Deep Neural Network As a Tool for Appraising Housing Prices: A Case Study of Busan, South Korea

S An\textsuperscript{1,2}, Y Song\textsuperscript{3}, H Jang\textsuperscript{4}, and K Ahn\textsuperscript{1,2, *}

\textsuperscript{1} Department of Industrial Engineering, Yonsei University, South Korea
\textsuperscript{2} Center for Finance and Technology, Yonsei University, South Korea
\textsuperscript{3} Department of Geography, Chonnam National University, South Korea
\textsuperscript{4} Department of Finance, Accounting and Risk, Glasgow Caledonian University, UK

\* E-mail: k.ahn@yonsei.ac.kr (K.A.)

Abstract. This study examines whether the number of hidden layers in a deep neural network significantly influences the model accuracy and efficiency for appraising housing prices. We provide empirical evidence that the deep neural network can achieve high accuracy with a small number of hidden layers on our dataset, which contains various hedonic variables. Furthermore, we show that adding layers does not necessarily guarantee the model’s accuracy and effectiveness of the computing time.

1. Introduction
Housing market is a crucial part of urban economy, and its price level reflects regional economic factors. At a personal or household level, housing accounts for a large portion of wealth \cite{1}. Therefore, various stakeholders, including governments, investors, and academics, have tried to maintain the stability and effectiveness of the housing market by developing a model for appraising housing prices \cite{2–4}.

Various approaches and variables have been introduced to appraise housing prices. Housing prices are determined by various factors, such as housing properties, local amenities, and local demographic characteristics. For instance, transportation accessibility was used as a key variable along with housing and local properties using a hedonic price model (HPM) \cite{5}. Furthermore, recent improvements in the image recognition technology enable us to generate novel variables such as green or sky view index, which can enhance the accuracy and effectiveness of the model \cite{6,7}. Datasets used to estimate housing prices cover various economic, functional, and amenity factors; therefore, they are often nonlinear and complex. However, traditional models such as HPM rely on a simple linear regression model, and whether HPM properly captures the complex nature of variables and their relationships is questionable \cite{8,9}.

Considering the complexity and nonlinear property of housing price data, this study develops an appraisal model with a deep neural network (DNN). The DNN has been proved suitable for nonlinear datasets and applied in diverse areas of study, including disease detection and energy consumption prediction \cite{10,11}. Thus, it can be used for appraising housing prices with high accuracy. Moreover, few studies have examined the optimal number of layers, which is important for an accurate and efficient appraisal of housing prices \cite{12,13}. Therefore, this study predicts housing prices using DNN and investigates how many layers are required to capture complex and nonlinear hedonic variables.
2. Method and data

2.1. DNN

DNN comprises an input layer, some hidden layers, and an output layer. We assume the structure of DNN with one to six hidden layers and investigate the optimal number of layers considering the accuracy and efficiency as model performance. Each hidden layer has 64 neurons and the same activation function, with a fully connected ReLU function, is applied, i.e., the most popular activation function in neural networks imposing nonlinearity on the layer structure. We define the ReLU function, which can enhance the model accuracy and prevent gradient from vanishing [14], as

$$\text{ReLU}(x) = \max(0, x),$$

where $x$ stands for an input of the current node. Only if input value $x$ is positive, the activation function conveys value $x$ to nodes in the next layer. Accordingly, ReLU function can significantly reduce the computation cost as it facilitates the calculation of propagating weights by deactivating neurons [15]. Through this feedforward propagation, DNN predicts the target value by aggregating whole weights and inputs of each node and computes the error generated by loss function. Then, DNN starts to update each weight for minimizing the bias through back propagation as follows:

$$w_{n+1} = w_n - \alpha \left( \frac{\partial l}{\partial w_n} \right),$$

where $l$ and $\alpha$ are the loss function and learning rate, respectively. Specifically, we set the learning rate to 0.01 and ‘early stopping’ to 10% of the maximum number of epochs. For a smooth training procedure, we use the Min–Max scaling instead of the Z-score normalization [16]. The model structure is presented in figure 1. To appraise the model’s accuracy, we select mean absolute error (MAE) and R-square, and further estimate the computation time in the model training.

