

Postgraduate Seminar Series

Venue: Graduate Seminar Room

Date & Time: July 18, 2026 at 2:30 PM

Speaker Information

Khandokar Md. Rahat Hossain (Std No. 0423052079) is a part time M.Sc. student in the department of CSE, BUET. He completed his undergraduate studies from Bangladesh University of Engineering and Technology (BUET) in 2023. His research interest lies in the field(s) of Machine Learning and Deep Learning. He is currently doing his postgraduate thesis under the supervision of Dr. Tanzima Hashem. He will be speaking about his ongoing research in this talk.

LLM-Enhanced Multimodal House Price Prediction using Structured, Textual, and Visual Features

Accurate house price prediction is essential for real estate buyers, sellers, agents, and financial institutions. Traditional methods rely primarily on structured data. While useful, these approaches often overlook rich contextual information embedded in textual descriptions and property photos, which can significantly affect how much buyers are willing to pay. Previous deep learning approaches have fused multimodal data but they fall short to extract structured spatio-temporal and contextual features from texts and images. This study introduces a novel multimodal framework that addresses these limitations by systematically integrating structured attributes, textual descriptions, and visual images of properties to achieve higher prediction accuracy while maintaining interpretability. From the textual descriptions, meaningful spatio-temporal and contextual features are extracted such as proximity to amenities, neighborhood characteristics, renovation status etc. In practice, a large language model is used to derive domain-specific features from textual descriptions, and an unsupervised vision-language approach is introduced to automatically categorize property photos into four semantically meaningful classes that are commonly present in real-world house listings. These categorized visual features help to capture aspects such as the interior quality and curb appeal of properties. The categories are created without manual labeling by aligning image clusters with carefully designed pseudo text prompts. The features from all three sources are then combined into a unified representation that is used to train a prediction model. An optimized regressor is used to predict the final prices. Experiments on large real-world property datasets demonstrate that this method significantly outperforms traditional methods and existing multimodal baselines. Ablation studies confirm that the novel text feature extraction and categorized visual embeddings provide the largest contributions to more accurate and reliable price predictions. Feature contributions are also analyzed in order, offering clearer explanations of the results. Overall, this study demonstrates how integrating structured, textual, and visual information in a structured and interpretable way can lead to more accurate and trustworthy house price predictions. The framework is designed to be adaptable to different markets and datasets.