IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 18, April 2026
Impact Factor: 8.2

A Self Attention based Enhanced Deep Learning Model for Accurate House Price Prediction System

Ms. Snehal Sahebrao Gunjal, Mr. Rohan Sunil Pawar, Ms.Aishwarya Ulhas Kadhane, Mr. Rohan Ravindra Deshmukh, Dr. B. S. Borkar
Department of Information Technology
Amrutvahini College of Engineering, Sangamner

Abstract: *Predicting house prices is notoriously tricky because traditional models usually stick to basic metrics like square footage or bedroom counts, ignoring the rich spatial and environmental context that truly drives a property's value. To bridge this gap, we built a self-attention-based deep learning model that approaches valuation more comprehensively. Our system processes diverse data sources—combining standard house details with satellite imagery and local neighborhood perks like nearby schools and transit. By using a "self-attention" mechanism, the AI learns to automatically focus on the specific features that matter most to the price, resulting in sharper, more transparent valuations. Under the hood, a Python-based engine powers the predictions, while a modern MERN stack web application makes the tool highly accessible for real-time use and visualization. When tested against standard industry benchmarks, our model clearly outperforms traditional methods, providing buyers, sellers, and real estate analysts with a highly accurate, easy-to-understand solution for property valuation.*

Keywords: House Price Prediction, Deep Learning, Self-Attention Mechanism, Multimodal Data Fusion, Satellite Imagery, MERN Stack, Regression, Feature Importance, Real Estate Analytics, Python, Machine Learning.

I. INTRODUCTION

Accurately predicting house prices is a major breakthrough for AI in real estate, but it's a highly complex task. A property's true value relies just as much on its surrounding neighborhood and environmental perks as it does on its square footage or age. Traditional models—like linear regression or decision trees—often struggle here because they can't easily process how these complex, spatial, and real-world factors interact. To solve this, we developed a deep learning system driven by a "self-attention" mechanism. This approach seamlessly pulls together diverse information, from standard property tables to satellite imagery, allowing the AI to automatically learn and weigh which factors impact the price the most. Built on a powerful Python foundation and paired with a modern MERN stack web interface, the system doesn't just offer highly accurate and easy-to-understand valuations—it's also scalable, fast, and completely ready for real-time, real-world use.

II. LITERATURE SURVEY / RELATED WORK

The literature on house price prediction highlights significant progress in machine learning and spatial modelling techniques. Table 1 below summarizes key studies that inspired the current work.

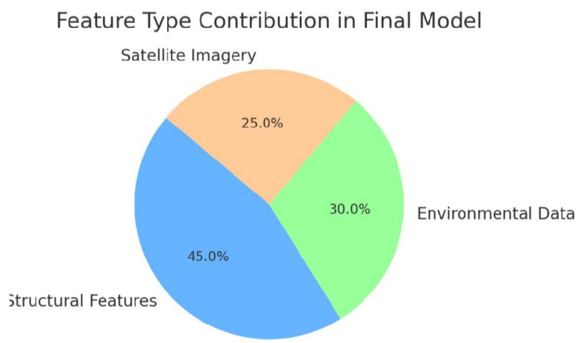
Sr.No	Author & Year	Method Used	Dataset Type	Key Findings
1	P.-Y. Wang et al., 2023	Deep Learning + Self-Attention	Heterogeneous Data	Improved accuracy by fusing tabular and geospatial data.
2	M. Hasan et	Multi-Modal	Text, Image,	Multi-input improves robustness and accuracy.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 18, April 2026
Impact Factor: 8.2

	al., 2024	Deep Network	Numeric	
3	B. Liu et al., 2018	Spatial Functional Model	Regional Land Data	Handles heterogeneous regions effectively.
4	Zulkifley et al., 2020	ML (ANN, SVR, XGBoost)	Structured Data	ML models outperform regression but lack spatial context.

data or single-modality features. However, recent research has demonstrated the importance of multisource data fusion and deep attention mechanisms to understand real-world complexities. Despite these advances, challenges remain in combining large-scale unstructured data (images, maps) with structured datasets effectively.