![Figure 1. Model structure.](image)

2.2. Data

We utilize the Mendeley data repository ([https://data.mendeley.com/datasets/d7grg846wv/3](https://data.mendeley.com/datasets/d7grg846wv/3)), which was introduced by Song et al. [17], to investigate the impact of extra layers on the DNN model for appraising housing prices. This dataset, which contains 61,152 observations in 2015, comprises four Korean metropolitans: Busan, Daegu, Daejeon, and Gwangju. To explore the impact of a subway network on housing prices in our study, we concentrate on Busan only because it showed the highest modal share of subway, i.e., 17.8%, among the four cities [5]. The prevalent type of housing in Busan is condominium, which is the dominant housing type in South Korea [18]. Herein, 17 variables are used for appraising the housing prices, as shown in table 1.
Table 1. Descriptive statistics.

| Variables          | Scale  | Min.  | Max.  | Mean  | Std.  | Skewness | Kurtosis |
|--------------------|--------|-------|-------|-------|-------|----------|----------|
| Higher degree Ratio | Ratio  | 10.16 | 59.10 | 35.40 | 10.77 | 0.26     | -0.89    |
| Year Date          | 1962   | 2015  | 1999.09 | 9.23 | -0.35 | -0.19    |
| Parking Ratio      | 0      | 15    | 1     | 0.50  | 0.94  | 12.88    |
| Area Ratio         | 14.74  | 250.85 | 79.27 | 30.27 | 1.13  | 1.93     |
| Households Ratio   | 4      | 5239  | 871.75 | 849.79 | 1.92 | 5.01     |
| CBD Ratio          | 244.28 | 23900.12 | 9003.08 | 4917.75 | 0.69 | -0.03    |
| Floor Interval     | -1     | 79    | 10.56 | 7.90  | 1.44  | 3.96     |
| ln_greenspacea     | Ratio  | -1.55 | 8.33  | 5.21  | 1.06  | -0.36    | 1.52     |
| ln_waterfronta     | Ratio  | 1.75  | 7.84  | 6.08  | 0.92  | -0.61    | 0.10     |
| ln_NET_DISTa       | Ratio  | 0.52  | 9.52  | 6.99  | 1.01  | -0.99    | 3.47     |
| SNU Ratio          | 0      | 12    | 1.17  | 1.85  | 3.27  | 13.49    |
| Dummy_Springb      | Nominal | 0    | 1     | 0.31  | 0.46  |          |
| Dummy_Fallb        | Nominal | 0    | 1     | 0.25  | 0.43  |          |
| Dummy_Winterb      | Nominal | 0    | 1     | 0.20  | 0.40  |          |
| PopDensity Ratio   | 140.41 | 38720 | 12978.93 | 7319.40 | 0.66 | 0.15     |
| Heating typec      | Nominal | 0    | 1     | 0.91  | 0.28  |          |
| Bus_Count Ratio    | 0      | 46    | 12.20 | 6.98  | 0.80  | 0.40     |

a This variable is transformed with logarithm.

b Control variables of sales period: spring, fall, and winter.

c Dummy variable: city gas 0, others 1.

3. Results and discussion

The model accuracy and computation time with additional layers are summarized in table 2. As the number of layers in DNN increases, it becomes more accurate in terms of MAE and its computation requires more time. However, adding more layers does not necessarily guarantee a higher accuracy than others in terms of R-square. For example, the model with four layers (i) achieves the smaller R-square and (ii) reduces the computation time of models with five and six layers by approximately 30%. This suggests that the optimal DNN model can be obtained with few layers when big data are used.

Table 2. Model accuracy and efficiency.

|                | MAE  | R-square | Computation time (s) |
|----------------|------|----------|-----------------------|
| Layer 1        | 0.017| 0.836    | 83.387                |
| Layer 2        | 0.011| 0.895    | 85.373                |
| Layer 3        | 0.009| 0.915    | 126.813               |
| Layer 4        | 0.009| 0.913    | 88.359                |
| Layer 5        | 0.009| 0.918    | 124.051               |
| Layer 6        | 0.008| 0.924    | 124.697               |

DNN optimizes the weight of each layer using various methods such as Adam or RMSProp and minimizes the bias using the stochastic gradient descent algorithm. Consequently, few layers can produce reasonable outcomes in terms of accuracy and efficiency. Most importantly, the performance
of DNN is closely related to preventing the overfitting issue. Multiple convolution kernel matrices and training parameters deteriorate the training time and space complexity, resulting in a low speed of network propagation and inaccuracy [19].

Figure 2 shows the average impact of each feature on the model output. We calculate the Shapley Additive exPlanations (SHAP) [20,21] values and compare model predictions with and without each feature. When an individual feature contributes more to appraising housing prices, SHAP value is high. We observe that after removing the subway-related variable, the model’s accuracy barely decreases. Moreover, the subway network depreciates the housing prices in contrast to Ahn et al. [5]. Thus, the expected positive effects of transit accessibility do not exist in the Busan metropolitan area.

