There is a great deal of literature on the subject of using machine learning models to predict housing and property prices. Ahtesham et al state in their study that using an XGBoost algorithm has yielded the best and most accurate result with a model accuracy of 98%. The author also suggests that XGboost is faster than some alternatives for their dataset as it is a suitable algorithm for tabular datasets and is capable of parallel processing making it up to 10 times faster than alternative models [3].

In a study conducted by Kumar et al, a system that automatically predicts the price of a house the most accurately was designed. Multiple regression algorithms were compared and the performance of each model was assessed Root Mean Square Error (RMSE). This metric was chosen as it punishes outliers greatly and would thus help differentiate between the performance of each model. They tested many models and found that Catboost regression yielded the lowest RMSE. Catboost regressor is an algorithm well suited for text, image, and historical, as well as requiring less training than other similar models. It was well suited for this study as the dataset was of historical sales data [4].

Some studies sought to use neural networks in conjunction with regression algorithms. One such study conducted by Varma et al utilized linear regression, forest regression, and boosted regression initially. The results were then fed into a neural network that was applied with boosted regression to increase the accuracy of the model. The neural network improves the results by computing the results fed into it from the previous regression models and computes the best result out of them. The study concludes that a combination of multiple regression models achieves better results than a single model. [5]

Work done by Truong et al looked at using multiple machine learning models to identify the differences amongst them. 5 models were tested: Random forest, XGBoost, LightBGM, Hybrid regression, and Stacked Generalization. Hybrid regression and stacked generalization utilize more than one regression algorithm in their model. Hybrid regression is a combination of, in the case of this study, 3 algorithms; Random Forest, XGBoost, and LightGBM. Stacked generalization consists of a 2-level stacking architecture where the first level consists of a regression algorithm whose results are fed into the next level as features for another regression algorithm. The results of the study found that Random Forest showed the best results but is prone to overfitting, while XGBoost and LightGBM weren’t prone to overfitting but not as accurate. While these 3 models required tuning and optimization to produce satisfactory results, Hybrid regression and stacked generalization did not require any tuning or optimization. Interestingly stacked generalization did poorly on the training set but performed well on the test set. This is likely due to the use of 5-fold cross-validation and a coupling effect with the regression models in use where they supported each other to produce better results. The only problem with it is that the time complexity is quite high. Overall the tested models all yielded satisfactory results with some advantages over the others, the best model would depend on the dataset being used. [6]

There was research carried out to evaluate the use of deep learning alongside XGBoost regression in estimating the price of real estate. Zhao et al sought to incorporate images of property in conjunction with tabular data to increase the accuracy of predictions. A hybrid neural network was built consisting of 4 components: A convolutional neural network was trained on AVA to give scores to property images. A multilayer perceptron model to analyze data from a table. One more CNN to extract specific visual features from the images. And finally, XGBoost regression to predict the final price of the property. The study found that a hybrid neural network improved the accuracy of the predictions by utilizing features that are typically not considered, such as images showing the aesthetic or quality of a house. [7]

Identifying the major factors that affect the price of a house can help in identifying major features. In [8], the Spearman correlation coefficient was used to identify the major factors that affect the price: Environmental factors (the surroundings of the house), House factors (the specifications of the house itself), and transportation factors. These factors were then used in a multiple linear regression to predict the prices of houses. The study shows that identifying key features can improve the performance of a regression model. [8]

In a study seeking to predict real estate prices in Finland, [9] sought to improve the existing artificial neural networks’ performance by automatically optimizing their hyperparameters. A baseline ANN model was used to compare against the optimized models. The fine-tuning of the hyperparameters was done automatically using a Bayesian algorithm, using the mean square error as the metric to measure the improvement. Utilizing Bayesian optimization improved the model in all metrics measured and did not show any over or underfitting. The paper concludes that using automatic tuning for the hyperparameters of a neural network is an effective way to improve the accuracy of results, and perhaps expanding the optimization process may yield even better results.

The Journal of Property Study featured a paper on using machine learning algorithms to predict property prices. 3 regression models were trained and tested on a dataset of 40,000 housing transactions over 18 years in Hong Kong. The 3 models tested were: Support vector machine (SVM), Random Forest (RF), and Gradient Boost Machine (GBM). The results of the study concluded that RF and GBM performed the best in terms of accuracy. SVM, while not as accurate, was faster than both models. [10]

[3] M. Ahtesham, N. Z. Bawany and K. Fatima, "House Price Prediction using Machine Learning Algorithm - The Case of Karachi City, Pakistan," 2020 21st International Arab Conference on Information Technology (ACIT), 2020, pp. 1-5, doi: 10.1109/ACIT50332.2020.9300074.

[4] G. K. Kumar, D. M. Rani, N. Koppula and S. Ashraf, "Prediction of House Price Using Machine Learning Algorithms," 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI), 2021, pp. 1268-1271, doi: 10.1109/ICOEI51242.2021.9452820.

[5] A. Varma, A. Sarma, S. Doshi and R. Nair, "House Price Prediction Using Machine Learning and Neural Networks," 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), 2018, pp. 1936-1939, doi: 10.1109/ICICCT.2018.8473231.

[6] Q. Truong, M. Nguyen, H. Dang, and B. Mei, “Housing price prediction via Improved Machine Learning Techniques,” Procedia Computer Science, vol. 174, pp. 433–442, 2020.

[7] Y. Zhao, G. Chetty and D. Tran, "Deep Learning with XGBoost for Real Estate Appraisal," 2019 IEEE Symposium Series on Computational Intelligence (SSCI), 2019, pp. 1396-1401, doi: 10.1109/SSCI44817.2019.9002790.

[8] Q. Zhang, “Housing price prediction based on multiple linear regression,” *Scientific Programming*, vol. 2021, pp. 1–9, 2021.

[9] J. Kalliola, J. Kapočiūtė-Dzikienė, and R. Damaševičius, “Neural network hyperparameter optimization for prediction of real estate prices in Helsinki,” *PeerJ Computer Science*, vol. 7, 2021.

[10] W. K. O. Ho, B.-S. Tang, and S. W. Wong, “Predicting property prices with machine learning algorithms,” *Journal of Property Research*, vol. 38, no. 1, pp. 48–70, 2020.