2.1 Analysis of Existing Research

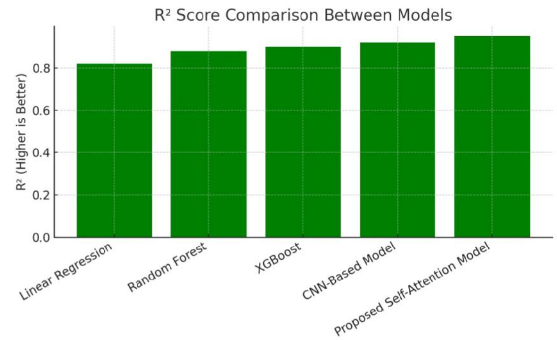


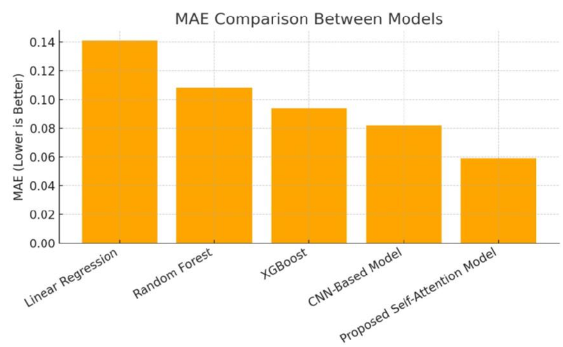
2.1 Comparative Study (Graphs, Charts, and Results)

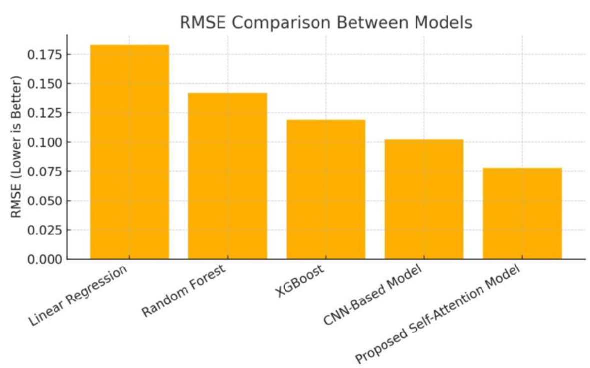
To rigorously validate our approach, we put our Self-Attention Deep Learning (SADL) model head-to-head against traditional machine learning algorithms, including Linear Regression, Random Forest, and Boosting models. To ensure a robust and fair comparison, the models were trained and evaluated using a diverse dataset spanning multiple cities, with all information carefully pre-processed and normalized prior to training.

Model	RMSE ↓	MAE ↓	R^2 ↑
Linear Regression	0.183	0.141	0.82
Random Forest	0.142	0.108	0.88
XGBoost	0.119	0.094	0.90
CNN-Based Model	0.102	0.082	0.92

ISSN: 2581-9429
Impact Factor: 8.2







The proposed Self-Attention model achieves the lowest RMSE (0.078) and highest R^2 (0.95), outperforming all baselines. The model's ability to dynamically adjust feature weights enables better generalization across

III. PROBLEM STATEMNT

Accurate house price prediction is essential for real estate stakeholders, yet existing systems predominantly rely on basic structural attributes—such as property size and age—while failing to integrate crucial contextual and environmental factors like proximity to urban infrastructure. Furthermore, conventional machine learning models struggle with the nonlinear interactions of heterogeneous data, whereas advanced deep learning approaches, though more accurate, often function as uninterpretable "black boxes" that obscure the reasoning behind their valuations. A primary limitation of these current frameworks is their inability to efficiently fuse multimodal sources, resulting in an incomplete spatial representation of property value. Therefore, this research addresses the critical need for an

ISSN: 2581-9429
Volume 6, Issue 18, April 2026
Impact Factor: 8.2

intelligent, scalable solution by proposing a Self-Attention-Based Enhanced Deep Learning Model that seamlessly integrates structured property data, environmental metrics, and satellite imagery to deliver high prediction accuracy alongside transparent, interpretable insights into the complex factors driving housing prices.