![SHAP value plot](image)

**Figure 2.** SHAP value plot.

4. Conclusion
This study investigates the number of efficient hidden layers for appraising housing prices in DNN. By comparing the accuracy and computation time, we find that DNN can effectively appraise housing prices with few layers, and an efficient approach involves adding more layers with enough number of neurons.

Acknowledgments
This work was supported by (i) the Technology Innovation Program ATC+ (20014125, Development of Intelligent Management Solution for Nuclear Decommissioning Site Characterization) funded by the Ministry of Trade, Industry & Energy (MOTIE, Republic of Korea) and (ii) the Future-leading Research Initiative (Grant Number: 2021-22-0306) funded by Yonsei University.

References
[1] Glaeser E L, Gyourko J, Morales E and Nathanson C G 2014 Housing dynamics: An urban approach Journal of Urban Economics 81 45–56
[2] Rizi M H 2021 What moves housing markets: A state-space approach of the price–income ratio
International Economics 167 96–107

[3] Xue C, Ju Y, Li S and Zhou Q 2020 Research on the sustainable development of urban housing price based on transport accessibility: A case study of Xi’an, China Sustainability 12 1497

[4] Pai P P and Wang W C 2020 Using machine learning models and actual transaction data for predicting real estate prices Applied Sciences 10 5832

[5] Ahn K, Jang H and Song Y 2020 Economic impacts of being close to subway networks: A case study of Korean metropolitan areas Research in Transportation Economics 83 100900

[6] Zhang Y and Dong R 2018 Impacts of street-visible greenery on housing prices: Evidence from a hedonic price model and a massive street view image dataset in Beijing ISPRS International Journal of Geo-Information 7 104

[7] Chen L, Yao X, Liu Y, Zhu Y, Chen W, Zhao X and Chi T 2020 Measuring impacts of urban environmental elements on housing prices based on multisource data: A case study of Shanghai, China ISPRS International Journal of Geo-Information 9 106

[8] Hong J, Choi H and Kim W S 2020 A house price valuation based on the random forest approach: The mass appraisal of residential property in South Korea International Journal of Strategic Property Management 24 140–152

[9] Chin T L and Chau K W 2003 A critical review of literature on the hedonic price model International Journal for Housing Science and its Applications 27 145–165

[10] Ahn K, Jang H and Song Y 2020 Detection and prediction of house price bubbles: Evidence from a new city Lecture Notes in Computer Science 10862 782–795

[11] Chen H, Geng L, Zhao H, Zhao C and Liu A 2021 Image recognition algorithm based on artificial intelligence Neural Computing and Applications 33 1–12

[12] Bickel D R 2020 Testing prediction algorithms as null hypotheses: Application to assessing the performance of deep neural networks Stat 9 e270

[13] Mayer M, Bourassa S C, Hoesli M and Scognamiglio D 2019 Estimation and updating methods for hedonic valuation Journal of European Real Estate Research 12 134–150

[14] Kalogirou S A and Bojic M 2020 Artificial neural networks for the prediction of the energy consumption of a passive solar building Energy 25 479–491

[15] Bickel D R 2020 Testing prediction algorithms as null hypotheses: Application to assessing the performance of deep neural networks Stat 9 e270

[16] Mayer M, Bourassa S C, Hoesli M and Scognamiglio D 2019 Estimation and updating methods for hedonic valuation Journal of European Real Estate Research 12 134–150

[17] Rahim T, Hassan S A and Shin S Y 2021 A deep convolutional neural network for the detection of polyps in colonoscopy images Biomedical Signal Processing and Control 68 102654

[18] Kulathunga N, Ranasinghe N R, Drincjencu D, Kinsman Z, Huang L and Wang Y 2021 Effects of nonlinearity and network architecture on the performance of supervised neural networks Algorithms 14 51

[19] Song Y, Ahn K, An S and Jang H 2021 Hedonic dataset of the metropolitan housing market-Cases in South Korea Data in Brief 35 106877

[20] Jabeur S B, Mefteh W S and Viviani J L 2021 Forecasting gold price with the XGBoost algorithm and SHAP interaction values Annals of Operations Research 1–21

[21] Lundberg S M and Lee S I 2018 A unified approach to interpreting model predictions International Conferences on Neural Information Processing Systems (CA, USA, December 2017) (Advances in Neural Information Processing Systems vol 30) ed U V Luxburg et al. pp 4766–4775