2.0346). RNN again showed the highest RMSE (2.6525), confirming its comparatively lower prediction accuracy.

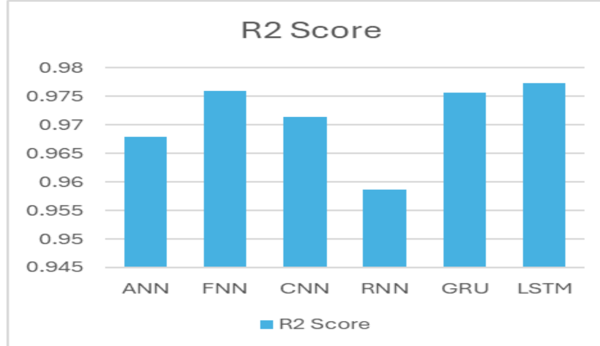
VI. CONCLUSION

In this study multiple deep learning models, including Artificial Neural Networks (ANN), Feedforward Neural Networks (FNN), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Gated Recurrent Units (GRU), and Long Short-Term Memory networks (LSTM) for the task of house price prediction. Each model was evaluated on a dataset containing over 10,000 real estate listings, with preprocessing steps such as one-hot encoding, normalization, and outlier management applied to improve data quality and prediction accuracy.

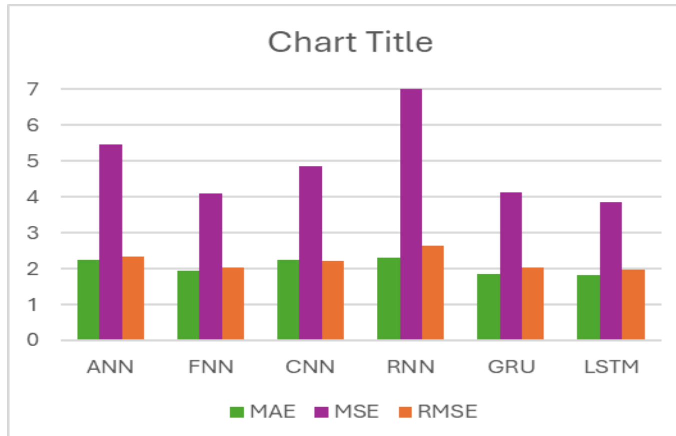


Actual vs Predicted House Prices Using Deep Learning Models

Among the models tested, the LSTM model achieved the highest performance across metrics, with an R^2 score of 0.9773, lowest Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE), indicating its superior ability to capture complex temporal dependencies in the data. The results show that while all models performed reasonably well, LSTM, GRU, and FNN models stood out for their accuracy, likely due to their structural capacity to capture intricate patterns in the data, particularly LSTM and GRU with their ability to manage sequence dependencies. CNN, despite its powerful feature extraction capabilities, showed slightly lower performance, suggesting that convolutional approaches may be more suited for spatial or visual data rather than structured data like tabular datasets in real estate. Future work can explore hybrid models that combine LSTM or GRU with CNN layers could enhance spatial-temporal learning capabilities, potentially improving the capture of complex relationships within the dataset. Further optimization through hyperparameter tuning and ensemble approaches, such as stacking or blending models, could boost the model's robustness and predictive power.



Comparison of R^2 Scores Across Models



Error Metrics (MAE, MSE, RMSE) Comparison

REFERENCES

[1] Wang, P.-Y., Chen, C.-T., Su, J.-W., Wang, T.-Y., Huang, S.-H.: Deep learning model for house price prediction using heterogeneous data analysis along with joint self-attention mechanism. IEEE Access **9**, 55244–55259 (2021).

[2] Yazdani, M. (2021). Machine learning, deep learning, and hedonic methods for real estate price prediction. *arXiv preprint arXiv:2110.07151*.

[3] Hasebul Hasan, Md, et al. "A Multi-Modal Deep Learning Based Approach for House Price Prediction." *e-prints* (2024): arXiv-2409.

[4] Mostofi, F., V. Toğan, and H. B. Başağa. 2021. "House price prediction: A data-centric aspect approach on performance of combined principal component analysis with deep neural network model." *J. Constr. Eng.* 4 (2): 106 116.

[5] Afonso, Bruno, et al. "Housing prices prediction with a deep learning and random forest ensemble." *Anais do XVI Encontro Nacional de Inteligência Artificial e Computacional*. SBC, 2019.

[6] Naz, Rabia, Bushra Jamil, and Humaira Ijaz. "Machine Learning, Deep Learning, and Hybrid Approaches in Real Estate Price Prediction: A Comprehensive Systematic

Literature Review." *Proceedings of the Pakistan Academy of Sciences: A. Physical and Computational Sciences* 61, no. 2 (2024): 129-144.

[7] Ayan, E. and Eken, S., 2021. Detection of price bubbles in Istanbul housing market using LSTM autoencoders: a district-based approach. *Soft Computing*, *25*(12), pp.7957-7973.

[8] Kim, H., Kwon, Y., & Choi, Y. (2020). Assessing the impact of public rental housing on the housing prices in proximity: Based on the regional and local level of price prediction models using long short-term memory (LSTM). *Sustainability*, *12*(18), 7520.

[9] Ragb, Hussin, et al. "Hybrid GRU-LSTM Recurrent Neural Network-Based Model for Real Estate Price Prediction." *Authorea Preprints* (2023).

[10] Şimşek, A. İ. (2024). USE OF MACHINE LEARNING AND DEEP LEARNING METHODS IN HOUSING PRICE INDEX ESTIMATION: AN ANALYSIS ON ANKARA AND ISTANBUL. *Fırat Universities Sosyal Bilimler Dergisi*, *34*(3), 1345-1353.

[11] Cho, K., Van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014). On the properties of neural machine translation: Encoder-decoder approaches. arXiv preprint arXiv:1409.1259.

[12] X Chen, L. Wei, J. Xu. House Price Prediction Using LSTM. Computing Research Repository (Corr). Abs/1709.08432, September 2017.

[13] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278–2324. doi:10.1109/5.726791.