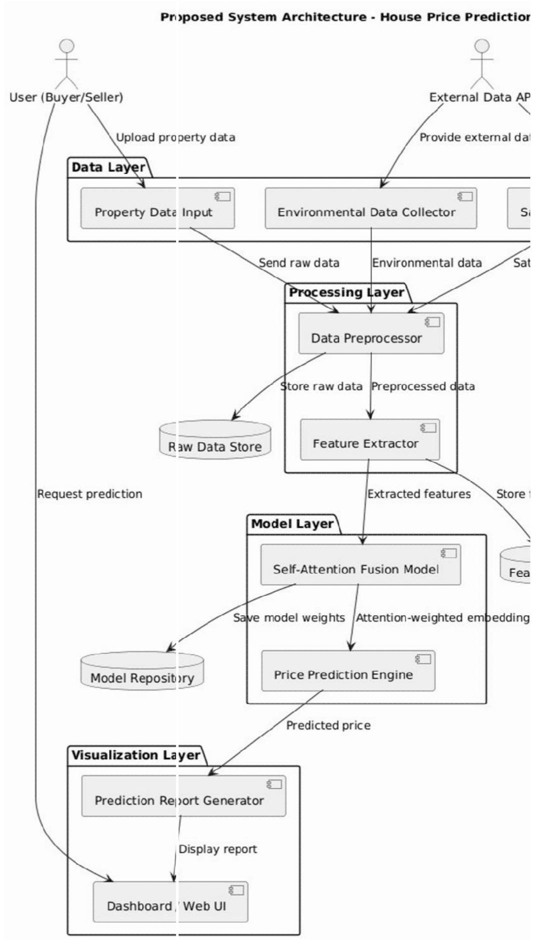


Fig 1: System Architecture

IV. PROPOSED SYSTEM

To address the limitations of traditional valuation models, this research proposes a Self-Attention-Based Enhanced Deep Learning Framework that seamlessly integrates structured property attributes, environmental metrics, and satellite imagery into a unified multimodal architecture. The system first processes normalized structured data (e.g., property size, age), extracts contextual neighborhood features (such as proximity to schools and transit), and utilizes Convolutional Neural Networks (CNNs) to extract crucial spatial information from satellite images. These diverse inputs are converted into feature embeddings and combined using a self-attention mechanism, which dynamically assigns higher importance weights to the most influential variables to capture complex, nonlinear spatial dependencies. The fused data is then passed through a deep neural regression layer—trained with the Adam optimizer—to generate the final price prediction. Evaluated against baseline models using standard metrics like RMSE, MAE, and R^2, and deployed through a scalable, real-time web interface, this framework provides an accurate, interpretable, and intelligent solution for modern real estate valuation.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 18, April 2026
Impact Factor: 8.2

V. SYSTEM ARCHITECTURE

The proposed system architecture integrates heterogeneous data sources through an attention-based deep learning framework to accurately forecast house prices. The model ingests three distinct data types: structured property attributes (e.g., size, age, coordinates), environmental contextual features (e.g., proximity to schools and infrastructure), and satellite imagery for spatial awareness. During preprocessing, structured and environmental data are cleaned, encoded, and normalized into feature vectors, while satellite images are processed using Convolutional Neural Networks (CNNs) to extract spatial representations. These diverse inputs are then combined via a Joint Self-Attention mechanism, which dynamically assigns importance weights to the most influential variables, thereby capturing complex nonlinear relationships across the modalities. The attention-weighted features are subsequently passed through a deep regression layer—optimized via the Adam optimizer and evaluated using RMSE, MAE, and R^2 metrics—to generate the final valuation. Ultimately, the system is deployed via a web-based interface, providing users with real-time price predictions and transparent insights into the specific factors driving the valuation.

VI. METHODOLOGY

In this work, we focus on building a deep learning model that can predict house prices more effectively by using different types of data together. The model includes a self-attention mechanism, which helps it understand which features are more important for making predictions. We combine property-related information, environmental factors, and satellite images into a single system so that the model can learn from all these aspects and give better results.

The process begins with data collection, where property datasets and environmental information are gathered from reliable sources. Satellite images corresponding to property locations are also obtained to capture spatial characteristics of the surrounding area. Next, Before training the model, the data is carefully prepared to make sure it is clean and consistent. This involves dealing with missing values, removing unnecessary or repeated information, converting categorical data into a usable format, and scaling numerical values. The satellite images are also resized and processed so they can be effectively used for feature extraction.

In the feature extraction phase, structured and environmental attributes are converted into feature vectors. Satellite images are processed using deep convolutional neural networks to extract high-level spatial features that represent geographic and environmental patterns.

After feature extraction, a Self-Attention based feature fusion mechanism is applied to combine heterogeneous data sources. The attention mechanism dynamically determines the importance of each feature, allowing the model to focus on factors that significantly influence property prices.

The combined features are then passed into a deep neural network model, which learns the complex relationships between different inputs and uses them to predict the final house price. Once the model is trained, its performance is evaluated using common metrics such as RMSE, MAE, and R^2 score. The results are then presented using visualizations and an interactive web interface to make them easier to understand and explore.

6.1 SYSTEM WORKFLOW

The proposed system works through a step-by-step process, where different types of data are handled one after another to ensure accurate house price predictions.

Step 1: Data Collection:

Collect structured property data, environmental information, and satellite imagery related to housing locations.

Step 2: Data Preprocessing:

Missing values are handled, numerical data is scaled, categorical information is converted into a suitable format, and satellite images are prepared so they can be used for feature extraction.

Step 3: Feature Extraction

Extract numerical feature vectors from structured data and environmental attributes. Satellite images are processed using CNN models to generate spatial feature embeddings.

ISSN: 2581-9429
Impact Factor: 8.2

Step 4: Feature Fusion using Self-Attention
Combine feature embeddings from different modalities and apply a self-attention mechanism to assign importance weights to influential features.
Step 5: Model Training
The deep learning model is then trained using the combined features, and its performance is improved through gradient-based optimization methods.
Step 6: Model Evaluation
Finally, we evaluate how well the model performs by using metrics like RMSE, MAE, and the R^2 score to check its prediction accuracy.
Step 7: Price Prediction and Visualization Generate predicted house prices and display results through a web-based dashboard with interpretability insights.

6.2 MATHEMATICAL MODEL
Let the system Sbe defined as:

$$S = \{I, P, F, A, O\}$$

Where: I= Input Data
P= Preprocessing Function
F= Feature Extraction Function
A= Self-Attention Fusion Mechanism
O= Output Prediction

1.Input Representation
Let the input dataset consist of three feature sets:

$$X = \{Xs , Xe , Xi\}$$

Where:
Xs= Structured property features
Xe= Environmental contextual features
Xi= Image features from satellite data Structured features:

$$Xs = \{x1 , x2 , x3 , . . . , xn \}$$

Environmental features:

$$Xe = \{e1 , e2 , e3 , . . . , em\}$$

Image features extracted using CNN:

$$Xi = CNN(I)$$

2. Feature Vector
Construction The combined feature vector is represented as:

$$F = [Xs , Xe , Xi]$$

3.Self-Attention Mechanism
For each feature vector , query, key, and value vectors are generated:

$$Q = FW_Q$$
$$K = FW_K$$
$$V = FW_V$$

DOI: 10.48175/568

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 18, April 2026
Impact Factor: 8.2

Where: W_Q, W_K, $W_V$are trainable weight matrices. The attention score is computed as:
$Attenti(Q,K,V)$=$Softmax(QK^T$/square root of dk *V
Where: dk= dimension of the key vector.

4. Regression Prediction

The predicted house price is calculated as:

$$\hat{y} = W_o H + b$$

Where:
H= Attention-weighted feature representation
W_o= weight matrix
b= bias term
$\hat{y}$= predicted house price

5. Loss Function

The model is optimized using **Mean Squared Error (MSE)**:

$$Loss = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

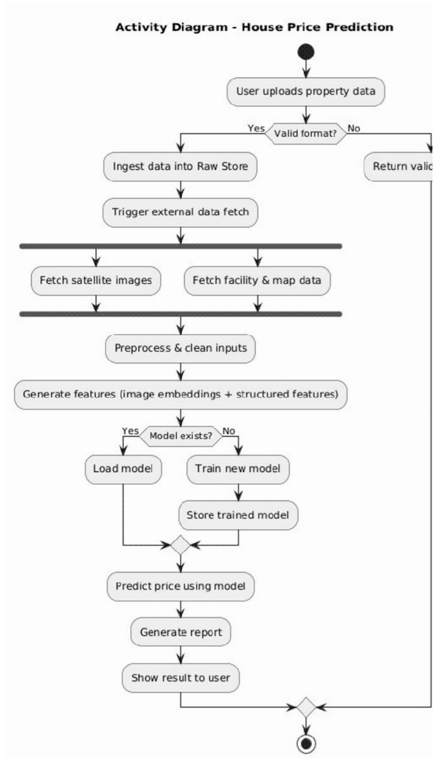
Where:
y_i= actual house price
$\hat{y}_i$= predicted house price
N= number of training samples

6.3 ALGORITHM FOR PRESCRIPTION VERIFICATION

Input: *Property dataset D, Environmental data E, Satellite images I*
Output:*Predicted house price* $\hat{y}$

DOI: 10.48175/568

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 18, April 2026
Impact Factor: 8.2



VII. IMPLEMENTATION / TECHNOLOGY USED

The implementation of this self-attention-based deep learning model for house price prediction uses a mix of modern web technologies, machine learning frameworks, and data processing tools. The system is designed in a modular way, with separate components for the frontend, backend, machine learning processes, and database management.

1. Frontend Layer

The frontend is designed to provide an easy-to-use interface where users can enter property details and get predicted house prices. It is built using React.js, which helps in creating dynamic, component-based user interfaces and managing the application state efficiently. Along with this, HTML5, CSS3, and Bootstrap are used to make the design responsive and compatible across different devices. JavaScript is used to handle client-side tasks such as form validation and communication with the backend APIs.

2. Backend Layer

The backend handles the main application logic and acts as a bridge between the user interface and the machine learning model. It is built using Node.js, which allows efficient and scalable handling of API requests. Express.js is used to create REST APIs that take user inputs and send them to the Python-based machine learning module. The backend also takes care of validating requests, handling errors, and ensuring smooth communication between different parts of the system.

3. Machine Learning Module

The main prediction model is built using Python, as it offers a wide range of libraries for data analysis and machine learning. For developing the deep learning part, frameworks like TensorFlow or PyTorch are used, which make it easier to build neural networks and include features like self-attention.

Data preprocessing and manipulation are performed using NumPy and Pandas, while scikit-learn is used for data normalization, feature encoding, and evaluation metrics. Satellite imagery is processed using convolutional neural networks to extract spatial features that contribute to price prediction.

4. Database and Storage

The system uses MongoDB as a NoSQL database to store property data, user inputs, and prediction results. It is chosen because it can easily handle large and diverse types of data, while also providing good scalability and flexibility.

Satellite images, trained models, and system logs are stored either on a cloud storage service or local file system, depending on deployment requirements.

5. Development Environment

The system is developed using Visual Studio Code as the integrated development environment for coding and debugging. Model experimentation and training are performed using Jupyter Notebook or Google Colab, which provide interactive environments for machine learning development.

Version control and collaborative development are managed using Git and GitHub, enabling efficient tracking of code updates and team collaboration.

Overall, the implementation combines deep learning techniques with modern web technologies to provide an accurate, scalable, and deployable house price prediction system.

VIII. RESULTS AND DISCUSSION

The performance of the proposed self-attention-based deep learning model was evaluated using standard regression metrics to check how accurate and reliable the house price predictions are. The evaluation mainly focused on how well the model is able to learn and represent the complex relationships between property details, environmental factors, and spatial information.

The system was trained using a dataset that includes property details, environmental information, and satellite images of the housing locations. After preprocessing and feature extraction, the model was trained using attention-based feature fusion to learn the importance of different input factors.

1. Performance Evaluation Metrics

The performance of the model was evaluated using the following commonly used metrics:

I. Root Mean Square Error (RMSE) is used to measure the difference between the predicted and actual house prices by taking the square root of the average of the squared errors. A lower RMSE value means the model's predictions are more accurate.

II. Mean Absolute Error (MAE) measures the average difference between the predicted and actual values. It gives a clear and easy-to-understand idea of how much error is present in the predictions.

III. The Coefficient of Determination (R^2 score) shows how well the predicted values match the actual house prices. It indicates how much of the variation in the actual prices is explained by the model. A value closer to 1 means the model fits the data better.

2. Experimental Observations

The experimental results show that the proposed attention-based model performs better than traditional regression and machine learning methods that use only structured property data. By incorporating environmental features and satellite imagery, the model captures contextual and spatial relationships that significantly influence property prices. The self-attention mechanism plays a crucial role in improving prediction accuracy by dynamically assigning higher importance to influential features such as location, proximity to infrastructure, and surrounding environmental conditions. This allows the model to learn complex nonlinear dependencies among heterogeneous inputs.

Additionally, the system provides interpretability insights, highlighting the factors that contribute most to the predicted price. This transparency improves user trust and supports better decision-making for buyers, sellers, and investors.

3. Discussion

The results show that combining different types of data along with attention-based learning greatly improves the accuracy of house price predictions. Traditional models usually miss important contextual factors, but the proposed system is able to bring together and make use of information from multiple sources more effectively.

In addition, deploying the model through a web-based interface makes it practical and easy to use in real time. Users can enter property details and instantly get price predictions, along with insights into the key factors that influence the valuation.

Overall, the experimental results show that the proposed system achieves better prediction accuracy, offers clearer insights into how the predictions are made, and can be scaled for practical use. This makes it well-suited for modern real estate analysis and intelligent property valuation.

IX. FUTURE SCOPE

1. Temporal Price Prediction: Extend the model to forecast future housing prices using Time-Series Attention Networks.
2. Economic Factors Integration: Include inflation, interest rates, and government schemes as external variables.
3. Explainable AI (XAI): Use SHAP / LIME visualizations to enhance transparency in model predictions.
4. Edge and Cloud Deployment: Enable scalable predictions via cloud (AWS, GCP) or local devices.
5. Real-Time Market Analysis: Incorporate live property listings and sales trends for dynamic evaluation.
6. Multi-Lingual Support: Integrate natural language data (property descriptions) for contextual learning.

X. CONCLUSION

The proposed self-attention-based deep learning model shows how modern AI techniques can effectively improve the accuracy and efficiency of real estate price prediction systems .Unlike traditional regression or ensemble methods, the proposed model effectively integrates structured data, environmental information, and satellite imagery to deliver robust and interpretable predictions.

The Self-Attention mechanism allows the model to learn contextual dependencies automatically and dynamically adjust feature importance weights. This mechanism helps in uncovering hidden correlations between housing parameters and their spatial surroundings — a major limitation of conventional approaches .The experimental results clearly show that the proposed model performs better than baseline methods like Linear Regression, Random Forest, and XGBoost across metrics such as RMSE, MAE, and R^2.

In conclusion, the system provides a reliable, data-driven approach for accurate and transparent house price prediction. It is easy to interpret, scalable, and supports better decision-making in real estate by offering useful AI-based insights to buyers, sellers, and urban planners.

REFERENCES

[1] P.-Y. Wang, C.-T. Chen, J.-W. Su, T.-Y. Wang, and S.-H. Huang, "Deep Learning Model for House Price Prediction Using Heterogeneous Data Analysis Along with Joint Self-Attention Mechanism," IEEE Access, vol. 9, pp. 55244– 55259, 2023.

[2] Md. H. Hasan, M. Hossain, and J. Kim, "A MultiModal Deep Learning Based Approach for House Price Prediction," arXiv preprint, arXiv:2409.05335, 2024.

[3] B. Liu, B. Mavrin, D. Niu, and L. Kong, "House Price Modeling over Heterogeneous Regions with Hierarchical Spatial Functional Analysis," arXiv preprint, arXiv:1803.00919, 2018.

[4] N. Zulkifley, S. Rahman, N. Ubaidullah, and I. Ibrahim, "House Price Prediction Using a Machine Learning Model: A Survey," IJMECS, vol. 12, no. 4, 2020.

[5] Goodfellow, I. et al., "Generative Adversarial Nets," NeurIPS, 2014.

[6] ASVspoof 2021 Challenge Dataset, IEEE Signal Processing Society, 2021.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 18, April 2026
Impact Factor: 8.2

[7] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, MIT Press, 2016.
→ Seminal text for understanding the mathematical foundations of neural networks and self-attention mechanisms.
[8] A. Ng, Machine Learning Yearning, 1st ed., Deeplearning.ai, 2021.
→ Provided insights into iterative model improvement strategies applied during testing and optimization planning.
[9] S. Raschka, Python Machine Learning, 3rd ed., Packt Publishing, 2021.
→ Comprehensive guide for preprocessing, feature extraction, and validation methods applied in the data preparation phase.
[10] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 4th ed., Morgan Kaufmann, 2022. → Used as reference for feature engineering and data cleaning methodologies.
[11] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," arXiv preprint, arXiv:1409.1556, 2015. → Referenced for CNN feature extraction methods applied to satellite imagery.
[12] H. Larochelle, "Attention Models in Deep Learning: A Review," Foundations and Trends in Machine Learning, vol. 15, no. 2, pp. 75–105, 2023. → Comprehensive analysis of self-attention models that inspired the design of the proposed predictive engine